Global Forum on Bioethics in Research (GFBR): Ethics of AI in Global Health Research

29 – 30 November 2022

Global Forum on Bioethics in Research

Contents

Introduction

Welcome to the Global Forum on Bioethics in Research (GFBR) meeting on Ethics of AI in Global Health Research.

Artificial intelligence (AI) is increasingly being used in global health research but frameworks, policy and best practice for the ethical review and oversight of AI heath studies is currently lacking. The Forum will discuss how traditional research ethics regulatory frameworks have responded to the rapid advances in AI technology, and what changes are required, including to the role and responsibility of research ethics committees (RECs).

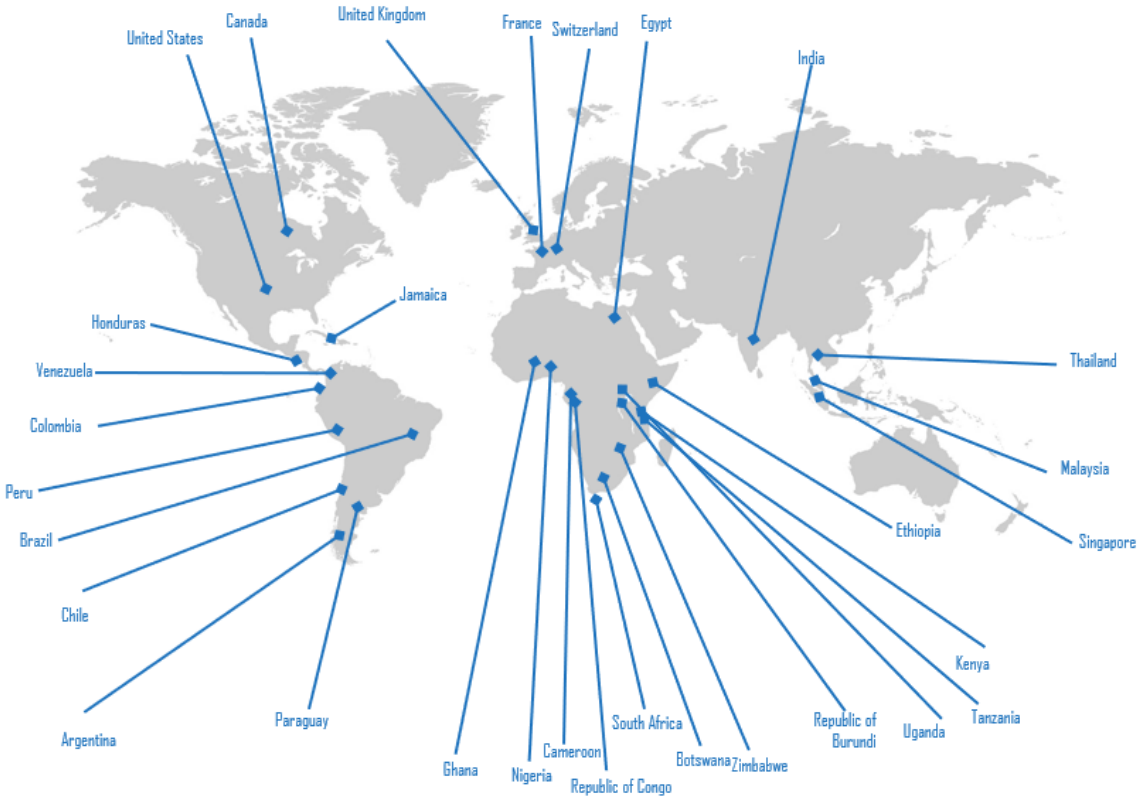
The theme of this meeting provides an exciting opportunity to build on the Forum's legacy as a global platform for debate on ethical issues in international health research. Specifically, the meeting will bring together the global bioethics and research community and regulators to explore the ethical challenges such as 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 will include 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, and low- and middle-income country (LMIC) perspectives have been underrepresented. The Forum will consider the 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.

We are very pleased to have participants from 31 countries (see map of participants' countries) and a range of disciplines. We would like to extend our thanks to our local host the South African Medical Research Council (MRC) for their support in the preparation of the meeting. We would also like to thank the Planning Committee of this meeting and the GFBR funders for their continuing support. We very much hope the meeting will be a positive experience for us all.

The GFBR Steering Committee

Caesar Atuire, Ghana and UK;
Anant Bhan, India;
Phaik Yeong Cheah, Thailand;
Anna Chiumento, UK;
Sharon Kaur, Malaysia
Rachel Knowles, UK (funder representative);
Carleigh Krubiner, UK (funder representative);
Katherine Littler, Switzerland;
Paul Ndebele, USA;
Ana Palmero, Argentina;
Michael Parker, UK;
Carla Saenz, USA;
Barbara Sina, USA (funder representative);
Ross Upshur, Canada;
Teck Chuan Voo, Singapore;
Jantina de Vries, South Africa.

Participants attending Cape Town 2022



Members of the GFBR Planning Committee for this meeting

Joseph Ali, USA;
Caesar Atuire, Ghana and UK;
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.

Map credit: The Pixel/Shutterstock.com

Background to the GFBR

The GFBR seeks to bring researchers, research policy makers and ethicists, among others together to share experiences and promote collaboration around research ethics. The Forum is built around case study presentations to ensure that discussion of the ethical issues remain grounded in the practical realities of how research is conducted 'on the ground', particularly in low resource settings. There are also sessions on governance issues. Compared to traditional meetings, GFBR is unique in that it is limited in size and built around small group discussions of case studies and governance papers that are submitted by participants. The Forum prioritises the participation of colleagues from LMICs, encourages networking and mentoring, and creates a venue for open and inclusive discussions.

Meetings began in Bethesda, USA in 1999 and subsequently convened in: Bangkok, Thailand in 2000; Cape Town, South Africa in 2002; Brasilia, Brazil in 2002; Paris, France in 2004; Blantyre, Malawi in 2005; Karachi, Pakistan in 2006; Vilnius, Lithuania in 2007; and Auckland, New Zealand in 2008.

Following a period to reflect on the structure and funding of the Forum between 2009-13, the GFBR was re-launched at a satellite meeting of the International Association of Bioethics in Mexico City, Mexico in June 2014. It renewed its emphasis on providing a platform for individuals from LMICs to bring forward ethical issues affecting their research practice for dialogue and discussion. Six full meetings have taken place since the re-launch:

- 'Emerging epidemic infections and experimental treatments', France, 2015
- 'The ethics of research in pregnancy', Argentina, 2016
- 'The ethics of alternative clinical trial designs and methods in LMIC research', Thailand, 2017
- 'The ethics of data sharing and biobanking in health research', South Africa, 2018
- 'Genome editing for human benefit: ethics, engagement and governance', Singapore, 2019
- 'Ethical issues arising in research with people with mental health conditions', online, 2021

GFBR is organised by the World Health Organization and is supported by a number of international research funders including Wellcome, UK MRC, the US National Institutes of Health and the South African MRC. A fellowships scheme was launched in 2015 and takes place annually (see page 113).

The key values of the GFBR are to:

- promote ethically conducted research;
- promote global development for health research ethics, particularly in LMICs;
- facilitate partnerships between the global north and south.

GFBR meetings aim to:

- maintain and strengthen the protection of human participants in health research;
- provide a forum for LMIC perspectives on ethical issues in research;
- explore opportunities to enhance capacity for the ethical review of research;
- create a context for scientists, ethicists, community representatives, policy-makers, industry and other relevant stakeholders to collaborate and talk in an environment of mutual cooperation and respect.

These aims are kept under review and refined by the Steering Committee.

Agenda

Tuesday 29 November 2022

08:00	Registration Somerset Conference Room
08:30	**Welcome and introduction** Welcome South African MRC representative **Introduction to GFBR and the meeting** Jantina De Vries, University of Cape Town, South Africa
08:40	**Keynote presentation** Effy Vayena, Health Ethics and Policy Lab, Institute of Translational Medicine, ETHZ, Switzerland
09:20	**Pecha Kucha** Chair: Phaik Yeong Cheah, Mahidol Oxford Tropical Medicine Research Unit, Thailand • **Analysing a local imbalance of power ethics: University of Ghana vs. Data Commission** Athanasius Egyarkoh Afful, University of Ghana, Ghana • **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 • **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 • **Future Nanomedicines: building a regulatory framework for the first-in-human nanoswarm cancer clinical trial** Matimba Swana, University of Bristol, UK • **International AI research: the issue of moral pluralism** Serene Ong, National University of Singapore, Singapore
10:00	*Tea/coffee break*

Theme 1 10:30	**Emerging issues in research methodology** Chair: Daudi Jjingo, Makerere University, Uganda
10 min	**Introduction to the theme**
20 min	**A silent trial is critical to accountable and justice-promoting implementation of artificial intelligence in healthcare** Melissa McCradden, The Hospital for Sick Children, Canada
20 min	**The PSYLECT study: opportunities and pitfalls of digitizing a clinical trial in a LMIC** Andre Brunoni, University of São Paulo Medical School, Brazil
20 min	**Discussion**
45 min	**Breakout group discussion**
12:25	*Lunch*
Theme 2 13:40	**Importance of local context and engagement when developing AI tools** Chair: Jay Shaw, University of Toronto Joint Centre for Bioethics, Canada
10 min	**Introduction to the theme**
15 min	**Ethical issues associated with the development of an ear biometric tool for patient identification in Zambia** Alinani Simukanga, University of Zambia, Zambia
15 min	**Adherence vs agency: AI for behaviour change in health** Niyoshi Shah, Quicksand Design Studio, India
15 min	**Feasibility, acceptance and ethical considerations of a mobile clinical decision support system in Botswana** Kagiso Ndlovu, University of Botswana, Botswana
20 min	**Discussion**
35 min	**Breakout group discussion**
15:30	*Tea/coffee break*

Theme 3 16:00	**Collaborative initiatives and data resources to support AI health research** Chair: Judy Gichoya, Emory University, USA
10 min	**Introduction to the theme**
20 min	**Ethical considerations in implementing the Data Advancing Wellness in Africa (DAWA) Project** Gakii Masunga, Harvard Medical School, USA
20 min	**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
20 min	**Discussion**
35 min	**Breakout group discussion**
17:45	*Meeting close*
18:30	*Meet in the venue reception for departure for dinner*

Wednesday 30 November 2022

08:30	**Summary – key themes from day 1** Sharon Kaur, University of Malaya, Malaysia Anna Chiumento, University of Edinburgh, UK
Theme 4 09:00	**Regulation of data for health research involving AI** Chair: Effy Vayena, Health Ethics and Policy Lab, Institute of Translational Medicine, ETHZ, Switzerland
10 min	**Introduction to the theme**
20 min	**Governance of cross-border transfer of data in Sub-Saharan Africa** Nezerith Cengiz, Stellenbosch University, South Africa
20 min	**Regulation of health data for AI in Uganda** Harriet Nankya, Makerere University, Uganda
20 min	**Discussion**
35 min	**Breakout group discussion**
10:45	*Tea/coffee break*

Theme 5 11:20	**Issues associated with research ethics frameworks and ethics review** Chair: Joseph Ali, Johns Hopkins Berman Institute of Bioethics, USA
10 min	**Introduction to the theme**
15 min	**Recommendations for the development of ethical guidelines for AI-related health research in Egypt** Ahmed Samir Abdelhafiz – National Cancer Institute, Cairo University, Egypt
15 min	**The 'proverbial' black box that is ethics of AI in global health research: are Kenyan RECs well equipped?** Brenda Odero, Strathmore University, Kenya
15 min	**Reframing research ethics frameworks to include environmental sustainability** Gabrielle Samuel, King's College London, UK
20 min	**Discussion**
40 min	**Breakout group discussion**
13:15	*Group photo and lunch*
14:30	**Pecha Kucha** Chair: Paul Ndebele, George Washington University, USA • **A shift to openness: open consent and open science in AI health research in South Africa** Meshandren Naidoo, University of KwaZulu-Natal, South Africa • **A regulatory framework for AI-health research in the Caribbean** Derrick Aarons, The Caribbean Public Health Agency, Jamaica • **How to translate universal principles to local realities: the Chilean experience in AI** Sofia Salas, Clínica Alemana Universidad del Desarrollo, Chile • **Developing a governance framework for data science health research in Nigeria** Oluchi Maduka, Center for Bioethics and Research, Nigeria • **Adaptability of India's Health Data Regulations** Rupanjali Karthink, Duke University, USA
15:05	**Concluding panel discussion** Chair: Katherine Littler, World Health Organization, Switzerland Panel: • Ross Upshur, University of Toronto, Canada • Teck Chuan Voo, National University of Singapore, Singapore • Keymanthri Moodley, Stellenbosch University, South Africa • Armando Guio, Harvard University, USA • Caesar Atuire, University of Ghana and University of Oxford
15:50	**Presentation of awards and announcement about next year's meeting** Katherine Littler, World Health Organization, Switzerland Ana Palmero, Ministry of Health, Argentina
16:00	*Meeting close*

Theme 1: Emerging issues in research methodology

Governance paper: A silent trial is critical to accountable and justice-promoting implementation of artificial intelligence in healthcare

Melissa D McCradden[1,2,3]
[1]The Hospital for Sick Children – Department of Bioethics
[2]Peter Gilgan Centre for Research & Learning – Genetics & Genome Biology Research Program
[3]Dalla Lana School of Public Health

Context
Artificial intelligence (AI) – a subset of machine learning (ML) – refers to a computational system that can run inference on novel cases based on a mapping of inputs (data) to outputs (labels). The ability to predict accurately the current or future state of a patient's health is a valuable property, yet to do so does not guarantee the patient will benefit. In a recent systematic review of clinical AI tools, only two fifths of those with good technical performance also were associated with a concomitant improvement to patient outcomes[1]. These trials are essential for two reasons: 1) to reliably establish knowledge about the causal effect of an AI tool on a given patient outcome; and 2) because AI can introduce novel and unanticipated errors, including automation bias[2] and algorithmic injustice[3]. Ensuring that AI adoption actually benefits patients reflects a commitment to evidence-based practice, responsible health data stewardship, organizational accountability, and efficient use of hospital resources. Additionally, the manner of technological integration and perception of value alignment can contribute positively or negatively to healthcare worker satisfaction with the tool, as we have seen with the introduction of the electronic health record. For all these reasons, the demonstration that an AI tool benefits patients is crucial to ethical integration for both patients and healthcare workers.

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, generalizability of an AI system's performance is a matter of controversy[4]. Attempts to utilize systems like Watson for Oncology have proven to be highly inappropriate when using them internationally. Assuming, rather than testing, that a model's performance will be retained in a new environment is a recipe for failure. Low- and Middle-Income Countries (LMICs) may reasonably wish to adopt AI technologies, some of which may be developed and trialed externally. In some cases, the population distribution will differ substantially from the training population, which poses additional concerns to the fairness properties of the system.

How should international healthcare organizations go about ascertaining that performance will be retained in their local population and promote rather than compromise fairness to patients? This governance paper proposes the silent trial as a critical process to protect patients' interests and serve justice.

<u>The silent trial</u>
In model development, an algorithm is initially developed and validated using historical data from a curated dataset as an initial proof-of-concept that some output (label) can be effectively mapped given particular inputs (data). While a valuable component of responsible AI evaluation[5], such validation is not sufficient to guarantee similar performance in the clinical environment. As such, the silent trial (aka shadow trial or silent mode) 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). The silent trial thus enables two goals: 1) demonstrating the ecological validity of the model and 2)

offers the kind of information that can establish clinical equipoise, which is the justificatory basis on which interventional trials are considered ethically permissible[6].

Silent trial and justice

In the context of LMICs, requiring that a silent trial be conducted onsite prior to committing to adopting a particular AI model can be a valuable process for protecting patients' interests. While compliance with standardized reporting guidelines and transparent reporting of clinical trials can help institutions identify whether a model could have utility for them[7], good models can still fail to generalize to new settings. The silent trial verifies whether a model will perform adequately on the local patient population prior to actually using the model to inform the care of patients. Institutions can thus demonstrable that they are accountable to their patients by verifying the AI's appropriateness in an empirical sense.

As is well-recognized by now, AI systems reflect the biases embedded in societal injustices, spitting out patterns of inequity, inaccessibility, and prejudice when not adequately overseen. Several case examples in the literature and public discourse describe the use of AI that resulted in systemic disadvantages to racialized, marginalized, and/or oppressed groups. These consequences are further problematic in their lack of recourse – affected persons typically have no avenue to dispute the output, and users do not have the tools to identify where the AI went wrong.

As such, increasing attention is being paid to the importance of algorithmic audits as a mechanism for the more robust characterization of algorithmic performance (i.e., with respect to clinical and demographic sub-groups)[7]. As fairness is a core concern of AI, the principle of distributive justice suggests two key steps must be taken: first, establishing the distributive benefits and burdens of a given system, and second, identifying opportunities for correction, revision, or redress. For AI, step 1 would look like an algorithmic audit at the silent trial stage to characterize model performance across the locally relevant population and subgroups of patients. Step 2 would involve reflection and engagement regarding the suitable options to address potential discrepancies in that distribution. Relational ethics highlights the importance of engaging those most affected by AI under-performance or discrimination, respecting their lived experience as valuable knowledge, and collaboratively identifying solutions that serve justice[3].

As previously noted[8], LMICs are not identical in population, character, or context. The silent trial provides additional opportunities to truly integrate (rather than 'deploy') a model by engaging those using the model in its implementation. One can imagine that human factors evaluation and stakeholder engagement will be increasingly important as the world outside of healthcare continues to be marred by ethical controversies involving AI. The silent trial provides a technical foundation to verify performance and an evaluative process whereby stakeholders are actively engaged in shaping AI implementation. The process thereby operationalizes relational ethics values that emphasize human interconnectedness and moral obligations to others[9] – a much-needed juxtaposition to the algorithmic coldness of AI.

Case example

Our institution is currently trialing a classification model to identify obstructive hydronephrosis in infants[10]. Despite the initially strong model performance (AUROC 90%) when moved into the silent trial stage we observed a decrease in performance (to an AUROC of 50%). Upon investigation, it was revealed that the patients in the silent trial were significantly younger and more likely to have right-sided obstruction of the kidney. Controlling for these differences improved the model performance. The team next addressed differences in image processing to improve the model performance to an AUROC of 85%. The model was additionally evaluated for performance across sex, laterality of hydronephrosis, ultrasound machine, and the patient's home postal code. The team found a retained performance of over 90% sensitivity across all variables. By conducting a silent trial, the team was able to establish the generalizability of the model to the live clinical setting and assure themselves of its performance across relevant patient subgroups. Notably, we were unable to assess race or ethnicity as such data is not yet routinely collected in Canada; despite the use of postal code as a proxy, we acknowledge the limitations of this work.

The need for the silent trial is evident when considering the known limitations of model generalizability. For example, Straw and Wu[11] assessed the performance of four classifier models that predict the presence of liver disease from a commonly used dataset. They observed a higher false negative among women generally across all studies, indicating that more women than men could have a missed diagnosis using the model. A caveat, however, is that although they addressed the issue of class imbalance, there is limited information into the effects of potential confounding variables (a noted limitation in many ML-based works exploring bias[12]).

An additional opportunity embedded in the silent trial paradigm is the chance to seek user feedback and engagement around model integration choices[10,13]. This can be done through quantitative or qualitative activities. When faced with model performance discrepancies, there is a need to make choices about how the effects will be mitigated. The silent trial provides the empirical characterization needed to ground discussions about the most appropriate actions to take for model integration to promote equity, transparency, and ethical decision-making.

Conclusion and recommendation

This paper advocates for widespread adoption of a silent trial prior to the integration of any AI tool. This step enables the operationalization of distributive justice and provides a reliable empirical foundation for ensuring AI tools will benefit patients across different contexts. Algorithmic auditing can be a key piece of the silent trial to identify failure modes, disproportionate error rates and other performance metrics, inform postdeployment safety monitoring, identify the need for postdeployment recalibration, and mitigate risks to groups[8]. Audits thereby are a way to operationalize distributive justice as well as healthcare's commitments to evidence-based integration and good clinical decision-making. A common goal across all contexts is the use of AI to alleviate health burdens and augment the humanistic aspect of medicine. Adopting processes that enable accountable decisions and provide an empirical foundation for informed decisions on fairness and justice is one step toward this end.

References

1. Zhou, Q., Chen, Z.H., Cao, Y.H. and Peng, S., 2021. Clinical impact and quality of randomized controlled trials involving interventions evaluating artificial intelligence prediction tools: a systematic review. *NPJ digital medicine*, *4*(1), pp.1-12.
2. Tschandl, P., C. Rinner, Z. Apalla, G. Argenziano, N. Codella, A. Halpern, M. Janda, A. Lallas, C. Longo, J. Malvehy, et al. 2020. Human-computer collaboration for skin cancer recognition. *Nature Medicine* 26(8): 1229–1234.
3. Birhane, A., 2021. Algorithmic injustice: a relational ethics approach. *Patterns*, *2*(2), p.100205.
4. Futoma, J., Simons, M., Panch, T., Doshi-Velez, F., & Celi, L. A. (2020). The myth of generalisability in clinical research and machine learning in health care. *The Lancet Digital Health*, *2*(9), e489-e492.
5. Wiens, J., S. Saria, M. Sendak, M. Ghassemi, V. X. Liu, F. Doshi-Velez, K. Jung, K. Heller, D. Kale, M. Saeed, et al. 2019. Do no harm: A roadmap for responsible machine learning for health care. Nature Medicine 25(9):1337–40.
6. McCradden, M.D., Anderson, J.A., A. Stephenson, E., Drysdale, E., Erdman, L., Goldenberg, A. and Zlotnik Shaul, R., 2022. A research ethics framework for the clinical translation of healthcare machine learning. *The American Journal of Bioethics*, *22*(5), pp.8-22.
7. Liu, X., Glocker, B., McCradden, M.M., Ghassemi, M., Denniston, A.K. and Oakden-Rayner, L., 2022. The medical algorithmic audit. *The Lancet Digital Health*.
8. McCradden, M.D. and Chad, L., 2021. Screening for facial differences worldwide: equity and ethics. *The Lancet Digital Health*, *3*(10), pp.e615-e616.
9. Ewuoso, C., 2021. An African Relational Approach to Healthcare and Big Data Challenges. *Science and Engineering Ethics*, *27*(3), pp.1-18.
10. Kwong, J. C., Erdman, L., Khondker, A., Skreta, M., Goldenberg, A., McCradden, M. D., ... & Rickard, M. (2022). The silent trial-the bridge between bench-to-bedside clinical AI applications. *Frontiers in digital health*, 163.
11. Straw, I., & Wu, H. (2022). Investigating for bias in healthcare algorithms: a sex-stratified analysis of supervised machine learning models in liver disease prediction. *BMJ health & care informatics*, *29*(1).
12. Bernhardt, M., Jones, C., & Glocker, B. (2022). Potential sources of dataset bias complicate investigation of underdiagnosis by machine learning algorithms. *Nature Medicine*, *28*(6), 1157-1158.
13. Tonekaboni, S., Morgenshtern, G., Assadi, A., Pokhrel, A., Huang, X., Jayarajan, A., ... & Goldenberg, A. (2022, April). How to validate Machine Learning Models Prior to Deployment: Silent trial protocol for evaluation of real-time models at ICU. In *Conference on Health, Inference, and Learning* (pp. 169-182). PMLR.

Case study: The PSYLECT study: opportunities and pitfalls of digitizing a clinical trial in a LMIC

Andre R Brunoni, MD, PhD, Associate Professor, University of São Paulo Medical School and Institute of Psychiatry - Hospital das Clínicas – FMUSP, São Paulo, Brazil

Brief description of the research project

The *Portable Transcranial Electrical Stimulation and Internet-based Behavioral Therapy for Major Depression Study* (PSYLECT) is a randomized, sham-controlled clinical trial with digital features that recruited 210 patients with unipolar depression. These patients will be randomized to one of three groups: sham, home-use transcranial electric stimulation (tES) and digital placebo intervention, active, home-use tES and digital placebo, and active, home-use tES and app-based behavior therapy. These patients will be followed up for 6 weeks, and, at the endpoint, we will compare which group achieved superior efficacy. Both interventions are designed to be performed at home and self-delivered, and were developed together with Flow Neurosciences™. Previously to the trial, we translated the iBT questionnaires to Portuguese and validated it in our local population. The trial started in May 2021 and finished in October 2022.

Background

Major depressive disorder (MDD) remains a leading cause of disability-adjusted life years, despite traditional pharmacological and psychotherapeutic options[1]. MDD affects more than 300 million people worldwide, with a chronic and recurrent course[2], particularly in LMIC. First-line treatments for MDD present significant caveats, as antidepressant medications are associated with modest efficacy[3] and adverse effects[4], while in-person cognitive-behavioral therapy lacks wide-range availability, and involves higher costs and logistical burdens[5]. Transcranial electrical stimulation (tES), is a non-invasive brain stimulation technique with excellent acceptability and moderate effectiveness for MDD. Thus, it could be a first-line intervention, especially in patients with low-drug resistance[6,7]. However, such an approach is hampered by the limited scalability of tES treatment. The relative scarcity of skilled personnel and the logistical burdens and transportation costs associated with daily visits to external facilities are probably associated with its suboptimal utilization in clinical practice. In this context, recent technological advancements are progressively allowing tES to be performed remotely, operated by patients themselves, therefore reducing costs and enhancing scalability. This could represent important gains in vulnerable groups with depression, such as in LMIC. Concomitantly, growing attention has also been directed towards the combination of tES and neurobehavioral or psychotherapeutic interventions[8,9]. They can also be delivered remotely, in an internet-based and self-directed manner, especially using interactive smartphone apps[10]. Meta-analyses that evaluated the effect of app-based interventions in MDD found superiority of these interventions over control conditions, with small to large effect sizes[10–12], and higher retention rates when there was human feedback and mood assessments through the apps[13]. Moreover, the recent research interest in mental health apps for the treatment of MDD is occurring within a larger framework encompassing the rapid development of digital mental health technologies, in great part, boosted by the social distancing restrictions imposed by the COVID-19 pandemic[14,15]. Therefore, the expansion of digital mental health interventions and their good usability enables better access to healthcare, cost reduction, personalized approaches, and adherence to treatment. While a few studies evaluating the combination of tES with psychotherapy have been performed in research facilities, to the best of our knowledge, no controlled trial has investigated the synchronous combination of portable transcranial electrical stimulation (ptES) and a remotely delivered, self-directed and internet-based behavioral intervention (iBT), for the treatment of MDD, in adult patients.

Digital clinical trial

Considering that the intervention (ptES + iBT) can be deployed fully at home, we designed a digital trial[16] to enhance its scalability. Therefore, we use social media and digital marketing strategies to advertise our trial and screen potential participants by telemedicine. All evaluations of the study, except the first and the last ones, are performed virtually. During the initial visit, participants are instructed on how to operate the device and install the accompanying app in their smartphones. Regarding data collection and protection, our staff was trained on how to use RedCap and has

taken classes on data protection law (the GDPR-Brazilian equivalent, LGPD). Communication with patients is centered on WhatsApp, a very popular app in Brazil, and randomization, allocation and blinding through the development of applications that carry out the intervention according to an encrypted code. Compared to previous studies performed by our group[17–19], we observed (as 60% of the sample has already been recruited) that the external validity / generalizability of these participants is higher compared to previous ones. Since tES sessions needed to be performed on-site (at the clinical center) before, the participation in our previous studies was restricted to those who had flexible working hours or lived near the clinical center. Additionally, the adherence to the study has been very high, as fewer than 5% of participants have dropped out so far.

Ethical issues

1. Selection process

We found that recruitment rates increased five times compared to our previous studies, due to online advertising, stimulating us to further increase trial scalability. Here, we found that a critical choke point is the screening process. Only 30% of screened patients meet eligibility criteria and can be invited to participate in the study. Moreover, during the initial interview, which is done on-site, about 30% of them are excluded as the online interview misses important information. Therefore, we considered options to enhance screening accuracy. In this context, we considered collecting more information by questionnaires that volunteers answer when subscribing to the study to train an algorithm for this purpose. However, here there is the ethical challenge of using screening data prior to the informed consent, which is only offered to the participant at a later step of the study. Ultimately, a new study, recruiting participants to collect data to allow future participants to be excluded in further trials, would have to be conducted. Nonetheless, the collection of new data that could enhance the algorithm would still not be allowed. Therefore, one possible solution is to collect the consent form as soon as the participant volunteers to participate in the trial. Additionally, usually more severe patients are not included in a study. However, more severe patients are also those in a more vulnerable socioeconomic position. An algorithm that is trained aiming to enroll less severe patients could use such information to enhance its accuracy, essentially digitizing exclusion. The same issue could occur whether the algorithm is trained to enroll participants that are less likely to drop out of the study. This could be related to people who have less digital literacy, which are usually people who are less educated or have a lower socioeconomic status. It is important to highlight that the selection process which eliminates people who are unlikely to be successful in the use of the tool means that the success rate of the trial will ultimately be higher than what will be obtained if administered outside the trial setting.

2. Issues arising during the trial

Another opportunity that also brings ethical challenges is using bots ("chatbots") for interacting with participants virtually. On one hand, this would improve scalability, considering that most questions from participants can be answered using decision trees (e.g., concerns regarding missing a session, or simple troubleshooting). On the other hand, participants might feel alienated if they do not have prompt access to the research team, which could in turn decrease adherence. In addition, ancillary care obligations that could require greater attention (perhaps human) to issues that participants may be facing might not be easily capturable by a chat box.

3. Ethical issues that may arise if successful and the tool is to become available population wide

Also, performing a digital trial brings the opportunity of collecting active and passive data using apps and wearables. Although using such an approach provides data granularity, this also brings ethical issues regarding the extent and how the data is being used. Participants and patients using apps that collect sensitive and personal data should be clearly informed regarding the extent and which type of data are being collected, and should be provided several options regarding privacy access.

Conclusion and recommendations

This is the first digital trial using portable tES devices combined with an app-based behavioral therapy system in a LMIC. We noticed ethical challenges in terms of recruiting participants to our study and using the data they provided to develop algorithms that could lead to their trial exclusion. To address this issue, we amended our informed consent term to guarantee that, even at the initial stage, participants would consent that the data they provide could be used for the specific process of screening. A second aspect is that advertising our study only in social media could exclude people with low digital literacy. Therefore, we are also using traditional media (newspapers and radio) and scheduling on-site screening visits to those who are not comfortable with the digital process of screening.

References

1. Collaborators GBD. … national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis …. 2018; Available from: https://leicester.figshare.com/articles/journal_contribution/Global_regional_and_national_incidence_prevalence_and_years_lived_with_disability_for_354_diseases_and_injuries_for_195_countries_and_territories_1990-2017_a_systematic_analysis_for_the_Global_Burden_of_Disease_Study_2017_/10200149/files/18383459.pdf
2. Nemeroff CB. Prevalence and management of treatment-resistant depression. *J Clin Psychiatry* 2007;68 Suppl 8:17–25.
3. Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry* 2006;163(11):1905–17.
4. Anderson IM, Ferrier IN, Baldwin RC, et al. Evidence-based guidelines for treating depressive disorders with antidepressants: a revision of the 2000 British Association for Psychopharmacology guidelines. *J Psychopharmacol* 2008;22(4):343–96.
5. Keller MB, McCullough JP, Klein DN, et al. A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression. *N Engl J Med* 2000;342(20):1462–70.
6. Razza LB, Palumbo P, Moffa AH, et al. A systematic review and meta-analysis on the effects of transcranial direct current stimulation in depressive episodes. *Depress Anxiety* [Internet] 2020; Available from: http://dx.doi.org/10.1002/da.23004
7. Fregni F, El-Hagrassy MM, Pacheco-Barrios K, et al. Evidence-based guidelines and secondary meta-analysis for the use of transcranial direct current stimulation (tDCS) in neurological and psychiatric disorders. *Int J Neuropsychopharmacol* [Internet] 2020; Available from: http://dx.doi.org/10.1093/ijnp/pyaa051
8. Sathappan AV, Luber BM, Lisanby SH. The Dynamic Duo: Combining noninvasive brain stimulation with cognitive interventions. *Prog Neuropsychopharmacol Biol Psychiatry* 2019;89:347–60.
9. Dedoncker J, Baeken C, De Raedt R, Vanderhasselt M-A. Combined transcranial direct current stimulation and psychological interventions: State of the art and promising perspectives for clinical psychology. *Biol Psychol* 2021;158:107991.
10. Li J, Theng Y-L, Foo S. Game-based digital interventions for depression therapy: a systematic review and meta-analysis. *Cyberpsychol Behav Soc Netw* 2014;17(8):519–27.
11. Josephine K, Josefine L, Philipp D, David E, Harald B. Internet- and mobile-based depression interventions for people with diagnosed depression: A systematic review and meta-analysis. *J Affect Disord* 2017;223:28–40.
12. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. *World Psychiatry* 2019;18(3):325–36.
13. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: A systematic review and meta-analysis. *J Affect Disord* 2020;263:413–9.
14. Psychiatry TL, The Lancet Psychiatry. Digital psychiatry: moving past potential [Internet]. *The Lancet Psychiatry*. 2021;8(4):259. Available from: http://dx.doi.org/10.1016/s2215-0366(21)00096-1

15. Torous J, Bucci S, Bell IH, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry* 2021;20(3):318–35.
16. Inan OT, Tenaerts P, Prindiville SA, et al. Digitizing clinical trials. *NPJ Digit Med* 2020;3:101.
17. Brunoni AR, Moffa AH, Sampaio-Junior B, et al. Trial of Electrical Direct-Current Therapy versus Escitalopram for Depression. *N Engl J Med* 2017;376(26):2523–33.
18. Sampaio-Junior B, Tortella G, Borrione L, et al. Efficacy and Safety of Transcranial Direct Current Stimulation as an Add-on Treatment for Bipolar Depression: A Randomized Clinical Trial. *JAMA Psychiatry* 2018;75(2):158–66.
19. Valiengo L da CL, Goerigk S, Gordon PC, et al. Efficacy and Safety of Transcranial Direct Current Stimulation for Treating Negative Symptoms in Schizophrenia: A Randomized Clinical Trial. *JAMA Psychiatry* [Internet] 2019; Available from: http://dx.doi.org/10.1001/jamapsychiatry.2019.3199

Theme 2: Importance of local context and engagement when developing AI tools

Case study: Ethical issues associated with the development of an ear biometric tool for patient identification in Zambia

Alinani Simukanga[1] (presenter), Lauren Etter[2], Wenda Qin[3], Rachel Pieciak[2], Margrit Betke[3], William Macleod[2], Jackson Phiri[1], Christopher J. Gill[2]

1 Department of Computer Science, School of Natural Sciences, University of Zambia, Zambia
2 Department of Global Health, Boston University, Boston, MA, USA
3 Department of Computer Science, Boston University, Boston, MA, USA

Brief description of the research project
In 2017, with funding from the Fogarty International Center at NIH, we launched Project SEARCH (Scanning EARs for Child Health). Our team has focused on uniquely solving the challenge of identifying patients – through biometric analysis of ears. Ears have significant advantages over other biometric targets: facial scans raise privacy concerns; iris scanners are expensive and can frighten small children, and fingerprint scans perform poorly in children <5 years. We created a powerful mHealth identification App (the SEARCH App) that runs on the ubiquitous Android operating system[1,2]. In field studies, the SEARCH App achieved near-perfect subject identification among Zambian infants as young as six months old, a significant advance over fingerprint technology[3,4]. Building on this foundation, the project intends 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).

Background
In 2005, the President's Emergency Plan for AIDS Relief (PEPFAR) funded a collaboration between Zambia's Ministry of Health (MOH) and the CDC to deploy an electronic management system to coordinate the delivery of HIV care. Since its initiation SmartCare has been scaled up and 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. Physically, the CareCard resembles an ATM bank card with a magnetic strip. In pediatric HIV care, these cards are issued to the child's guardians.

Relying on the CareCard for identification has proven a critical limitation of SmartCare. As with paper Under-Five Carecards (UFCs), the cards are easily lost, damaged, or used inadvertently by another individual (the cards do not display a name or photograph, just the assigned ART number). Additionally, they suffer from a high rate of technical error when damaged cards cannot be scanned, when the internet goes down, or when the software is incompatible between card and reader – all of which occur frequently. If the patient has lost or forgotten their CareCard or is using someone else's card, or when there is a technical failure, demographic information (such as names and dates of birth) is used in an often-fruitless attempt to find the patient in the SmartCare database. Each failure requires that a new identity be created for that individual and a new CareCard issued. With a 30-40% combined failure rate, patients quickly accumulate handfuls of these cards, spreading their medical history across multiple unlinked aliases in SmartCare's database. Given these limitations, some clinics have reverted to using paper registries to duplicate the electronic, undermining the motivating rationale for an EMR. The work of disambiguating double documentation for paper and electronic forms creates backlogs going back months. Even when eventually entered electronically, the data are so late and incomplete as to be practically useless for programmatic monitoring. And the problem is not unique to Zambia: in South Africa, our team encountered similar challenges when trying to track ART retention due to the proliferation of aliases in the national EMR[5].

Ethical Issues

Issue #1: Fairness and equity

Bias in data is a challenge that presented itself early on in the project. The team was initially limited to using datasets collected at the Museum of Science in Boston to finetune the biometric tool. Publicly accessible datasets of darker-coloured ears captured in controlled conditions could not be found at the time. Initial tests conducted with the tool on a dataset in Lusaka (Zambia) showed a drop-off in performance. It was clear at this time that a major limitation to future work would come from the fact that the training datasets in use at the time were not representative of the intended use-case population. This necessitated a data collection exercise primarily focused on creating a dataset of darker-coloured ears captured from young Zambian infants. Data from 224 infants were captured while attending vaccination visits at Chawama First Level Hospital in Lusaka from November 2019 to April 2020. Images were taken by one data collector, who was thoroughly trained in the use of the data collection tool. Two images were taken of both the left and right ear were taken at each visit.

Written consent forms (approved by both the Boston University IRB and the University of Zambia Biomedical Research Ethics Committee) translated into two local languages, were provided to all the study participants. The intended use of the data was laid out in the consent form. Participants who provided consent and intended to attend well-child visits at the facility in the future were included in the study.

The various tests conducted on the datasets early on in the project allowed us to identify data bias and the resulting algorithmic harm early in the development process. The main issue identified was that we would have ended up developing a tool that would perform poorly in its targeted setting. Our images were captured with the assistance of a 3D-printed opaque plastic cylinder we call the 'Donut'. The Donut is mounted to a phone and allows for the standardisation of conditions during image capture - angle, distance to ear and lighting. The contrast between the darker ears and the light-coloured material of the Donut would lead to the camera's auto-exposure feature underexposing (darkening) the ear to avoid having too bright of a background. This was not a problem with the earlier datasets that had a majority of light-coloured ears. At that time, our tool could not extract enough detail from the images as some of the ear features could not be identified in the overexposed images. We went on to employ a few post-image capture techniques that would assist our tool in feature extraction in later versions of the tool. It is indeed possible that while relying on datasets collected in Boston, we could have stumbled on some version of this problem. The performance loss we saw on initial tests with the locally collected dataset meant that we tackled this problem earlier in the project.

Issue #2: Transparency and engagement

A series of focus group discussions and interviews were conducted with mothers and health officials in Zambia's rural and urban settings. The University of Zambia Biomedical Research Ethics Committee granted ethical approval for this activity. The key focus of this activity was to engage the stated stakeholders and gain an understanding of how receptive the community would be to the app being developed.

Participants were recruited from three health facilities in Lusaka and Southern provinces. Focus group participants were recruited using a set of inclusion criteria: 1) they were mothers younger than 45 years of age, 2) had one or more children, and 3) had experience using the UFC (paper-based under-five card). Participants were approached during their clinic visits, had the study described to them, and underwent a consent discussion. In total, focus group discussions were conducted with 59 mothers across the three health facilities. Participants were recruited until budgetary and timing constraints didn't allow for further recruitment. Participants received information on the study in their preferred language.

In-depth interviews were conducted with the nurses in charge and clinicians at each of the three facilities as well as the District Health Director (DHD) or Information Officers (IO) at the district level. We felt that this group would help us answer these questions:

- What are their perceptions on the use of biometrics as a tool for child identification in place of a child health card?
- Would a biometric identification system generate value for healthcare workers?
- What concerns could they anticipate about the community's acceptance of a biometric system?
- In what way would end-users (clinicians, parents and children) react to a biometric system?

Potential clinic staff participants had the purpose of the study described to them and underwent a consent discussion.

Partner acceptance was an issue that some mothers expressed concern over. Mothers highlighted their partner's aversion to western world technologies, fear of malicious intent, and general unwillingness to support change.

Being aware that we were inserting ourselves into an environment with a tool that the key stakeholders were encountering for the first time meant that community engagement activities had to be conducted. With that in mind, these activities were included in the grant application that was submitted to the funding agency. It might have seemed a bit premature to be conducting this level of engagement since we were not yet at the implementation stage, but we thought it was important to get the views of the likely end users and the mothers earlier in the process. An EHR system has been used in public health facilities across the country for at least the past decade, but we felt that there would be a distrust of our tool which was more people-facing than an EHR system running in the background that patients never have to interact with. The main takeaway from these activities was that further engagement was needed to get over the various sociocultural barriers that stood in the way of an eventual rollout[6].

Conclusions

Adequate community sensitisation will be key in tackling the sociocultural issues that cause hesitance toward proposed digital solutions. The invasive nature of capturing biometric data goes further to heighten fears around technology. We generally feel that the task of community engagement must be shared by both researchers and the Ministry of Health. The Ministry has an important role in dispelling any fears that the populace might have towards AI tools and new technologies as a whole. If this is communicated to clinicians and the patient populace, an eventual rollout would occur with minimal issues. Researchers still have a very important role to play in clearly laying out how the tool will benefit the end users.

References

1. Ragan, E.J., Johnson, C., Milton, J.N. *et al.* Ear biometrics for patient identification in global health: a cross-sectional study to test the feasibility of a simplified algorithm. *BMC Res Notes* **9**, 484 (2016). (https://doi.org/10.1186/s13104-016-2287-9)
2. Etter, L.P., Ragan, E.J., Campion, R. *et al.* Ear biometrics for patient identification in global health: a field study to test the effectiveness of an image stabilization device in improving identification accuracy. *BMC Med Inform Decis Mak* **19**, 114 (2019). (https://doi.org/10.1186/s12911-019-0833-9)
3. Etter L, Simukanga A, Qin W *et al.* Project SEARCH (Scanning EARs for Child Health): validating an ear biometric tool for patient identification in Zambia [version 1; peer review: 2 approved]. *Gates Open Res* 2020, **4**:168 (https://doi.org/10.12688/gatesopenres.13197.1)
4. Simukanga A, Kobayashi M, Etter L *et al.* The impact of ear growth on identification rates using an ear biometric system in young infants [version 1; peer review: 1 approved with reservations]. *Gates Open Res* 2021, **5**:179 (https://doi.org/10.12688/gatesopenres.13459.1)

5. Fox MP, Bor J, Brennan AT, MacLeod WB, Maskew M, et al. (2018) Estimating retention in HIV care accounting for patient transfers: A national laboratory cohort study in South Africa. PLOS Medicine 15(6): e1002589. (https://doi.org/10.1371/journal.pmed.1002589)
6. Lissin P, Hamapa A, Kobayashi M *et al.* Relative advantages and compatibility of a biometric patient identification tool in Zambia: a qualitative analysis [version 1; peer review: 1 not approved]. *Gates Open Res* 2021, **5**:102 (https://doi.org/10.12688/gatesopenres.13265.1)

Case study: Adherence vs agency: AI for behaviour change in health

Niyoshi Shah - Consultant, Quicksand Design Studio, India

Background

International evidence suggests that health workers are not as diligent about handwashing as they should be given how crucial it is in preventing hospital-acquired infection (HAI)[1]. The global average compliance rate for handwashing is only about 38.7% with some countries falling below an acceptable threshold[2]. Direct supervision and feedback are usually used to improve staff habits but these methods are resource-intensive, prone to bias, and the Hawthorne effect[3-6]. Some studies suggest that new technologies like artificial intelligence (AI) can be used instead[7-10] but there is a dearth of information on their implementation in low- and middle-income countries (LMICs). Through this project we therefore employed human-centred design (HCD) to explore how AI may be deployed for the quality improvement of hand hygiene in India's public health system. The team consisted of partners and advisors from the social and commercial sectors of which Jhpiego (for ground support),[11] Quicksand (for HCD), and Datakalp (for technology) were main.

Brief description of the project

An AI system called *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[12]. 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 the World Health Organization (WHO). This was supported by a display monitor where the staff got live feedback for each episode. (A green tick for the steps they got right, a red cross for the ones they didn't.) (2) All the data captured at the basin was aggregated day-wise and made available to the management on a dashboard for better monitoring and evaluation. We applied the HCD process to understand the hospital's experience of both these components.

HCD can be described as a creative process of problem solving with co-creation as its cornerstone. It begins with ethnography where designers learn about the needs, preferences, social context and constraints of their end users. This research is translated into rough solutions that are tested on the ground early on and often to come up with a final product, service, system or strategy that is truly people-centred. In recent years, this hands-on approach has gained a special significance in the development sector for it allows projects to fail and learn at a smaller scale before growing into a clinical trial. HCD can therefore be applied as 'an ingredient' to a larger study or from 'end to end' as our project demonstrates[12].

Main activities of the project

I. **Foundational Research:** We conducted field visits, in-depth interviews, and focussed group discussions at all the project sites with 30+ stakeholders across all staff categories to learn about their handwashing practices and their initial experience of the AI system.

II. **Ideation & Prototyping:** We used the Manoff Group's *Toolkit for Behaviour Integration* to isolate the factors that influenced the uptake of Vajrahands from other research findings. These behavioural levers became the linchpin of our brainstorming sessions, where the most promising ideas were refined through multiple cycles of feedback from our end users until we arrived at a final set of interventions to make the AI system more friendly and useful in improving hand hygiene.

III. **Delivery:** In the last phase of the project, we introduced a new user interface for the display monitor and summary reports for the dashboard, as described in the next section. These product changes (collectively called Vajrahands 2.0) were supported by a bundle of non-digital interventions to encourage different forms of motivation and accountability (like staff meetings, self-driven targets, rewards, and feedback on hygiene from patients) along with team cohesion.

IV. **End Assessment:** The acceptability and the impact of the new interface, the summary reports, and the supporting interventions were evaluated in the following ways -

- Usage and performance data from Vajrahands (All sites)
- Staff surveys (95 respondents, all sites)
- Management surveys (21 respondents, all sites)
- In-depth interviews and group discussions with the staff, management, and ground team (30 respondents, all sites)

Ethical issues with a brief commentary on each issue

A. Vajrahands was programmed according to the handwashing technique recommended by the WHO to align with global standards but this was not well-received by the staff because they were used to another protocol called SUMAN-K. They had to un-learn it and perform each step of the WHO sequence in a very specific way to be marked right by the algorithm. In addition to that they found the display monitor confusing. Its graphics were too small, and the red crosses for the missed or incorrect steps only appeared at the end of a hand wash so people would inadvertently linger on the same movement wondering why the AI had not acknowledged it. We worked on these challenges amongst others to improve the user experience and make it more engaging. Our new version almost worked like a video tutorial[13]. The staff simply had to follow a series of GIFs on the handwashing steps by the timer to get a perfect score. This, coupled with the supporting interventions, greatly improved the practice of handwashing — the average compliance rate at our weakest site jumped from 2.7% to 20.1% — but in making the interface more directive, we felt we had further reduced the room for variations that people naturally practise when WHO's recommendations are just that: they are not meant to be rigidly enforced.

> How might we tread the line between adherence and agency as AI is increasingly deployed for behaviour change?

B. Through our foundational research, we learned that the hospitals wanted an easier alternative to the dashboard. We therefore created PDF reports with data visualisation to give them a high-level summary of the ward's performance every fortnight. The reports were shared with the management on Whatsapp with a nudge to circulate them further amongst the staff. In the end evaluation, most respondents felt this intervention was crucial to behaviour change: for the first time, they could quantify and track their adherence to the handwashing protocol. But it also led to top-down supervision, where the higher-ups at some sites used fear to motivate their staff. They told the non-medical workers (who have the lowest status in the staff hierarchy) that the ward's performance was being watched by the government, leaving them anxious to comply with the algorithm. Some of the staff members were scared of making a mistake at the basin.

> How might we protect public health employees against soft coercion as computer visioning and AI are increasingly used to monitor their adherence at work?

C. The staff often de-prioritized the handwashing protocol because there is no incentive to follow it. We therefore introduced a short competition cycle with rewards but some sites had a blinkered view of it: they were bent on getting their scores right instead of using the intervention to inspire learning. They asked their senior staff to use Vajrahands more often to balance out the day's compliance rate if it dropped. We even had instances where the access to the project basin was altogether curtailed for those who accompanied the expectant mother to the labour room because they did not know the WHO sequence, which leads us to ask:

> How might we encourage a more honest relationship with numbers and data as AI is increasingly used in public health?

Conclusions and recommendations

The data generated by Vajrahands was analysed using an interrupted time series segmented regression analysis across a total period of seven months (Dec 2021 - June 2022) that was divided into two phases for (1) the original AI system, and (2) Vajrahands 2.0 + supporting interventions. This analysis confirmed that AI can be used to improve micro-behaviours like handwashing that are imperative to infection control and prevention. The average compliance rate at our best performing site touched 50.1% after the second iteration of Vajrahands was introduced with its retinue of non-digital interventions. But we need to pay more attention to the challenges at the adoption phase. There is a soft link between the ethical issues outlined above if we take a bird's eye view.

The Indian health system is highly hierarchical where development projects such as ours may be sanctioned by the state without consulting the implementation sites in a meaningful way. This leads to lower buy-in, and severe teething issues where the technology in question may not click with its end users [Issue A]. Which in turn creates a fertile ground for soft coercion [Issue B] and number play [Issue C] as the middle rung of decisionmakers (between the state and staff) feel the pressure to meet programme outcomes.

1. A social understanding of AI is therefore crucial to ensure that it does not exacerbate the structural disadvantages against the subaltern. HCD has the skills and mindset to closely work with people and tip the power relations by including a wide range of stakeholders from the margins to the mainstream in its participatory approach.
2. Regulatory frameworks like the Department of Health & Human Services, USA also need to account for this political reality. For example, our project was granted a non-research determination and exempted from ethical oversight because we were (i) collecting non-identifiable information from the ground (ii) investigating a trend of public importance (iii) under the aegis of the state when its decisions are not always consensual[14].
3. There is also an urgent need to develop checks and balances against tech solutionism from within. This can be done by creating an evaluation tool for policymakers to measure the need for AI in their jurisdiction from a rights perspective, in a democratic way.

References

1. Summary of WHO Guidelines on Hand Hygiene in Health Care (2009). Retrieved from https://www.who.int/publications/i/item/9789241597906
2. Ibid.
3. Livorsi DJ, Goedken CC, Sauder M, Vander Weg MW, Perencevich EN, Reisinger HS (2018). Evaluation of Barriers to Audit-and-Feedback Programs That Used Direct Observation of Hand Hygiene Compliance: A Qualitative Study. *JAMA Netw open*. 2018;1(6):e183344. doi:10.1001/jamanetworkopen.2018.3344
4. Fuller C, Michie S, Savage J, et al (2012). The Feedback Intervention Trial (FIT) — Improving Hand-Hygiene Compliance in UK Healthcare Workers: A Stepped Wedge Cluster Randomised Controlled Trial. Cowling BJ, ed. *PLoS One*. 2012;7(10):e41617. doi:10.1371/journal.pone.0041617
5. Gould DJ, Creedon S, Jeanes A, Drey NS, Chudleigh J, Moralejo D (2017). Impact of Observing Hand Hygiene In Practice and Research: A Methodological Reconsideration. *J Hosp Infect*. 2017;95(2):169-174. doi:10.1016/j.jhin.2016.08.008
6. Hagel S, Reischke J, Kesselmeier M, et al (2015). Quantifying the Hawthorne Effect in Hand Hygiene Compliance Through Comparing Direct Observation with Automated Hand Hygiene Monitoring. *Infect Control Hosp Epidemiol*. 2015;36(8):957-962. doi:10.1017/ice.2015.93
7. Dyson J, Madeo M (2017). Investigating The Use of an Electronic Hand Hygiene Monitoring and Prompt Device: Influence And Acceptability. *J Infect Prev*. 2017;18(6):278-287. doi:10.1177/1757177417714045

8. Fisher DA, Seetoh T, May-Lin HO, et al (2013). Automated Measures of Hand Hygiene Compliance among Healthcare Workers Using Ultrasound: Validation and a Randomized Controlled Trial. *Infect Control Hosp Epidemiol*. 2013;34(9):919-928. doi:10.1086/671738
9. Armellino D, Hussain E, Schilling ME, et al (2012). Using High-Technology to Enforce Low-Technology Safety Measures: The Use of Third-Party Remote Video Auditing and Real-Time Feedback in Healthcare. *Clin Infect Dis*. 2012;54(1):1-7. doi:10.1093/cid/cir773
10. Cheng VCC, Tai JWM, Ho SKY, et al (2011). Introduction of an Electronic Monitoring System for Monitoring Compliance with Moments 1 and 4 of the WHO "My 5 Moments For Hand Hygiene" Methodology. *BMC Infect Dis*. 2011;11. doi:10.1186/1471-2334-11-151
11. Jhpiego is a nonprofit associated with Johns Hopkins University. It was initially called the Johns Hopkins Program for International Education in Gynecology and Obstetrics, but is now simply referred to as Jhpiego.
12. We also selected labour wards because this project is part of the MOMENTUM Country and Global Leadership (MCGL) programme that works to reduce maternal, newborn, and child mortality and morbidity in high burden countries.
13. Johnson T, Das S, Tyler N (2021). Design for Health: Human-centred Design Looks to the Future. *Global Health: Science and Practice.* 2021;9(2). doi:10.9745/GHSP-D-21-00608
14. Datakalp LLP. (2022, Feb 12) Vajrahands - AI For Smart Hand Washes, Because Clean Hands Mean Safe Hands [Video]. YouTube. https://www.youtube.com/watch?v=rvlgk3ZIIks
15. We were granted a "non-research determination" by the John Hopkins Bloomberg School of Public Health, which follows the federal regulations laid out by the Department of Health & Human Services, USA.

This project is made possible by the generous support of the American people through the U.S. Agency for International Development (USAID) under the terms of the Cooperative Agreement #7200AA20CA00002, led by Jhpiego and partners. The contents are the responsibility of MOMENTUM Country and Global Leadership and do not necessarily reflect the views of USAID or the United States Government.

Case study: Feasibility, acceptance and ethical considerations of a mobile clinical decision support system in Botswana

Kagiso Ndlovu, University of Botswana, Botswana
Collaboration partner institution**:** VisualDx, USA

Brief description of the research project
Globally, the healthcare industry is undergoing transformation to mitigate the rising costs of healthcare provision and shortages of medical experts. The industry is looking to implement new technology solutions and processes. In Botswana, these efforts are coordinated by the Ministry of Health (MOH) with guidance from the National eHealth Strategy (2020-2024). The MOH acknowledges shortage of health human resources, most significantly in primary healthcare[1]. Moreover, healthcare workers in remote areas have limited training and insufficient reference materials to support diagnosis and management of diseases in dermatology and other subspecialties. Over the years, this has resulted in unnecessary patient referrals and increased burden on the few dermatologists in the country. In order to contribute to addressing this challenge, the University of Botswana collaborated on a research study with an international private organization (VisualDx), to assess feasibility and acceptance of a mobile clinical decision support system in Botswana.

VisualDx is an artificial intelligence (AI) driven mobile clinical decision support system with documented benefits. Previous studies have demonstrated that implementation of AI systems is commonly associated with challenges such as algorithm bias, privacy, and the protection of all beneficiaries. Prior to implementation of VisualDx in Botswana, ethical review processes at both the University of Botswana and the MOH Research Unit were followed.

Overall, study participants' responses indicated acceptance of the VisualDx platform. The ability to access information quickly without internet connection is crucial in resource constrained environments such as in Botswana. User confidence on the VisualDx platform was likely increased by, 1) prior approval by Institutional Review Boards, 2) the informed consent option prior to participation, 3) adherence to data protection standards such as the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR), 4) collection of de-identified and generalized demographic information about the patient and 5) discarding of patient images immediately after analysis. In order to inform future adoption strategies for VisualDx in Botswana, it is important to evaluate how the platform aligns with standard ethical considerations. Findings could inform policy decisions towards adoption of AI systems in Botswana and similar developing countries.

Background
Botswana is a low-middle-income-country (LMIC) situated in southern Africa, and sharing borders with South Africa, Namibia, Zimbabwe and Zambia. The majority of Batswana live in urban villages (villages surrounding urban areas; 43%), followed by rural villages (36%), and then cities and towns (21%)[2]. The Ministry of Health (MOH) in Botswana is mandated with the oversight and delivery of healthcare services and has recently launched a National eHealth Strategy which recognizes eHealth (the cost-effective use of ICT's for health) as a means of improving healthcare provision and delivery[3]. The number of dermatology specialists in Botswana's public health sector has varied from none to most recently 2 full time MOH employees and three contract specialists from Cuba. However, the demand for dermatology care continues to be much higher than can be provided by the current specialists resulting in six or more months of waiting times for appointments[4]. This 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[5]. It further suggests a critical need for a clinical decision support system (CDSS) to ameliorate current challenges.

Although CDSS offer some documented benefits, it is essential to consider ethical issues arising in their use. Botswana has an established research governance and oversight system for research involving human subjects that has existed since 1980. There exists a mandatory requirement for

a research permit before commencement of any research in Botswana. There are established country-wide Institutional Review Boards (IRBs) at academic and institutional levels as well as Community Advisory Boards (CABs) mostly linked to IRBs.

In 2020, the University of Botswana (UB) collaborated with VisualDx on a research study funded by the Bill & Melinda Gates Foundation (grant number INV003773) to assess the feasibility of VisualDx usage in patient care settings in Botswana and also gather feedback to inform further improvements of the platform. Prior to VisualDx implementation in Botswana, research ethical clearance was sought through UB and MOH. A total of 20 dermatology clinics in Botswana participated and these were nominated by the Gaborone District Health Management Team (DHMT). The DHMT is a local authority under MOHW tasked with overlooking management and staffing of primary care clinics. Two VisualDx employees supported the research project by attending weekly update meetings and also supporting virtual user training. No feature modifications were introduced on the VisualDx platform prior to implementation in Botswana and product intellectual property rights remained with VisualDx.

The recently launched Botswana national eHealth Strategy (2020-2024) notes potential value of emerging technologies, including utilizing mobile devices and IoT (subsection 2.2.3) as is the value of “Sensors to populate digital devices with data” (subsection 2.2.3). It however lacks guidance towards ethical considerations while using AI in healthcare research studies. This shortcoming must be addressed. This study aims to assess the extent to which VisualDx conformed to ethical practices while implemented in Botswana.

Brief history of VisualDx

VisualDx has over 20 years of experience in supporting healthcare providers with their clinical decision making. It employs over 70 full-time team members all dedicated to maintaining accurate, up to date content with user friendly functionality. The platform has become a standard professional resource at more than 2,300+ universities, hospitals, and clinical sites globally. It combines expert knowledge, problem-oriented search, the world’s best curated medical image library, and technology to support differential diagnosis, treatment recommendations, and patient education. VisualDx is available on the web, native iOS and Android applications and most recently includes off-line capability on Android devices.

VisualDx has the potential to contribute to increased provider confidence and a reduction in diagnostic errors in primary care settings[6, 7]. The platform 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 (CNN) to estimate diagnosis and lesion categories from an input image. CNNs are data-driven models that require a large dataset of labeled pairs to train and validate[8].

Study participants

Healthcare workers supporting dermatology clinics and medical students participating in dermatology coursework or rotations at health facilities and universities across Botswana were sent an email and WhatsApp invitations to participate in the study through the eHealth Research Unit at the UB. Consent forms were also provided via email to confirm participation. A total of 28 participants enrolled from 20 sites (healthcare facilities and UB) in Botswana. Participants were based at 6 health districts (Greater Gaborone (21), Greater Palapye (1), Greater Phikwe (2), Greater Francistown (2), Maun (1) and Chobe (1)). All participants used personal smartphones or tablet devices to download and install the VisualDx mobile application, with account credentials provided by VisualDx. They were offered mobile data vouchers to assist with the cost of data for the mobile application download and subsequent usage. Initial training of participants was conducted using the Zoom platform and later in-person training session at the UB eHealth Research Unit. Training sessions covered information technology skills, demonstrations of VisualDx application features and the practical application of VisualDx to common dermatologic and general medical conditions seen in Botswana. All training sessions were recorded and provided to participants who were unable to attend on the training day. Throughout the study

duration, six case-based training sessions were provided to demonstrate the successful use of VisualDx to guide the clinical reasoning process. Participants used VisualDx at their own discretion throughout the study period. A WhatsApp group was created to offer a platform for sharing announcements and seeking support related to the research study.

Ethical issues and commentary on each issue

According to the World Health Organization (WHO), "Non nocere!" (do no harm) is the indispensable principle of the healthcare profession, meant to encourage healthcare practitioners to desist from actions that may result in causing more harm than good[9]. Consequently, in the age of digital health, the new definition of "do no harm" may include that AI driven digital health technologies should "do no harm". If properly implemented, AI in healthcare could uncover clinical best practices from electronic health records (EHRs) by analyzing clinical trends over time, thus assisting in the development of new clinical practice models of healthcare delivery such as precision medicine[10]. A recent study identified that in order "to fully achieve the potential of AI in healthcare, four major ethical issues must be addressed: (1) informed consent to use data, (2) safety and transparency, (3) algorithmic fairness and biases or discrimination, and (4) data privacy are all important factors to consider"[11]. Although the authors acknowledge that these recommendations are related to AI in healthcare, similar considerations can be applied to the use of AI in a research context. The following section outlines how the four ethical considerations were adhered to while using VisualDx in Botswana.

Informed consent to use data

All study participants gave informed consent prior to participation. In fact, the study protocol and consent form were approved by IRBs at the University of Botswana (UB: UBR/RES/IRB/BIO/223) and the Ministry of Health in Botswana (MOHW: HPDME: 13/18/1) prior to implementation.

Safety and transparency

To ensure safety, VisualDx uses peer-reviewed and experts' validated content. The platform is also complaint with the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA). In order to ensure correct and safe use, a series of training and sensitization sessions enabled continuous knowledge exchange. Study site teams allowed interactions among clinicians and specialists. While in use, the VisualDx AI model suggests possible diagnosis and treatment options, and is not prescriptive.

Algorithmic fairness and biases

VisualDx's DermExpert utilizes convolutional neural networks (CNN)[12], a popular deep learning architecture used for computer vision applications. In order to insure fairness and eliminate any biases, the CNN model was trained with over 80 million image variations with different ethnicities consisting of dark and light skins. The model was primarily tested on dark skin colors in Botswana, hence contributing towards the federated learning approach[13], thereby improving its fairness.

Data privacy

VisualDx collects only de-identified and generalized demographic information about the patient to provide a differential diagnosis. Even when using the 'DermExpert' AI tool, the image of the patient remains on the device at all times and is discarded immediately after the analysis is complete. This alleviates any data security concerns and allows the tool to conform to data protection standards such as the HIPAA and the GDPR[14,15].

Limitations and barriers to AI adoption

Despite the benefits that come with ethical use of AI in healthcare, more often health systems' infrastructure is not ready to support such innovations, especially in LMICs[16]. In Botswana, healthcare systems (across and within public and private sectors) are unique in technology adoption and care processes[17]. The private sector is ahead while the public sector is in the process of catching up on digital transformation. This hinders widespread adoption or use of AI algorithms. The lack of locally generated quality data to continuously train AI models is a significant barrier. In our research study, we used an AI model trained using external datasets hosted on VisualDx cloud

servers. The regulatory role for AI systems in Botswana is still at an infancy stage, as well as the establishment of "good machine learning practices," and robust oversight mechanisms.

Conclusions and recommendations

VisualDx was a well-received amongst the study population in Botswana and has the potential to upskill and empower general practitioners to do more at the point of care. While the AI capabilities of VisualDx may not be able to completely replace clinical judgment, it can help clinicians make better decisions. The need for increased algorithmic transparency, privacy, and protection of all the beneficiaries is essential. Similar considerations are outlined in the Botswana Data Protection Act (DPA)[18] (Botswana Data Protection Act, 2018). In fact, efforts to utilize AI in Botswana should align with the DPA for guidance on security, privacy and confidentiality considerations. The national eHealth strategy should further guide ethical use of AI systems to support healthcare provision. VisualDx image security could be enhanced by disabling 'screen-capture' feature while using mobile devices.

References

1. Nkomazana, O., Mash, R., Shaibu, S. and Phaladze, N., 2015. Stakeholders' perceptions 2851 on shortage of healthcare workers in primary healthcare in Botswana: focus group 2852 discussions. *PloS One*, 10(8), p.e0135846.
2. Statistics Botswana, 2017. Botswana Demographic Survey Report 2017, Gaborone: Statistics Botswana.
3. Ministry of Health and Wellness Botswana. The eHealth Strategy of Botswana (2020–2024), "A healthy and productive nation, that enjoys the benefits of high quality eHealth services." Gaborone , Botswana; 2020. https://ehealth.ub.bw/bhdc/ehealthstrategy.html.
4. Williams VL, Narasimhamurthy M, Rodriguez O, Mosojane K, Bale T, Kesalopa K, Kayembe MA, Grover S. Dermatology-Driven quality improvement interventions to decrease diagnostic delays for Kaposi sarcoma in Botswana. *Journal of Global Oncology*. 2019 Nov;5:1–7
5. Mosam A, Todd G. Dermatology Training in Africa: Successes and Challenges. *Dermatol Clin*. 2021 Jan;39(1):57–71. https://doi.org/10.1016/j.det.2020.08.006. PMID: 33228862.
6. Breitbart EW, Choudhury K, Andersen AD, Bunde H, Breitbart M, Sideri AM, et. al. Improved patient satisfaction and diagnostic accuracy in skin diseases with a Visual Clinical Decision Support System—A feasibility study with general practitioners. *PloS one*. 2020 Jul 29;15(7):e0235410.
7. Dulmage B, Tegtmeyer K, Zhang MZ, Colavincenzo M, Xu S. A Point-of-Care, Real Time Artificial Intelligence System to Support Clinician Diagnosis of a Wide Range of Skin Diseases. *J Invest Dermatol*. 2021 May 1;141(5):1230-5. https://doi.org/10.1016/j.jid.2020.08.027.
8. M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in Proceedings of the 36th International Conference on Machine Learning, Jun. 2019, vol. 97, pp. 6105–6114. [Online]. Available: http://proceedings.mlr.press/v97/tan19a.html
9. World Health Organization. "Non nocere!" (Do No Harm) - Preventing and Addressing Medication Errors. https://www.afro.who.int/news/non-nocere-do-no-harm-preventing-and-addressing-medication-errors ; 2018 (accessed June, 2022)
10. Mesko, B., 2017. The role of artificial intelligence in precision medicine. Expert Review of Precision Medicine and Drug Development, 2(5), pp.239-241.
11. Naik N, Hameed BZ, Shetty DK, Swain D, Shah M, Paul R, Aggarwal K, Ibrahim S, Patil V, Smriti K, Shetty S. Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility? *Frontiers in Surgery*. 2022;9.
12. Albawi, S., Mohammed, T.A. and Al-Zawi, S., 2017, August. Understanding of a convolutional neural network. In 2017 international conference on engineering and technology (ICET) (pp. 1-6).
13. Li, T., Sahu, A.K., Talwalkar, A. and Smith, V., 2020. Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, 37(3), pp.50-60.
14. Contracting with Vendors that are NOT HIPAA Business Associates: Best Practices. [Internet]. 2022. https://www.hipaa.com/. Accessed 15 June 2022.

15. Complete guide to GDPR compliance. [Internet]. 2022. https://gdpr.eu/. Accessed 15 June 2022.
16. Guo J, Li B. The application of medical artificial intelligence technology in rural areas of developing countries. *Health equity*. 2018 Aug 1;2(1):174-81.
17. Ndlovu, K., Scott, R.E. & Mars, M. Interoperability opportunities and challenges in linking mhealth applications and eRecord systems: Botswana as an exemplar. *BMC Med Inform Decis Mak* 21, 246 (2021). https://doi.org/10.1186/s12911-021-01606-7.
18. Government of Botswana, Data Protection Act. Law 32 of 2018. https:// www.bocra.org.bw/sites/default/fles/documents/DataProtectionAct.pdf.

Theme 3: Collaborative initiatives and data resources to support AI health research

Case study: Ethical considerations in implementing The Data Advancing Wellness in Africa (DAWA) Project

Joan Gakii Masunga, MBE, Postgraduate Research Fellow, Harvard Medical School, Boston, USA

Brief description of the research project

The **D**ata for **A**dvancing **W**ellness in **A**frica (**DAWA,** Swahili for "Medicine"**)** hence named **The DAWA Project** 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.

The specific aims of the DAWA project are:

1. *To create a tool, an African NCD Big Data Commons, for the generation of timely health information and wellness products at varied levels of complexity, be it in the patient, practice, policy and/or economic frameworks.*

   The integration and harmonisation of data sources from various health organisation databases, geotagged smartphone devices as well as environmental and geospatial datasets from public sources will allow for the creation of an African NCD Big Data Commons. This will allow complex levels of data to be subsequently mined and enable us to deliver public health gains through the insights that big data analytics can leverage. This platform will be enhanced using automated data retrieval and harmonisation techniques and validated for future real-time collection and update of data into the Big Data Commons.

2. *To study the multifactorial determinants of, and outcomes from, cardiometabolic disease and mental illness in SSA using big data and its analytics.*

   Data science will allow precise identification of high-risk populations, and comprehensive understanding of the study of diseases, including the interactions between behavioural, social, environmental, and economic determinants of health risk factors. The results can improve population surveillance and thus improve and target interventions for health promotion and disease prevention.

Background

Cardiometabolic diseases and mental illness continue to present a formidable burden of disability and premature mortality in sub-Saharan Africa. These populations also face disparities in healthcare and an ever-increasing cluster of behavioural, social, environmental and economic risk factors. Countries across SSA vary extensively in respect to their social and economic development and historical trajectories, each with their own specific health challenges and outcomes. Interventions and health promotion activities touted from traditional epidemiological studies from the US and Europe have been limited by the lack of African data available and variables studied and have not impacted sufficiently at the population level. These methods also pose significant timeliness and efficiency barriers from research to implementation, efficiency limitations and lack of spatial resolution. Currently, major impediments to advancing our battle in non-communicable disease (NCD) prevention and control in SSA are:

a. The power to find unexpected associations, though potentially without substantive relevance
b. The capacity to assess complex interactions with more complicated variable selection
c. The potential to design dynamic interventions.

The DAWA project is being initiated to address these issues in collaboration with four African institutions located in South Africa, Nigeria, Tanzania, Uganda and three US institutions located in Boston, Massachusetts**.** This will provide an ecosystem comprising data scientists, engineers, software technicians, global health experts, epidemiologists, biostatisticians, and community representatives to advance data science applications for curbing the NCD burden. It will also further develop much needed human capacity to strengthen data science and computational thinking platforms on the African continent.

Ethical issues and commentary on each issue

Trust and trustworthiness

1. The DAWA project will collect considerable data to be analysed from participants that will be used by at least four distinct groups of stakeholders namely, health care providers, researchers, entrepreneurs and policy-experts. This data will include sensitive health information which raises questions of autonomy and privacy. Situated in the context of historical exploitation of the African continent, questions of trust, privacy, protection, and autonomy are paramount. For this project to be impactful, participants need to trust the researchers and staff requesting their personal data. It is therefore imperative that the research team establish trust and transparency with the participants[1]. One way to do this is by elaborating how the collection of their health data will benefit them. A study on data sharing in health research in Kenya showed that research participants deem this to be of primary importance as they consider sharing their data for research[2]. Additionally letting the researchers from the respected local institutions who have established relationships with the community members lead the project may help allay fear and apprehension of potential participants.

Fairness and equity

2. Identification and mitigation of algorithmic and data bias has received considerable attention in High Income Countries as AI use has become more widespread and societies have had more time to identify the problems. In Low-and Middle-income countries, however, because adaptation is still not widespread, the efficiency gains from AI interventions receive a lot more attention than the potential risks. Therefore, in implementing the DAWA project, bias identification and mitigation needs to be thought about pre-emptively[3].

3. The DAWA study also seeks to elucidate underlying moral viewpoints within participants that may inform their health-seeking, political, and communal behaviours. The philosopher John Mbiti defines African personhood in a relational and communal context where having a sense of duty to others is normative[4]. It is therefore unwise to assume that Western individualistic moral viewpoints inform the behaviours and thought processes about justice and fairness of Africans. It is also equally misguided to assume that just one theory could be generalized to the entire continent[5]. Therefore, this study seeks to understand the underlying values specific to each cultural context to begin to understand what the epidemiological results mean in this context.

4. The DAWA project will be a collaboration between four institutions based in Africa and three based in the US. There is an undeniable power differential between the institutions which could alter in-group relations and participant willingness and disclosure. These factors could affect knowledge generation and production and therefore must be critically examined. One example of how this could present is for researchers from the US based institutions to replicate research designs from their home institutions that may render the study irrelevant for the target populations where the study is being conducted. For DAWA to be successful, mutual trust and respect must be established among researchers and staff across these organisations and all other stakeholders involved in the project.

5. The widespread penetration of cellular phones across Africa makes conducting the DAWA project possible. However the WHO Ethics and Governance of Artificial Intelligence for Health report shows that women in LMICs have less access to cellular phones and the internet[6]. Concerted effort to ensure that women are sufficiently recruited to participate in the study is therefore needed to ensure that they are represented in the datasets.

Conclusions and recommendations

For the DAWA project to be impactful, it is important to incorporate the perspectives of all the relevant stakeholders pertinent to this project in the design and development of AI systems, to ensure that the issues that are relevant to them are captured. Another important exercise to demystify AI in health research is to hold focus groups with proposed study participant groups to gauge their level of understanding and provide subsequent education to fill knowledge gaps.

Definitions of sensitive/protected groups need to be widened for an Africa specific context, so that such identifier data is collected proactively, and used for auditing later for potential bias. For example, whereas in the West, race is a significant sensitive feature, in Africa other sensitive features may include tribe, clan and religion. After identifying these potential dimensions of discrimination, conscious effort can then be made to ensure data collected covers all these groups.

In addition, in stipulating the type of projects that can utilize the data from DAWA, methodologies to document the provenance, creation, and use of machine learning datasets such as “datasheets for datasets”, should be encouraged to avoid discriminatory outcomes. This is an example of a best practice that can be embedded to inform the technical design and development of AI for health research and to mitigate potential unforeseen risks. Further, the machine learning community has in recent years developed various definitions of fairness. Examples include individual fairness and group fairness. Specific metrics include demographic parity, equality of opportunity and equality of odds. Depending on specific applications, a collaborative effort between researchers and subjects will be needed to identify the most applicable metrics on a case-by-case basis.

Finally, international bodies such as the WHO should consider instituting rules and regulations to guide the implementation of AI research in LMICs by external parties. Establishing of regulatory frameworks of operation should not be left to the goodwill or trustworthiness of researchers, global tech companies and their AI developers. A globally established framework will ensure that local populations are not exploited and that their rights and benefits are prioritized and served[7].

References

1. Kerasidoua A. Artificial intelligence and the ongoing need for empathy, compassion and trust in healthcare: La inteligencia artificial y la continua necesidad de empatía, compasión y confianza en la atención sanitaria. *Intell Artif Besoin Constant Empathie Compassion Confiance Dans Sect Santé*. 2020;98(4):245-250. doi:10.2471/BLT.19.237198
2. Jao I, Kombe F, Mwalukore S, et al. Research Stakeholders’ Views on Benefits and Challenges for Public Health Research Data Sharing in Kenya: The Importance of Trust and Social Relations. *PLOS ONE*. 2015;10(9):e0135545. doi:10.1371/journal.pone.0135545
3. Gebru T, Morgenstern J, Vecchione B, et al. Datasheets for Datasets. Published online December 1, 2021. doi:10.48550/arXiv.1803.09010
4. Mbiti JS, Mbiti. *African Religions & Philosophy*. Heinemann; 1990.
5. Atuire CA. African perspectives of moral status: a framework for evaluating global bioethical issues. *Med Humanit*. 2022;48(2):238-245. doi:10.1136/medhum-2021-012229
6. World Health Organization. *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*. World Health Organization; 2021. Accessed October 1, 2022. https://apps.who.int/iris/handle/10665/341996
7. Kerasidou A. Ethics of artificial intelligence in global health: Explainability, algorithmic bias and trust. *J Oral Biol Craniofacial Res*. 2021;11(4):612-614. doi:10.1016/j.jobcr.2021.09.004

Governance paper: Responsible research and development in AI for healthcare: what we are learning from establishing a national collaborative platform in the UK

Dr Kate Devlin, King's College London, UK

Context: National good practice/guidance/frameworks for research – setting up a government-funded research hub on trusted and trustworthy systems.

Introduction

Today, easy access to large amounts of data and of computing power means machine learning is embedded and used in many of the applications we take for granted. However, power inequalities in AI at a global level – the dominance of corporate control in the US, UK and Europe – mean that AI systems and their creators are designing and deploying products that inherently benefit them and often actively harm outgroups.

In healthcare, AI shows clear promise in automating diagnoses and personalise treatments. This includes the use of autonomous systems: software and machines that are able to take actions with little or no human supervision. For example, during the Covid-19 pandemic, autonomous systems such as mobile robots were able to significantly reduce the risk of infectious disease transmission[1]. But automation is not without its social impact downstream: autonomous systems can threaten clinicians' agency, could lead to reduced funding in staffing, and could result in job loss, as well as facing scepticism and concern from patients.

As the social repercussions of AI emerge, the ethics community has paid closer attention to methods of mitigation, such as guidelines, principles and frameworks[2-5] and a wider call for attention to responsible research and innovation. Actual regulation, however, is not easy given issues around jurisdiction, corporate pushback, and lack of algorithmic transparency[6].

While there is much agreement on the need for ethical AI, the tangible steps – the ethics in action – are very much a work in progress. There is no quick software fix for AI's flaws, nor can we engineer out deep-rooted and systemic biases overnight. However, now is the time to implement practical guidelines and approaches from the ground up. An additional responsibility is to earn the trust of the end user via transparent development, clear communication, and informative engagement.

The Trusted Autonomous Systems Hub

With the above in mind, three UK universities – Southampton, Nottingham, and King's College London – applied to establish a national network on trust in AI that foregrounded ethical, responsible and inclusive development. The resulting UK's Trusted Autonomous Systems Hub[7] is the world's largest research programme in Trustworthy AI and Autonomous Systems. 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 has over 100 international industry partners. Health and AI is a large part of our remit.

From the start, we actively sought to create a programme that would begin to address these issues by making our work contingent on adhering to fairness and equity in all that we do. Our goal is to deliver world-leading best practices for the design, regulation and operation of socially beneficial autonomous systems which are both trustworthy in principle and trusted in practice.

TAS is built around the core principles of responsible research and innovation (RRI) with equal attention to equality, diversity and inclusion (EDI) – ethical by design. These are issues that are key our research, and any projects we fund must centre these criteria in named, practical ways before we will consider supporting them. We fund projects, set up networks, advise on policy, and invite researchers, industry, NGOs and the public to engage and contribute use-cases/datasets or collaborate on research projects, tech transfer, and training activities.

TAS carries out research internally and also awards open grants to UK academics and industry partners. For our pump-priming funding – pilot studies to explore areas of AI and trust – our co-investigators, network members and industry partners worked on the following health research:

Trustworthy light-based robotic devices for autonomous wound healing – Robotic technologies have the potential to guide wound healing at the cellular level. Machine learning allows us to tailor the control to individual cellular dynamics on the go, enabling personalised solutions. This raises questions about how to ensure these systems are trustworthy and safe.

Identifying conflict and confluence in stakeholder imaginaries of autonomous care systems – Identifying the conflicts and confluences in the imaginaries of robotic and autonomous systems in the health-social care ecosystem by conducting of expert interviews and workshops with stakeholders and members of the public.

A participatory approach to the ethical assurance of digital mental healthcare – Developing a novel approach to assurance through participatory methodology, to underwrite the responsible design, development, and deployment of autonomous and intelligent systems in digital mental healthcare.

Co-designing Trustworthy Autonomous Diabetes Systems – Designing algorithms for diabetes management during life transitions using co-design, provenance and explainable AI, bringing together clinicians, data scientists, and people with type-1 diabetes.

Trustworthy autonomous systems to support healthcare experiences – How trustworthy autonomous systems embedded in devices in the home can support decision-making about health and wellbeing.

Diagnostic AI System for Robot-Assisted A&E Triage – Prototyping a robot-assisted A&E triage solution for reducing patient waiting time and doctor workload.

TAS Hub is in its second year now and we are already gaining valuable insight in how to set up a national framework for conducting ethical, responsible research in AI.

Conclusion and recommendations

This practical governance paper focuses on the bigger picture: the move from theory to practice in terms of AI ethics and responsible research development. There are two aspects to our work which we seek to highlight at GFBR:

- First, *trustworthy in principle*, where we share what we have learned from setting up a large, nationally-funded, multidisciplinary programme on safe AI – one that is novel in its approach by centering ethics and ethical development ahead of results and outputs. This involves the successes and also the failures, and the course corrections we have had to take. This includes:
    - devising criteria for grant reviews that reward projects which centre stakeholder engagement;
    - promoting early career leadership opportunities;
    - ensuring tangible ethical approaches;
    - writing actionable equality, diversity and inclusion strategies – and using them;
    - forming an operational framework;
    - collaborating with industry (and researchers' varying reactions to who we work with – some people have strong views on who invests in our work); and
    - the choices we made when setting up our Board, our Strategy Advisory Network and our International Scientific Committee.

- Second, *trusted in practice*, where we are embarking in collaboration not only with industry and academia but also with end-users and the public. This includes such aspects as participatory and collaborative project design, as well as the resulting education and

outreach, as we strive to be transparent about a future with autonomous systems. This includes:

- working with stakeholders, such as charities and mental health service-users, to explore ideas around trust in digital mental healthcare systems;
- providing resources such as video conversations, podcasts, and teaching materials; and
- commissioning and developing interactive creative artwork that leads to thought-provoking encounters – for example, our Cat Royale art project that involves pets being cared for and played with by a robot arm[8].

We will describe what is working well for us as we establish a national platform, and what struggles we have faced, particularly when operating in a sector which often prioritises results over method, and one where industry collaboration may mean differing expectations or priorities. We hope that this information can be shared widely and freely to encourage similar initiatives, and we seek collaboration on this type of work from existing or nascent programmes. We are particularly keen for anything people find useful from TAS to be reused, adopted and adapted as needed by those working towards the same goal.

References

1. Tavakoli, M., Carriere, J., & Torabi, A. (2020). Robotics, smart wearable technologies, and autonomous intelligent systems for healthcare during the COVID-19 pandemic: An analysis of the state of the art and future vision. Advanced Intelligent Systems, 2(7), 2000071.
2. Bartoletti, I. (2019). AI in Healthcare: Ethical and Privacy Challenges. In: Riaño, D., Wilk, S., ten Teije, A. (eds) Artificial Intelligence in Medicine. AIME 2019. Lecture Notes in Computer Science(), vol 11526. Springer, Cham. https://doi.org/10.1007/978-3-030-21642-9_2
3. Dignum, V. (2019). Responsible artificial intelligence: how to develop and use AI in a responsible way. Springer.
4. Winfield, A. (2019). Ethical standards in robotics and AI. *Nature Electronics*, 2(2), 46–48.
5. Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). How to Design AI for Social Good: Seven Essential Factors. Science and Engineering Ethics, 26(3), 1771–1796.
6. Hoffmann-Riem, W. (2020). Artificial intelligence as a challenge for law and regulation. In *Regulating artificial intelligence* (pp. 1-29). Springer, Cham.
7. Trusted Autonomous Systems Hub web site: https://www.tas.ac.uk/ (accessed Sept 2022)
8. Cat Royale by Blast Theory https://www.youtube.com/watch?v=WLAy3GWBvzQ (accessed Sept 2022)

Theme 4: Regulation of data for health research involving AI

Governance paper: Governance of cross-border transfer of data in Sub-Saharan Africa (SSA)

Nezerith Cengiz (presenter)**,** Centre for Medical Ethics and Law, WHO Collaborating Centre in Bioethics, Department of Medicine, Faculty of Health Sciences, Stellenbosch University, South Africa
Co-authors: Dr Dirk Brand, Prof Jerome Amir Singh, Prof Annelize McKay, Prof Keymanthri Moodley

Brief description of the context
Large research networks and collaborative projects in sub-Saharan Africa (SSA) mean that data must often be transferred or shared amongst African and international countries. SSA's most research-intensive countries are characterised by diverse data management and privacy governance frameworks. Such regional variance can impede time-sensitive data sharing and highlights the need for urgent governance reforms to facilitate effective decision-making in response to rapidly evolving public health threats. 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. We explore governance considerations that ought to apply to the collection, transfer, and use of data; and provide an overview of the prevailing data-sharing governance landscape in SSA's most research-intensive countries. As a result, we identify some key limitations and gaps that impede effective data collation, sharing and analysis. A range of stakeholders such as data scientists, researchers, artificial intelligence (AI) coders, and government decision-makers may benefit from and find this paper useful. The issues explored here are of universal concern and therefore of relevance to the African context as well as a broader international audience.

Commentary
The sharing/transfer of data between countries and national institutions in SSA significantly strengthens research capacity[1]. However, concerns regarding the cross-border flow of data and privacy protection have been raised. Africa lacks the capacity and resources to build, maintain, and analyse large data sources and datasets required by AI systems; consequently hindering the continents' ability to make informed, evidence-based decisions in healthcare or policy development to describe related challenges[2]. Data protection legislation in SSA does not, often, adequately address the lawful use of data in the development of AI tools although it is required to guide its ethical use in healthcare and offer guidance to software developers and other stakeholders.[2] Since large datasets are required in the development of AI tools in healthcare, concerns about privacy, accountability, and transparency among others are raised as its misuse could adversely impact individual data subjects and/or society[1;2;3]. Accordingly, data ethics plays a vital role in developing AI applications and evaluating large datasets and related activities (collection, analysis, sharing/transfer, and use).

Table 1 categorises the rigour of national data protection laws concerning the cross-border transfer of personal data[4].Table 1 is not aimed at providing a strict overall categorisation of various data protection laws, but rather, is focused on the scope of legal protection afforded to data subjects regarding the cross-border transfer of their personal data. Countries with stringent rules require notification of, or approval by, a relevant data protection authority, and/or special conditions (such as proof of appropriate safeguards concerning the protection and security of personal data), as well as consent from the data subject.

South Africa and Kenya count among the countries that could be described as providing stringent data export protection to data subjects. For example, Kenya's Data Protection Act of 2019 complies with the European Union (EU) legal standards, which are generally regarded as being stringent in nature. For data to be transferred out of Kenya, the data processor must verify to the data commissioner that the third-party recipient's jurisdiction is bound by appropriate safeguards for the security and protection of the data. It is also important that the data transfer be purposeful, such as necessary for the conclusion or performance of a contract, a legal claim, and public or data subjects' interests. In addition, consent from the data subject is also required for cross-border data transfers[4].

Countries falling in the moderate category allow for more than one possible legal ground to permit data export, such as consent of the data subject, but do not require notification or approval by the data protection authority. Nigeria counts amongst countries providing moderate data export protection to data subjects as the country's data protection law does not require third-party recipients of data to be bound by adequate data protection law, agreements, or corporate rules if the data subject provides consent after being informed of possible risks of inadequate data protection or if the transfer meets a certain exception. One example of such an exception is the public's or data subject's interest. Beyond obtaining consent from data subjects for data transfers, the Nigeria Data Protection Regulation 2019 requires the National Information Technology Development Agency (NITDA) or Honourable Attorney General of the Federation (HAGF) to ensure that the third-party recipients of the transferred data have adequate data protection standards in place[4].

Ghana's data protection legislation does not contain any provisions pertaining to the cross-border transfer of personal information and could thus be described as providing inadequate protection to data subjects in relation to the export of their personal data[5].

The diverse legal landscape governing data sharing in sub-Saharan Africa – including the stringency of data export provisions – highlights that cross-border data transfers will have to be evaluated on a case-by-case basis as there is no uniform law across the continent akin to the General Data Protection Regulation (GDPR) (2018), which constitutes a common legal framework for all EU Member States. Although the AU Commission is developing a data policy framework for Africa to harness digital technologies and innovation in an attempt to bridge the digital divide, this process is ongoing and will take time to implement[6].

Table 1: *Sub-Saharan Africa country rankings by research output ("Public Health, Environment, and Occupational Health")*[7].

Rank and country	Legal Requirements	Legislation	Data export protection classification
South Africa	A responsible party may only transfer personal data outside South Africa if the recipient is subject to a law, binding corporate rules or the binding agreement that provide adequate protection. Or the data subject consents to the transfer; or The transfer is necessary for the terms of the provisions of the Act.	Sec. 72 of the Protection of Personal Information Act, 4 of 2013 (South Africa)	Strict
Nigeria	Cross-border transfer of personal data is subject to authorisation by the Attorney General or the National Information Technology Development Agency (NITDA) based on an adequate level of protection. In the absence of authorisation by the Attorney General or the Agency, personal data transfer may only take place if the data subject gave consent, or the data transfer is necessary in terms of the Regulation.	Reg. 2.11 and 2.12 of the Nigeria Data Protection Regulation, 2019.	Moderate
Kenya	Only allowed if there is proof of adequate data protection safeguards or consent from the data subject. Data controller or data processor must provide proof to Data Commissioner on appropriate safeguards. The data transfer must be necessary in terms of the Act.	Sec. 25(h) 48 of the Data Protection Act, No. 24 of 2019 (Kenya)	Strict
Ethiopia	Cross-border data transfer may only take place subject to an adequate level of data protection in the recipient country. Data controller or data processor must provide proof to Data Protection Commission of appropriate level of protection, or the data subject has given consent to the proposed transfer, or the transfer is necessary, or the transfer is made from a register and intended to provide information to the public.	Sec. 27-30 of the Draft Proclamation to Provide for Personal Data Protection, 2021 (Ethiopia)	Strict
Uganda	Data processors or data controllers must ensure that there are adequate measures in place for the protection of personal data, or the data subject must provide consent.	Sec. 19 of the Data Protection and Privacy Act, 2019 (Uganda)	Strict

Conclusion and recommendations

Given the lack of data protection legislation in SSA, we aim to provide guidance on ethical data sharing. Harmonised data sources and their integration into national health information systems will create a comprehensive dataset. A holistic approach to data management should underpin evidence-based decision-making. To facilitate cross-border data transfers involving personal data, standard contractual provisions and templates for cross-border data transfers should be developed by data protection authorities in Africa. Doing so will facilitate not just scientific cooperation between countries, but also facilitate an integrated cross-border approach to the management of future pandemics. SSA countries should aim to strengthen their digital infrastructure for capturing and storing data to aid in building appropriate analytical capacity. To enhance both the use of and access to data in the context of AI, principles of transparency, fairness, and accountability would strongly aid with the establishment of a reliable and accessible digital ecosystem in SSA.

References

1. The World Health Organisation. Ethics and governance of artificial intelligence for health. 2021 June 28 [cited 2022 June 15]. Available from: https://www.who.int/publications/i/item/9789240029200
2. Owoyemi A, Owoyemi J, Osiyemi A, Boyd A. Artificial Intelligence for Healthcare in Africa. Front Digit Health. 2020 Jul 7;2:6. doi: 10.3389/fdgth.2020.00006. PMID: 34713019; PMCID: PMC8521850.
3. Abebe R, Aruleba K, Birhane A, Kingsley S, Obaido G, Remy S, et al. Narratives and Counternarratives on Data Sharing in Africa. In: 2021 FAccT ACM Conference on Fairness, Accountability, and Transparency. ACM FAccT; 2021. p. 329–41.
4. Suominen K, Vambell E. Alliance for E-Trade Development: Toward an African Data Transfer Regime to Enable MSMEs' Cross-border Ecommerce [Internet]. 2021 Sep [cited 2022 June 15]. Available from: https://www.allianceforetradedevelopment.org/_files/ugd/478c1a_72021e35a826441db0723642a79e65e5.pdf
5. The Parliament of the Republic of Ghana. Data Protection Act, 2012. Data Protection Act, 2012, 843 Republic of Ghana; Oct 16, 2012.
6. Africa Union. The digital transformation strategy for Africa (2020-2030) [Internet]. Addis Ababa; 2021 Aug [cited 2022 June 15]. Available from: https://au.int/sites/default/files/documents/38507-doc-dts-english.pdf
7. Scimago Journal & Country Rank. Country Rankings: Public Health, Environment, and Occupational Health [Internet]. Scimago Journal & Country Rank. 2022 [cited 2022 Aug 23]. Available from: https://www.scimagojr.com/countryrank.php?region=Africa&category=2739 from: https://www.allianceforetradedevelopment.org/_files/ugd/478c1a_72021e35a826441db0723642a79e65e5.pdf

Governance paper: Regulation of health data for AI in Uganda

Harriet Nankya, Makerere University, Uganda

Context: Assessing Uganda's regulation of health data in reference to the World Health Organization's recommendations for health data protection in the development and application of AI.

Commentary

Artificial Intelligence (AI) holds great promise to improve health. It can enable more accurate diagnosis and treatment of diseases, support pandemic preparedness and response, inform the decisions of health policy-makers or allocate resources within health systems[1]. However, to fully reap the benefits of AI, ethical challenges to its development and application must be addressed. Important issues to consider arise in the informed consent to use data, data safety and transparency, algorithmic fairness and biases, and data privacy[2]. This calls for the need to prioritize the ethical principles and human rights obligations by those who fund, design, regulate or use AI technologies for health, to avoid potential serious negative consequences.

The World Health Organization (WHO) in 2021 issued a report after analysing many opportunities and challenges of AI, and recommended policies, principles and practices for ethical use of AI for health and means to avoid its misuse to undermine human rights and legal obligations[1]. WHO hoped that these principles would be used as a basis for governments, technology developers, companies, civil society and inter-governmental organizations to adopt ethical approaches to appropriate use of AI for health. One of the key principles endorsed was protecting human autonomy; this principle 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 WHO endorsement points to the fact that the development of a successful AI system for health relies on high-quality data but systems can suffer with uneven management of such sensitive health data[3]. This presents several risks, for example, one's personal data may end up in the wrong hands or be used contrary to the owner's wish. Therefore, there is a need for data privacy and security in the research and implementation of AI-based health technologies, for compliance purposes and to build public trust in these solutions[4]. Currently in Uganda, just like in some other countries where technology advancement is still low, there are scant well-defined regulations in place to address the legal and ethical issues that may arise in the research and use of AI in health settings[2]. AI systems have been subject to sector-specific laws or subject-specific guidelines such as data-protection acts, cyber-security laws, anti-discrimination regulations. These measures have been applied on a haphazard and piecemeal basis creating large regulatory gaps and ethical implications of AI usage[3].

By basing on the WHO recommendations for health data protection in the development and application of AI, this paper describes how Uganda is positioned to comply with some of these recommendations on the regulation of health data for AI, as explained herein.

Recommendation 1*. Governments should have clear data protection laws and regulations for the use of health data and protecting individual rights, including the right to meaningful informed consent.*

Uganda passed the Data Protection and Privacy Act, 2019 ('the Act') in 2019 and the Data Protection and Privacy Regulations, 2021 ('the Regulations') in May 2021. The Act and Regulations are intended to support the protection of privacy and personal data through regulation of its collection, processing and storage. These privacy protections are already guaranteed to Ugandans under the Constitution and complement sectoral laws for regulated activities. The Act also guarantees the protection of privacy in the digital world. This Act mirrors the UK Data Protection Act, 1998 which revolves around several principles concerning data protection and collection. The Act is also in line with a number of international conventions including; the Universal Declaration of Human Rights to which

Uganda is a signatory, the African Union Convention on Cyber Security and Personal Data Protection and the GDPR. It is also in line with the European Union General Data Protection Regulation (GDPR).

It is, therefore, very important for companies and other persons using AI and big data systems to abide by the strict requirements of the Act before collecting or processing personal data. This Act, manifests as a comprehensive law in regards to the AI technological advancements that could affect the right to privacy.

Recommendation 2. *Governments should establish independent data protection authorities with adequate power and resources to monitor and enforce the rules and regulations in data protection laws.*

Uganda has the Personal Data Protection Office (the Office) which is the national independent data protection authority. It is established as an independent office under the National Information Technology Authority, Uganda (NITA-U) responsible for overseeing the implementation of and enforcement of the Data Protection and Privacy Act No. 9 of 2019.

Section 3 of the Regulation stipulates that the Office, in the performance of its functions, is independent and not subject to the direction or control of any person or authority. Section 3(3)b of the Regulations points out that the affairs of the National Information Technology Authority, Uganda are managed separately from the affairs of the Office.

Section 5 of the Regulation stipulates the power of the office in carrying out the functions specified under the Act; In Section 5(a), the Office may establish a mechanism for collaboration and promotion of partnerships between various categories of players in the data protection and privacy aspects; and section 5(b) the Office will charge fees for services provided by the Office. In enforcing the regulations of the Act, Section 6 of the regulations stipulates that the Office shall cooperate with other government authorities like ministries, departments and agencies. The Office would, therefore, cooperate with agencies like; the Uganda National Council for Science and Technology (UNCST). UNCST is a government of Uganda Agency, established under the Ministry of Finance Planning and Economic Development to coordinate the formulation of national policy on all fields of science and technology, and for assisting in the promotion and development of indigenous science and technology[5]. UNCST works in collaboration with Uganda National Health Research Organization (UNHRO) for health research. With that stance, health research guided by UNCST and UNHRO is under the oversight of the Office.

In June 2022, The Office and the United Nations Capital Development Fund (UNCDF), launched a data protection and privacy portal that would ease reporting, processing, and resolving of data protection and privacy complaints and breaches[6]. The portal includes SMS/USSD functionality to enable universal access and usage by most citizens. UNCDF's support to the Office to develop the data protection portal is part of its 'Leaving No One Behind in the Digital Era Strategy'. The portal strategy, therefore, aims to empower millions to use digital services that will leverage innovation and technology to improve their wellbeing, while contributing to the Sustainable Development Goals.

The advent of AI and big data is set to raise a number of human rights issues. For example, the requirement for large volumes of data is likely to see the right to privacy of data being forsaken since the personal data of individuals is being shared and/or processed without their knowledge or consent. To counter this, NITA-U will have to step up its regulatory function to protect the integrity of personal data.

Recommendation 3*. Governments should require entities that seek to use health data to be transparent about the scope of the intended use of the data.*

Section 12 of the Act states that; A person who collects personal data shall collect the data for a lawful purpose that is specific, explicitly defined, and is related to the functions or activity of the data collector, or data controller. Section 17 guides more on further processing of the data. It allows further processing of data but this should be specific to the purpose for which the data was collected (section 17 (1)). Section 17 (2) stipulates what should be put into account when further processing the data including; the relationship between the purpose of the intended further processing and the purpose for which the data was collected; the nature of the data concerned; the manner in which the data has been collected; the consequences that the further processing is likely to have for the data subject; and the contractual rights and obligations between the data subject and the person who processes the data (section 17 (3)(c)). Further processing of data is allowed for forensic purposes including national security, law enforcement etc. Further processing of data is also allowed for historical, statistical or research purposes (section 17 (3)(e)). Section 19 stipulated that for processing personal data outside Uganda**,** the data processor or data controller shall ensure that; (a) the country in which the data is processed or stored has adequate measures in place for the protection of personal data at least equivalent to the protection provided for by the Act; or (b) the data subject has consented.

Recommendation 4. *Mechanisms for community oversight of data should be supported. These include data collectives and the establishment of data sovereignty by indigenous communities and other marginalized groups.*

The Act doesn't stipulate community oversight of data. However, Section 9 highlights prohibitions on the collection and processing of special personal data which relates to the religious or philosophical beliefs, political opinion, sexual life, financial information, health status, or medical records of an individual.

Recommendation 5. *Data hubs should meet the highest standards of informed consent if their data might be used by the private or public sector, should be transparent in their agreements with companies, and should ensure that the outcomes of data collaboration provide the widest possible public benefit*

Neither the Act nor the Regulation stipulates data handling by the data hubs. However, Section 34 (1) of the Regulation stipulates that a data collector, data processor, or data controller who collects or processes personal data without the prior consent of the data subject in contravention of section 7(1) of the Act, commits an offense and is liable, on conviction to a fine not exceeding three currency points for each day that the contravention continues or to imprisonment not exceeding six months or both. Section 34 (2): Where the offense in sub-regulation (1) is committed by a corporation, the corporation and every officer of the corporation who knowingly and willfully authorizes the collecting or processing of personal data in contravention of section 7(1) of the Act, commits an offense and is liable, on conviction; to a fine specified in sub-regulation (1).

Recommendation 6. *Governments should enact laws and policies that require government agencies and companies to conduct impact assessments of AI technologies, which should address ethics, human rights, safety and data protection, throughout the life-cycle of an AI system.*

In line with data protection impact assessment, Section 12 of the Regulation stipulates that;

- Subsection (1): Where the collection or processing of personal data poses a high risk to the rights and freedoms of natural persons, the data collector, data processor or data controller shall, prior to the collection or processing, carry out an assessment of the impact of the envisaged collection or processing operations on the protection of personal data.
- Subsection (2): Every data protection impact assessment shall include (a) a systematic description of the envisaged processing and the purposes of the processing; (b) an assessment of the risks to personal data and the measures to address the risks; and (c) any other matter the Office may require.

- Subsection (3): The Office shall establish and make public a list of the processing operations which are subject to the requirement for a data protection impact assessment under subregulation (1).

Recommendation 7. *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.*

The Regulation has no clause on ownership nor benefit sharing of the outcomes. However, Section 26 of the Act stipulates that;

- Subsection (1): A data subject may by notice in writing to a data controller, require the data controller to stop processing his or her personal data for purposes of direct marketing.
- Subsection (3): Subject to subsection (1) a data subject may enter into agreement with a data controller for purposes of using or processing his or her personal data for pecuniary benefits.

Recommendation 8. *Governments should consider adopting models of co-regulation with the private sector to understand an AI technology, without limiting independent regulatory oversight. Governments should also consider building their internal capacity to effectively regulate companies that deploy AI technologies and improve the transparency of a company's relevant operations.*

In regards to data protection even for AI technologies, Section 6 of the Regulation stipulates the power of the Office to cooperate with other authorities.

- Subsection (1): The Office shall cooperate with other government ministries, departments and agencies in the implementation of the Act and regulations.
- Subsection (2): For the purpose of subregulation (1), all ministries, departments and agencies of government shall accord to the Office such assistance as may be necessary to ensure proper discharge of the functions.

Conclusion and recommendation

Principally, the rules and principles of the Act and the Regulation apply both in the phase of AI research and development and with regard to its use for analysing and decision-making about individuals. They contain important rights for data subjects relating to any processing of their personal data as well as obligations of processors, which will shape the way AI will be developed and applied. However, because AI, in a manner analogous to Big Data, presents a challenge for the application of traditional data processing principles, there is need to necessitate the elaboration of new applicative solutions to safeguard informational privacy and other fundamental right.

References

1. Guidance W. Ethics and Governance of Artificial Intelligence for Health. World Health Organization. 2021.
2. Naik N, Hameed BZ, Shetty DK, Swain D, Shah M, Paul R, et al. Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility? Frontiers in Surgery. 2022;9.
3. Väänänen A, Haataja K, Vehviläinen-Julkunen K, Toivanen P. AI in healthcare: A narrative review. F1000Research. 2021;10(6):6.
4. Sallstrom L, Morris O, Mehta H. Artificial intelligence in Africa's healthcare: ethical considerations. ORF Issue Brief. 2019(312).
5. Uganda U. Uganda National Council for Science and Technology (UNCST) Act (CAP 209). Kampala: UPPC. 1990.
6. Poor UPaPFft. Uganda Launches Data Protection & Privacy Portal to Streamline Management of Personal Data 2022.

Theme 5: Issues associated with research ethics frameworks and ethics review

Governance paper: Recommendations for the development of ethical guidelines for AI-related health research in Egypt

Ahmed Samir Abdelhafiz, National Cancer Institute, Cairo University, Egypt

Introduction

In November 2019, the National Council for Artificial Intelligence (AI) was founded by the Egyptian government as a partnership between the government, academic institutions, and leading companies in the field of AI. The AI strategy enablers included, among others, the development of governance to monitor the execution of the strategy, laws, and regulations, ethical principles, and guidelines. Healthcare was among the priority sectors where the government was looking for partners to develop AI solutions for triage and diagnosis of diseases, smart management of healthcare data, mental health, and bioinformatics[1].

A personal data protection law was issued by the Egyptian parliament in 2020[2]. Another law for clinical research was approved in the same year[3]. However, both laws are specific for certain aspects and don't cover several ethical, legal, and social (ELSI) issues related to AI, especially in the field of healthcare research. While the data protection law is not related to medical research, the clinical research law is mainly related to clinical research, including clinical trials, and doesn't cover many aspects of pre-clinical research[4].

Some ELSI issues are common in different types of medical research, including AI research. However, some challenges are more related to the nature of AI research. These challenges could arise before, during and after conducting this research. First: Knowledge about the nature of AI research is not known among Institutional Review Board members. While these members are familiar with traditional types of research, this could not be the case with AI research, which could affect their decision. The informed consent model to be used with participants in AI research represents another challenge. Third: During and after the research process, questions about data protection, confidentiality, and privacy also arise. While these issues represent a concern in all types of medical research, the risks associated with privacy could be different or greater in AI related research for different reasons. Since AI research usually involves the partnership of many new private bodies, they may not stick to the privacy roles followed in academia or pharmaceutical companies. Moreover, AI research itself is associated with high risk of privacy breaches through AI-driven methods, where the process of anonymization could be compromised or fail using AI algorithms[5]. Fourth: When research is done, issues related to fairness and equitable benefit sharing could arise, both at the macro-level (sharing benefits between governments, companies and research institutions), and at the individual level (return of results). This governance paper aims at identifying some ethical issues that may arise during AI related health research on humans, and to provide recommendations to deal with these issues. Three issues/themes represent the concern of this paper, namely; informed consent, commercialization and benefit-sharing and institutional review boards. Each issue will be discussed in brief, followed by providing some recommendations based on the knowledge of the author about the local laws and regulations. These recommendations could guide researchers, ethicists, and policymakers to develop guidelines or laws to govern this research in the future in Egypt.

The informed consent

The traditional concept of informed consent is challenged in AI-related research for many reasons. First: it is difficult to predict who will have access to the data in the future, and in what type of research this data will be used. Second: the nature of research and expected outputs could be too difficult for patients/participants and doctors to understand and sometimes for researchers to explain[6-8]. Third: Growth of the use of AI based health applications, including those used for assessment of symptoms, to improve compliance to treatment, or to guide healthy decisions such

as diet, raises new ethical concerns. Before using these apps, users agree to terms of participation, and in many cases may not read them at all. This agreement is different from the traditional informed consent used before data collection for treatment or research. What kind of informed consent should be used if research is conducted using data collected through these applications?[8] Fourth: How will secondary use or analysis of data collected for another purpose be explained in the informed consent, especially if data is stored in large accessible databases? Finally: The risks related to privacy, including potential re-identification risk could be greater in this type of research. This risk should be explained in detail to potential participants during discussion of informed consent. In Egypt, a previous study showed that among Egyptians who were willing to participate in medical research, many preferred a consent that gives them control over the selection of diseases in which their samples will be used for research[9]. These preferences should be taken into consideration during the discussion of the consent models to be used in AI research.

Recommendations: Research should be carried out to understand participants' attitude and preferences regarding AI research, including the type of informed consent that can be used in this type of research. Experts in the field should also discuss this issue with Institutional Review Board members and other concerned parties to select appropriate consent models. Although it is premature to recommend a mode for consent since little is known about which model best aligns with public preference and scientific practicability, we have some primary suggestions for selection among the available consent models. Specific consent that explains all details about proposed research is a suitable option. If research will include multiple future studies, other options should be considered. Among these options, we think that the use of broad consent, which is implemented in biobanks in Egypt and other countries for future use of samples and data, is difficult for use in AI related research due to the unknown nature of future research in this research[6]. Tiered or dynamic consent models, or a mix of both, could be preferred options for use in this AI research. However, the use of dynamic consent is restricted by the limited access to the internet in some regions of Egypt. Interestingly, there is a significant growth in the number of internet users in Egypt[10], which could support the use of this type of consent in the future.

Commercialization and benefit-sharing

AI research requires the use of advanced technology and sometimes "supercomputers". This necessitates collaboration with technology partners and private sector companies which can support this research[1]. Such collaboration raises issues about the commercialization of data and benefit-sharing. While commercial entities usually look for profits, governments and academic institutions look for other types of benefits as well. These include, among others, capacity building, authorship over scientific publications, sharing in patents and intellectual property rights, and getting final research products (such as new technologies) at affordable prices. These issues could be even more complicated if collaboration is done between researchers from high income countries or international companies or on one side and researchers from low or low middle income countries on the other side[11].

Recommendations: Material and data transfer agreements used in biobanks for sample and data sharing are interesting examples for documents that explain the rights of each involved party in this scientific collaboration. This model can be adopted in other types of research that includes benefit sharing. In Egypt, we think that guidelines, regulations, and policies regarding commercialization and benefit sharing should be developed and implemented. Since the government proposes binding law and regulations, usually look for the interests of the local community, and has better negotiation power than individual researchers or institutions when it comes to the discussion about benefits, we think that the government and its representatives can negotiate benefit sharing with technology partners to reach the best deals with them in this aspect. Alternatively, the government may set the boundaries of acceptable practice, and specific terms can be negotiated by the involved parties.

Institutional Review Boards (IRBs)

There is no available data about the knowledge of Institutional Review Board members in Egypt about AI and ethics related to its implementation in health research. This novel type of research could represent a specific challenge for IRB members. It has been reported that the lack of national research ethics guidelines and the need for training of IRB members in research ethics were among the challenges faced by members of ethics committees in Egypt[11].

Recommendation: Evaluation of the knowledge, perceptions and attitude of IRB members about this type of research is necessary. Based on the results of this assessment, proper education and training about issues and challenges associated with this novel type of research are needed. The Supreme Council for the Review of Clinical Research Ethics established based on the clinical research law[3] and The Egyptian Network of Research Ethics Committees[12] can play a major role in coordinating these activities. Additionally, experts in the field should be available to explain some technical aspects of this research to IRB members upon request, which could help them to make an informed decision about research proposals submitted to them.

Conclusions and recommendations

Egypt is trying to improve its capacity building capabilities in the field of AI. As science leaps forward, ethics should not lag behind! Many ethical committees have been working in different academic institutions for years. However, many of them have not been updated with new types of research that go beyond traditional medical research. In general, Egyptians are a bit skeptical about certain types of research and collaborations (e.g. genetic research and collaboration with western countries and commercial entities). Moreover, many of them think that medical research be conducted under some level of government oversight[9,13]. The development of laws, guidelines and recommendations to support AI health research is necessary to equip IRB members with tools to monitor this type of research, and to preserve the rights of the involved parties. Training of IRB members on how to use these tools is essential to allow them to create a balance between advancing medical research and the protection of the community. Encouraging and supporting AI research in the medical field will allow real-time detection of problems and discussions to find solutions for them.

References

1. The Egyptian Ministry of Communications and Information Technology. Artificial Intelligence page. Available at: https://mcit.gov.eg/en/Artificial_Intelligence. Accessed 29th of October2022
2. Egyptian Court of Cassation. Data Protection law. Available at: https://www.cc.gov.eg/legislation_single?id=404869. Accessed 29th of October2022
3. Egyptian Court of Cassation. Clinical research law. Available at: https://www.cc.gov.eg/i/l/407276.pdf. Accessed 29th of October2022
4. Matar A, Silverman HJ. Ethical Analysis of Egypt's Law Regulating Clinical Research. *Journal of Empirical Research on Human Research Ethics*. 2022;17(4):494-503. Doi:10.1177/15562646221096188
5. Murdoch, B. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Med Ethics* 22, 122 (2021). https://doi.org/10.1186/s12910-021-00687-3
6. Blasimme, Alessandro and Vayena, Effy, The Ethics of AI in Biomedical Research, Patient Care and Public Health (April 9, 2019). Oxford Handbook of Ethics of Artificial Intelligence, Forthcoming, Available at SSRN: https://ssrn.com/abstract=3368756 or http://dx.doi.org/10.2139/ssrn.3368756
7. Naik N, Hameed BMZ, Shetty DK, et al. Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility? *Front Surg*. 2022;9:862322. Published 2022 Mar 14. Doi:10.3389/fsurg.2022.862322
8. Gerke, S., Minssen, T., & Cohen, G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. *Artificial Intelligence in Healthcare*, 295–336. https://doi.org/10.1016/B978-0-12-818438-7.00012-5

9. Abou-Zeid A, Silverman H, Shehata M, et al.: Collection, storage and use of blood samples for future research: views of Egyptian patients expressed in a cross- sectional survey. *J Med Ethics*. 2010;36(9):539–47. 10.1136/jme.2009.033100
10. Dataportal. DIGITAL 2021: EGYPT. Available at: shorturl.at/inAQ3. Accessed 29th of October 2022
11. Sleem H, El-Kamary SS, Silverman HJ: Identifying structures, processes, resources and needs of research ethics committees in Egypt. *BMC Med Ethics*. 2010;11:12. 10.1186/1472-6939-11-12
12. The Egyptian Network of Research Ethics Committees website. Available at: http://www.enrec.org/. Accessed 29th of October2022
13. Abdelhafiz AS, Sultan EA, Ziady HH, et al.: What Egyptians think. Knowledge, attitude, and opinions of Egyptian patients towards biobanking issues. *BMC Med Ethics*. 2019;20(1):57. 10.1186/s12910-019-0394

Governance paper: The 'proverbial' black box that is ethics of AI in global health research: are Kenyan RECs well equipped?

Brenda Odero (presenter), Strathmore University, Nairobi, Kenya; South African Research Ethics Training Initiative (SARETI), University of KwaZulu-Natal, South Africa
David Nderitu, Egerton University, Njoro, Kenya

Brief description of the context
The lack of adequate regulatory and policy frameworks for integrating artificial intelligence (AI) in research in low and middle-income countries (LMICs) raises ethical and legal issues. These issues include discrimination caused by algorithmic bias[1], lack of transparency when seeking informed consent[2] and emerging high risks that may cause harm to participants[3]. This is coupled with the inadequate capacity of research ethics committees (RECs) to review research protocols involving AI in health and the lack of knowledge from researched communities. The lack of legal and ethical frameworks in regulating the use of AI in health leads to the need for accountability not by the machine, but by the people who built it and the protection of those who use it[4].

In Kenya, the Data Protection Act (DPA)[5] makes provision for 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 (NGEBMR) (2020)[6], 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 (NACOSTI) (2021), demonstrates significant optimism in regulating the use of digital technology, including AI in science, technology, and innovation in Kenya. However, all these regulations and guidelines give minimal attention directly to the use of AI in research. Simultaneously, some researchers in Kenya have been using AI for health in diagnosis, storage of electronic medical records, disease outbreaks, surveillance, health policy and planning[7]. Furthermore, the role of ethical review in AI research is poorly examined or explained in the DPA. This paper examines the gaps in the DPA and NGEBMR as a resource for reviewing AI in global health research in Kenya. It assesses the challenges RECs face when reviewing AI-based health research protocols and offers recommendations on how traditional research ethics procedures may be adapted to respond to AI-based health research.

Commentary
In terms of readiness to uptake AI, Kenya ranks third in Sub-Sahara Africa with a score of 45.5% according to the Government Artificial Intelligence Readiness Index 2021 by Oxford Insights. The index was defined by three pillars: government, technology, data and infrastructure. The government pillar incorporates governance and ethics, digital capacity, and technological adaptability. Governance and ethics are understood to include indicators related to data protection and privacy legislation, cyber security, the national ethics framework, legal frameworks, and business models. Under this domain, the Readiness Index sought to answer the question, are there regulations and ethical frameworks in place to implement AI in a way that builds trust and legitimacy?

NACOSTI, the body that accredits and offers training to all RECs in Kenya, has not adopted the mandatory inclusion of members with expertise in AI[8]. Hence there is very little knowledge in the review of ethics in AI.

Traditional research ethics procedures may apply in sections 28–30[5] of the DPA on the collection of data. This section permits the collection of personal data through consent by the data participant only. However, there is a need to specify in what context the data is being collected from the data participant and how it is being used and stored. Whether it is for marketing, healthcare, or business. This is because some data are more sensitive than others, especially in the health sector. Section 35 states that a data participant should always offer consent on any automation of their data being processed or profiled. Additionally, section (35) (3) (a) (b) states that the data controller/processor should notify the data participant of the use of any data collected. Unfortunately, it is not explicit

how informed consent from the data participants before data automation, profiling and general collection of data will be sought during research in AI.

Furthermore, the focus on the prevention of ethics dumping in Kenya is solely based on biological samples in health research isolating regulation of potential ethical dumping of AI health data. There are no clear indications or procedures in sections (48), and (49) that allow personal data transfer and safeguards outside Kenya. There is no clear indication of continuous respect for the data participant on what would happen to their data if it was continuously used, and it may be discarded. An adequate regulatory and policy framework would ensure the ethical soundness of research involving AI. Moreover, it would increase the capacity of RECs to review such research and, ultimately, create awareness in communities under research in Kenya.

Recommendations

Effective AI research models should be built on law, policy, and ethical guidelines. This model will incorporate ethical guidelines and review processes to discover, assess, or track the impact of AI health research. We propose the adoption of the Emanuel *et al*. framework[9] and some insights into the traditional Kenya National ethical guidelines for biomedical research in interpreting and reviewing AI-based health research. This is as follows:

Community engagement

While there is a need to consider and commend AI's attention around and positivity in global health research, most of its adoption is centred around the European context. This excludes the African context, causing exclusion and discrimination in the communities' use of AI in healthcare. AI research needs to incorporate Africa's communitarian philosophies of making decisions together and a sense of inclusion in the changes we face. Community consultations to gather feedback and offer public access to the understanding of Algorithmic Impact Assessment (AIA) procedures will offer accountability and build trust in the community[10]. Such assessment is done to assess the possible societal impacts of the AI system before it is used. RECs should ensure that a human-centred technology design for AI health research is incorporated that includes the community's desires and public concerns. This form of engagement should be done early and continued throughout the study.

Collaborative partnerships

It would be commendable to establish *ad hoc* Committees on AI to conduct inclusive multi-stakeholder consultations to determine the feasibility and potential elements of a legal framework for the design and application of AI according to Kenyan law[11]. RECs in Kenya must intensively engage research communities, the government, experts and other stakeholders in digital innovations in the policy review process. RECs should also ensure that all stakeholders submit data-sharing agreements for review to ensure that data rights are understood and respected.

Social value

RECs should ensure that all research involving AI in health fits into the Kenyan context while identifying who the beneficiaries are. This is because different counties in Kenya have different health needs and finite resources. It could also ensure that if a vulnerable population is included, there should be justification in the study. RECs should ensure that researchers clearly state how they will disseminate and share potential benefits with the community. This can be done through an algorithmic impact evaluation report where researchers give monthly progress reports on the impact of the use of AI in healthcare can be looked up when in use.

Scientific validity

Data is the nourishment that AI algorithms should survive and thrive[12]. However, there needs to be justification and considerations as to why AI is required for a study. The choice of study design and procedures must be rigorously reviewed to yield valid and reliable data. The REC must ensure that research participants' healthcare interventions align with the scientific objectives in the proposal. This is because sometimes the interventions to which the research participants are entitled may go beyond what is sustainable or feasible for the study objectives.

Fair selection of study population
To ensure the interests of research participants are defended, RECs can offer a checklist of non-bias and non-discriminatory guidelines. This would entail selection based on scientific importance and not convenience. RECs looking at the research on AI systems in the Kenyan context should review the inclusivity and diversity of the data to be collected. By and large existent research AI models are predominantly designed using western epistemologies and worldviews, which can limit their applicability to African contexts[13]. Moreover, the use of western languages, graphics, and aesthetics inherent in AI designs may contradict the Kenyan reality[14,15] especially in clinical research.

Favourable risk-benefit ratio
There is a need for Kenyan RECs to observe potential harms that may stem from inadequate benefits or high risks of AI in the community. The REC should weigh whether an AI-led study's risks, burdens, or benefits are needed. This is because research participants in the Kenyan context have a higher chance of facing algorithmic bias, stigma, and physical or psychological harm.

Independent review
When reviewing AI health research proposals, an ad hoc AI consultancy expert should always be brought in. For independent oversight to occur, training should be done, and national guidelines dedicated to AI health research should be created. Reviews by RECs should be independent of public or private deployers of AI, equipped with interdisciplinary expertise and training. This also includes monitoring and evaluation of risk assessment and non-infringement of human rights[16]. There need to be adequate standard operating procedures that clearly show the impact assessment procedure adapted to respond to AI-based health research and its impact on the community before it is approved. Lastly, the NGEBMR should ensure basic training in AI for all ethics review members.

Informed consent
The goal of informed consent is to generally respect the participants' and communities' decision-making interests. As suggested above, working with various stakeholders, especially the research community, could reveal new aspects and levels of effective consenting, risks, and privacy issues in research involving AI. RECs should draft transparency requirements for AI health-based research proposals. Consent on all AI systems should not be a 'terms and conditions' jargon document. A short, simple write-up to aid the understanding of the participant should be required for review by RECs[17]. We recommend that RECs informed consent forms for AI use in global health research closely resemble user agreements where necessary[10].

Ongoing respect of participants and communities
To ensure health equity while still ensuring ongoing respect for the community, when using AI, monitoring and evaluation are required. Every research relating to AI in global health should have a framework for publishing and sharing with the public. This will allow an effective monitoring and evaluation system and reduce research fatigue of the same study in a community. Here, we recommend RECs create an AIA model template that focuses on the impacts of AI and whether they will be ongoing, reversible, short-term, or perpetual[18].

Conclusion
In Kenya, AI readiness is at its peak, but its regulation unfortunately is lacking. The recommendations made above may inform RECs' ethics review of AI-based global health research. We hope that these recommendations may be used to inform law and policy in the future.

References

1. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399.
2. Andreotta, A.J., Kirkham, N & Rizzi, M. (2021). AI, big data, and the future of consent. Springer Nature: AI & Society. https://doi.org/10.1007/s00146-021-01262-5
3. Nicholson Price II, W. (2019). Risks and remedies for artificial intelligence in health care. Brookings: AI Governance, Thursday, November 14, 2019
4. Naik, N., Hameed, B. Z., Shetty, D. K., Swain, D., Shah, M., Paul, R., ... & Somani, B. K. (2022). Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility? Frontiers in Surgery, 9.
5. DPA (2019). The Data Protection Act. [online] https://www.odpc.go.ke/dpa-act/#
6. NACOSTI (2020). National Commission for Science, Technology, and Innovation Guidelines for Accreditation of Institutional Ethics Review Committees in Kenya (NGEBMR), January 2020.
7. Wahl, B., Cossy-Gantner, A., Germann, S., & Schwalbe, N. R. (2018). Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? BMJ global health, 3(4), e000798.
8. NACOSTI. (2017). National Commission for Science, Technology, and Innovation Guidelines for Accreditation of Institutional Ethics Review Committees in Kenya October 2017. October 1–14.
9. 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.
10. Gerke, S., Minssen, T., & Cohen, G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. In Artificial intelligence in healthcare (pp. 295-336). Academic Press.
11. WHO (2021). Ethics and Governance of Artificial Intelligence for Health: WHO guidance. In World Health Organization. http://apps.who.int/bookorders.
12. Ormond, E. O. (2020). The Ghost in the Machine: The Ethical Risks of AI. Papers.Ssrn.Com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3719745
13. Miller K. (2022). The Movement to Decolonize AI: Centering Dignity Over Dependency. Ethics and Justice, March 21, 2022, https://hai.stanford.edu/news/movement-decolonize-ai-centering-dignity-over-dependency
14. Bossmann, J. (2016). Top 9 ethical issues in artificial intelligence. World Economic Forum, 21 Oct 2016 https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/
15. Nwaodike, C. (2020). The AI Digital Divide: An African Perspective. Internews Commentary, March 17, 2020. https://internews.org/commentary/ai-digital-divide-african-perspective/
16. de Europa, C. (2019). Unboxing artificial intelligence: 10 steps to protect human rights. Recuperado desde https://www. com. int/en/web/commissioner/-/unboxing-artificial-intelligence-1 0-steps-to-protect-human-rights.
17. Amrita Sengupta, 2020 Algorithmic Consent: Why Informed Consent Matters in An Age of Artificial Intelligence. https://www.capgemini.com/insights/expert-perspectives/algorithmic-consent-why-informed-consent-matters-in-an-age-of-artificial-intelligence/
18. Government of Canada. (2020). Directive on Automated Decision-Making. Available at: https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592

Governance paper: Reframing research ethics frameworks to include environmental sustainability

Gabrielle Samuel, King's College London; Oxford University, UK

Brief description of the context

This paper addresses a specific governance issue associated with research ethics frameworks that are currently used for artificial intelligence (AI) research - their lack of a normative consideration for the adverse environmental impacts associated with AI research endeavours. In the Commentary section, this paper makes a case for why environmental considerations are important to include in an AI research ethics framework. In the Recommendations section, it draws on the commonly used international research ethics framework proposed by Emanuel et al. (2008)[1] to propose what such a framework could look like.

Commentary

Dominant research ethics paradigms have historically revolved around ethics principles that are concerned with the protection, rights, safety, and welfare of individual research participants. Strong criticism has long existed about the appropriateness of placing individual risk at the focus of research ethics frameworks. Much of this criticism has pointed to the need to consider *communitarianism*[2-5], i.e., the need to consider the moral status of the community in research ethics considerations. Community harms are viewed as more than the sum of individual values and interests and relate to questions associated with whether communities will be beneficiaries of the research, or even whether they share the same goals as the researchers[6, 7]. For example, Tsosie et al. (2019) argue that in genomics research, individualising risk dismisses a deeper examination of the systemic barriers to health that are imposed on minorities, and by doing so, collective health status is overlooked[8]. In AI research specifically, the community is a central consideration for research ethics because many of the potential harms that can come from AI research are likely to be group-based harms. Consider, for example, how an AI algorithm to detect skin cancer was shown to have been optimised for fair skin, being less able to detect Melanoma on darker skin[9]. To address these concerns, Emanuel and Weijer (2005) have emphasised an ethical principle of 'respect for community' to sit alongside other more individually focused ethical principles. This requires researchers to devote attention to understanding the socio-political impact of research on communities as a whole and not only on individuals[6].

While considerations of community harm have expanded the moral status considerations of research ethics frameworks beyond those focused on individual risk alone, many (though not all) have stopped short of considering the adverse environmental (and consequential human health) harms generated from the manufacturing, use, and waste disposal of equipment, tools, and technologies associated with research. In the AI research field, these adverse environmental/health harms are associated with the large amounts of electricity consumed to power and cool equipment in data centres – the large warehouse scale buildings where the data that underpins the digital revolution and AI methodologies is located. They are also associated with the electricity needed to power the training of algorithms being developed during health-related research: some of these algorithms are particularly energy hungry (for example, the training of one particular AI algorithm has been calculated to be equivalent to the energy needed to power a trans-American flight[10]). This electricity consumption contributes towards climate change when fuelled by non-renewable energy sources, and climate change is characterised by both environmental and human health harms. These harms are becoming particularly acute in lower-to-middle-income countries where there are less resources to help communities to withstand extreme climate effects.[11] Furthermore, in lower-to-middle income countries, where electricity supply is relatively unstable, the electricity demand could mean that local communities may experience harms due to further exacerbation of brownouts and/or black-outs. While the converse could be true–that the investment in data centres improve electricity infrastructure for local communities, ethnographic research has questioned the promise of data centres to bring benefit to communities, and there have been reports of, for example, data centres drawing resources away from farmers in areas of low water supply.[12]

Other adverse environmental/health harms are associated with the extraction of minerals necessary to manufacture digital technologies upon which data can be stored and processed, which can adversely impact biodiversity in local mining areas. 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 and therefore raise environmental justice issues. Further along the product pathway, manufacturing digital objects produces toxic emissions that can leach into local environments, also posing health issues. Finally, obsoleteness is a concern: AI and digital research often need to run on the most up-to-date software, meaning that digital servers need relatively frequent replacing. Many digital objects are not recycled formally, and often end their lives in electronic-waste (e-waste) dumps in lower-to-middle income countries (possibly after secondary use (or not)). Individuals and families come to these dumps to make a living because they can extract precious minerals for re-sell. However, environmental concerns have been raised because doing so requires the use, or leads to the leaching of, toxic (including many carcinogenic) chemicals that have been shown to now be present in these landfills in dosages far above those recommended[13].

Although some scholars expect that the continuing efficiency improvements in digital technologies will address many of these concerns, others expect efficiency improvements to lead to consumption increases rather than decreases. This has been a historical pattern known as a rebound effect[14]. Given these environmental impacts, 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. This is now something that is recognised in the AI ethics community and there have been calls to consider these environmental impacts in decision-making. However, these issues have not been discussed within the research ethics literature. This is a concern: health research has a special interest in addressing environmental impacts, not only as a matter of international priority, but also as a commitment to health[15]. In the below section, this is addressed through modifying specific, relevant aspects of Emanuel et al.'s (2008)[1] international research ethics framework.

Recommendations

A research ethics framework that includes considerations associated with the adverse environmental impacts of AI research endeavours requires modification of the following substantive principles in Emanuel et al.'s (2008)[1] research ethics framework:

> -*Social value*. Health research must have a reasonable potential to benefit participants, community, and/or society. Consideration must also be given to potential harms/benefits (mentioned in the previous framework), including to the environment (not made explicit previously). Research that promises potential health benefit to a small number of individuals/communities, but which does not consider how this benefit will be accessible to all, nor how the adverse risks associated with this benefit – such as those towards the environment – have been considered, should not be considered as having social or environmental value.
>
> *-Respect for persons, communities, and environment* (stated as 'respect for participants' and 'community partnership' in previous framework). For AI health research, 'respect for persons and communities' entails respecting all of those affected by the research. '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'). For AI model development and training, fair collection, storage, use, linkage, and sharing of data is vital. Researchers must also 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?). Furthermore, attention should be focused on benefit sharing of research outcomes.

-*Fair consideration of those affected by the research process* (previously 'fair participant selection' - additional recommendation). Consideration must 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. This is particularly the case because those individuals and communities most affected by the adverse environmental and health impacts of research are the least likely to benefit from any potential health benefits that may or may not arise from the research, meaning that there is an inequitable burden of adverse research outcomes.

-*Favourable risk/benefit ratio.* Risk benefit considerations for AI research need to go beyond including those affected by partaking in the research and/or affected by the research outcomes (as previously stated in the framework), but also those affected by the manufacture of digital products used during the AI research process, and the subsequent disposal of digital research products and e-waste.

It is proposed that these adaptations to the Emanuel et al. framework can and should be applied by all researchers, research ethics committees that review AI (health) research, and those that shape the research policy agenda more broadly. However, the context of these principles will vary dependent on each of the practices:

-*For researchers and research ethics committees.* Attention should be paid to where data is going to be stored, with the use of differential storage of data (long and short latency times) to reduce energy costs where possible. Algorithms must be optimised for environmental considerations. Considerations of obsoleteness require new computers to be bought only when necessary and, where possible (institution permitting) these should be repurposed. A recycling plan should be put in place for the research. See Lannelongue (2021) for more in-depth guidelines[16].

-*For research policymakers.* Policymakers must not *solely* rely on the increasing efficiency of digital technologies to reduce the adverse environmental impacts associated with digital technologies. Rather, they must put constraints in place to ensure that as efficiencies improve, consumption does not increase. This could be achieved by constraining the level of resources provided to AI researchers. Resources could be shared more equally with those research proposals that use methodologies that have lower environmental costs. Such research often focuses on addressing the social/political/economic determinants of health, which, if addressed, have been shown to lead to more significant positive population health outcomes compared to those produced through clinical medicine.

There are limitations to implementing such a framework, including the incomplete data associated with changing practices to address specific environmental impacts, which is compounded by the often lack of transparency from private data storage and processing companies, or their incomplete knowledge. Nevertheless, the above changes could be implemented without this evaluative data with the driving goal of reducing consumption.

References

1. 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.
2. Buchanan DR, Miller FG. A public health perspective on research ethics. *Journal of medical ethics* 2006;32(12):729-33.
3. Adashi EY, Walters LB, Menikoff JA. The Belmont Report at 40: Reckoning With Time. *Am J Public Health* 2018;108(10):1345-8.
4. Gbadegesin S, Wendler D. Protecting communities in health research from exploitation. *Bioethics* 2006;20(5):248-53.
5. Cragoe NG. Oversight: Community vulnerabilities in the blind spot of research ethics. *Research Ethics* 2017;15(2):1-15.

6. Emanuel EJ, 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.
7. Friesen P, Kearns L, Redman B, Caplan AL. Rethinking the Belmont Report? *Am J Bioeth* 2017;17(7):15-21.
8. Tsosie KS, Yracheta JM, Dickenson D. Overvaluing individual consent ignores risks to tribal participants. *Nat Rev Genet* 2019;20(9):497-8.
9. Lashbrook A. AI-Driven Dermatology Could Leave Dark-Skinned Patients Behind *The Atlantic* 2018 August 16.
10. Strubell E, Ganesh A, McCallum A. Energy and Policy Considerations for Deep Learning in NLP. In: Linguistics AfC, editor. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*; 2019; Florence, Italy; 2019. p. 3645–50.
11. IPCC. *Sixth Assessment Report of the Intergovernmental panel on climate change*: Intergovernmental panel on climate change; 2022.
12. Solon O. Drought-stricken communities push back against data centers. *NBC news* 2011.
13. Ngo HTT, Watchalayann P, Nguyen DB, Doan HN, Liang L. Environmental health risk assessment of heavy metal exposure among children living in an informal e-waste processing village in Viet Nam. *Science of The Total Environment* 2021;763:142982.
14. Hilty LM, Köhler A, Von Schéele F, Zah R, Ruddy T. Rebound effects of progress in information technology. *Poiesis & Praxis* 2006;4(1):19-38.
15. Samuel G, Richie C. Reimagining research ethics to include environmental sustainability: a principled approach, including a case study of data-driven health research. *Journal of Medical Ethics* 2022:jme-2022-108489.
16. Lannelongue L, Grealey J, Bateman A, Inouye M. Ten simple rules to make your computing more environmentally sustainable. *PLoS computational biology* 2021;17(9):e1009324-e.

“Pecha Kucha” presentations

“Pecha Kucha” translates from Japanese roughly as “chit-chat”. Pecha Kucha presentations are designed to be delivered quickly and concisely, with slides **automatically** advancing every 20 seconds. They are an informal opportunity for GFBR participants to find out about each other’s research, viewpoints and experience.

The format does not allow for questions at the end of each presentation, but you are welcome to discuss the presentations after the session or during breaks.

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

1. Analysing a local imbalance of power ethics: University of Ghana vs. Data Commission

Athanasius Egyarkoh Afful, University of Ghana, Ghana

Data is power; and whoever, whatever, controls data these days, wields power, or too much power. In conversations about the ethics of power relative to the governance of data, we give considerable attention to international imbalances of power, and how states, multinational organisations and peoples on opposite sides of the technology divide either exploit, or are exploited. I present, however, a case of a local or domestic imbalance of data power, and how it affects the conduct of AI-dependent bio-research at a local level. The domestic case I present on is from Ghana; and I present it as two tensions:

- Community engagement
- Institutional control

In October, 2021, the University of Ghana's School of Public Health, in collaboration with the International Food Policy Research Institute (IFPRI) advertised, and subsequently commenced the Accra Urban Adolescent Nutrition Study. The objective of the study was to describe the nutritional status, dietary intake, physical activity patterns, and food environment of about 1000 adolescents aged 12-19 years, from low- and middle-class households in 10 selected areas in the Greater Accra Region[1]. A major component of the study required participants to wear a GPS/accelerometer-integrated belt to collect data on their physical activity and movement patterns for the period of the study, except when they were sleeping, after which their blood samples would be taken for analysis. The study had been approved by the Noguchi memorial Institute for Medical Research's Institutional Review Board, the Ghana Health Service, and the Ghana Education Service. Written informed consent and assent had also been obtained from both the parents/caregivers of the participants, and the participants themselves, prior to participation.

On the 8th of June, 2022, the Executive Director/Commissioner of the Data Protection Commission of Ghana, in a press statement, directed the University of Ghana to among other things, halt, with immediate effect, the said study, and conduct a Data Protection Impact Assessment (DPIA) to ascertain potential data risks which could have arisen from the conduct of the study since its commencement. The reason for the injunction? Some parents/guardians and teachers had raised concerns over the intrusive and sensitive nature of the data to be collected with the AI devices, and how these would be processed. The Commission investigated the concerns, and determined that the University of Ghana was not registered with the Commission as a data controller, and that it had breached a contract agreement with its sponsors (IFPRI), which required partners to comply with data regulations in their home countries.

The posture of the Data Commission, which has a mandate to "protect the privacy of the individual and personal data by regulating the processing of personal information", and "provide the process to obtain, hold and use or disclose personal information"[2] has extensive powers to make the administrative arrangements it considers appropriate for the discharge of its duties, and investigate any complaint and determine it in the manner it considers fair. The DC's powers put it in a position of uncontrolled control over other data handlers, including the University of Ghana, and leave a lot of room for arbitrary government interventions and intrusions in the name of 'appropriate' and 'fair', with negative consequences, sometimes, for the conduct of bio-data-dependent research.

The concerns raised by some other parents/caregivers and teachers highlights a question of the ethics of community engagement: How much of community engagement should be done, the extent and limits in the definition of 'community', and what communities should be engaged for the collection of AI data for research.

References

1. Press briefing, University of Ghana https://www.ug.edu.gh/sites/default/files/documents/Statement%20on%20the%20Accra%20Urban%20Adolescent%20Nutrition%20Study_revisedc.pdf
2. Data Protection Act, 2012. Act 843

2. Ethical concerns in the use of AI in patient safety research: an examination of the adequacy of Nigerian laws

Dorcas A. Akinpelu (presenter), Simisola O. Akintola. Faculty of Law, University of Ibadan, Nigeria

Brief description of the context

Available literature reveals that artificial intelligence (AI) could be useful in researching to the causes of medical errors and developing effective contextual solutions to improve health outcomes and patient safety particularly in low and middle income countries such as Nigeria, where there are not enough skilled medical professionals to care for the rapidly growing population. Nevertheless, using AI to improve patient safety raises ethical concerns and complexities such as algorithmic fairness and biases, erosion of confidentiality, breach of privacy and data security, complications around obtaining informed consent, assignment of responsibility and liability, among others. These concerns are fundamental and require careful consideration and attention; non-consideration of these concerns can also impact patient safety in an adverse manner. Although Nigeria has a legal framework for health research, the extent to which it addresses ethical concerns in the use of AI for patient safety research is unclear.

Ethical concerns in AI-assisted patient safety research

Patient safety research, like other types of health-care research, is fraught with ethical issues. These issues are exacerbated when AI is used as a research tool because AI has its own set of ethical implications, some of which will still unfold as AI evolves. In the first place, the quality of data entered into the AI system determines the predictability of the outcomes. When an AI device with underrepresented data is utilized for patient safety research, there is a possibility of bias and discrimination, which is an ethical concern that must be addressed. Also, disclosure of AI predictions about an individual, particularly to third parties, can lead to discrimination against that person.

The use of AI in patient safety research also has the potential to erode confidentiality, violate privacy and data security, and disregard autonomy, particularly when research subjects' data is made available to third parties. Addressing privacy and data security concerns is critical, and strict oversight of data use and transfer will be required to protect personal information and interests of data subjects. In addition, obtaining informed consent from data subject could be challenging because of the black-box problem arising from the novelty and technological intricacies of AI. Furthermore, the use of AI in patient safety research makes assigning responsibility and liability onerous. It's worth noting that AI forecasts are far from perfect. Also contributing to this onerousness is the opaque nature and unpredictable effects of black-box AI, as well as the problem of many hands, which obfuscates blame attribution.

Nigerian laws on the use of AI in patient safety research

Given the ethical concerns raised above, the need for governance and regulation cannot be overstated. The Constitution of the Federal Republic of Nigeria 1999, the National Health Act, the National Code of Health Research Ethics, the Cybercrimes (Prohibition, Prevention, etc.) Act 2015, and the Nigeria Data Protection Regulation of 2019 can be relied on for ethical guidance in patient safety research. A review of these laws reveals that although they provide guidance for some ethical issues in the use of AI for patient safety research, they do not adequately address the ethical concerns raised and those that may arise in the future as the use of AI for patient safety research advances. This is because they were not developed with patient safety and/or AI in mind, as these are emerging disciplines in Nigerian health research.

Recommendation and conclusion

It is highly expedient to have a framework to guide the use of AI in patient safety research. This is needed to fill the lacunae of the existing laws and to specifically address the identified ethical issues while leaving room for future issues that may arise. There should also be established a body which will oversee the activities of stakeholders in the use of AI for health research. Members of this body should be well-versed, knowledgeable and skilled in both AI and health research.

3. Who minds the machines? Developing a governance framework for pre-market authorisation of responsible AI applications in healthcare in South Africa

Irvine Sihlahla (presenter)
MBCHB (UZ) MMed (UCT) FcRAD (CMSA)
LLB (UNISA), Ph.D. Fellow, University of KwaZulu-Natal, South Africa
https://orcid.org/0000-0003-3576-6711

Dusty-Lee Donnelly
BA LLB (UND) LLM Ph.D. (UKZN)
Senior Lecturer in the School of Law, University of KwaZulu-Natal, South Africa
https://orcid.org/0000-0002-5574-7481

Artificial intelligence (AI) is a rapidly evolving set of technologies that are set to radically transform the healthcare system. AI will considerably improve workflow efficiency, and diagnostic accuracy, reduce health costs and help with the alleviation of staff shortages in low-resource settings. However, AI design, development, and deployment within the high-risk healthcare sector are fraught with ethical, legal, and human rights constraints. The constraints have imperative value within the South African context considering its history of economic inequalities, exploitation, and racial disparities. Immature regulation and governance frameworks in LMIC including South Africa further compound the constraints of AI.

This paper proposes enforceable governance reforms of the regulatory authorisation (by SAHPRA in South Africa) for the manufacture, wholesale, or marketing of AI software as a medical device (SaMD). The two key recommendations are 1) to develop ethical and human rights impact assessment tools that aim to mitigate against derogation of ethical principles and human rights and can guide researchers, AI developers, regulators, and clinicians in decisions about the design, development, and deployment of AI; and 2) the reform of the current single-stage regulatory oversight mechanism to provide for total product lifecycle regulatory oversight of AI SaMD. A key element of local regulatory approval must be satisfactory evidence (through impact assessments and post-deployment audits) that the training and validation of the AI algorithms were performed on local population data, and if this is not the case, that measures are in place to detect biased and inaccurate algorithmic outputs. Statutory established national and provincial health research ethics committees (RECs) set standards of practice, review the research protocol for patient safety, and provide guidance on human rights and ethical issues that may affect research at the design and development phase[1]. The implementation of ethical and human rights impact assessment broadly aligns with the core responsibility of the RECs during clinical trials and health research involving AI SaMD. Regulatory approval pathways, data protection laws, and REC approval processes comprise distinct legal and regulatory compliance measures. Multi-stakeholder pluralistic participation will aspire to align these frameworks and provide for total product lifecycle coverage of AI in health care through ethical and human rights impact assessment, audits, and post-deployment surveillance mechanisms. The governance reforms will be valuable to local legislators, regulatory agencies, AI developers, health researchers, research ethics committees, healthcare personnel, and ultimately patients, as the end-user, of the need for responsible human-centric AI within the healthcare system.

Reference

1. National Health Act 61 of 2003 Chapter 9; Department of Health (DOH). *South African Good Clinical Practice: Clinical Trial Guidelines* (2020) and DOH *Ethics in Health Research Principles, Processes and Structures* (2ed.)(2015)

4. Future Nanomedicines: building a regulatory framework for the first-in-human nanoswarm cancer clinical trial

Matimba Swana, University of Bristol, UK. PhD Supervisors: Prof Jonathan Ives & Dr Sabine Hauert

This paper will provide an overview of future nanomedicines and focus on what is needed to build regulations for nanoswarms in clinical trials in the UK, Europe and the US. As and when this technology becomes ready for first-in-human tests, decisions will have to made about how it can and should be safely tested.

Overview: Cancer nanomedicine

Cancer nanomedicines can be used as drug carriers that can target tumours more effectively with anti-cancer agents, while leaving normal tissues untouched. Swarm behaviour, present in social animals such as birds, ants, fish and termites, can be designed using a systems approach as *in silico* modelling is an effective tool that can minimise costly trial-and-error methods[1, 2]. Researchers can use simulations for selecting nanoparticles so drugs can more effectively reach the tumour while avoiding side effects. Advancements in nanomaterial-based approaches and artificial intelligence (AI) offer unique opportunities for researchers to go beyond nanoparticle selection. Researchers are investigating controlling the movement of nanoparticles to establish an intelligent drug delivery system. These intelligent nanoparticles or nanorobots are nano-sized entities that can control their motion and interactions with the environments[3]. Nanoswarms are multiple nanoparticles or nanorobots that can interact with each other or their environment to achieve a task (e.g. deliver chemotherapy to a tumour without killing healthy cells), exhibiting collective behaviour inspired by swarms[1, 2, 4].

Commentary: Nanoswarm classification

The classification of nanoswarms as a drug delivery system is likely to fall under the medical device category, however, this will be dependent on the application and each country's regulatory body. Medical device and drug trials have very different requirements in most countries. There is the added complexity of different regulations for software as a medical device and other guiding AI principles. Policymakers need to co-design regulations with developers, patients, healthcare workers and the public as this will be essential for innovation, development and to reap the benefits of nanoswarm technology. Nanoswarms do not generate entirely new categories of ethical issues, but they do require us to think carefully about whether our current theories and approaches can provide the guidance we need. A concern amongst researchers is that overly restrictive legislation, arising from ethical concerns, will stifle research at a time when it is increasingly needed as cancer incidence rises[3]. Additionally, nanomedicines could increase the divide between high- and middle- or low-income countries, leading to a so-called 'nano-divide'[3]. It also seems clear that nanoswarms could in principle be deployed to harm, as well as heal. This 'dual use' problem is not unique to this technology and cannot be managed by regulators alone; additional legal mechanisms must be in place.

Conclusion: Regulating the next frontier of nanomedicines

Nanotechnologies have been used in the clinic for years. For all their potential medical applications, nanoswarms are still largely in the research and development stage. The next frontier of nanomedicines in clinical trials have yet to be approved[3], but we ought to start thinking about what the first-in-human clinical trials of nanoswarms could or should look like, and ask how we will regulate the development of this new medical technology, in order to anticipate ethical controversies that may arise, and to mitigate risk[3]. We need to go beyond the usual process and discuss everything from concept, design, testing, and mode and mechanism of action, to manufacturing and waste disposal and management. To aid clinical adoption of nanoswarms, a harmonised nanomedicine vocabulary is essential, and this is a pre-requisite for an effective regulatory framework[3]. We suggest, for now, that there are 6 central domains that need to be explored in order to draft guidance for regulation in this area, these are stakeholders, social/economical, in silico, *in vitro* and *in vivo* analysis, nanomaterials, legal, and approval process[3].

References

1. N. R. Stillman, M. Kovacevic, I. Balaz, and S. Hauert, “In silico modelling of cancer nanomedicine, across scales and transport barriers,” npj Computational Materials 2020 6:1, vol. 6, no. 1, pp. 1–10, Jul. 2020, doi: 10.1038/s41524-020-00366-8.
2. N. R. Stillman et al., “Evolutionary computational platform for the automatic discovery of nanocarriers for cancer treatment,” NPJ Comput Mater, vol. 7, no. 1, p. 150, Sep. 2021, doi: 10.1038/S41524-021-00614-5.
3. M. Swana, J. Blee, N. Stillman, J. Ives, and S. Hauert, “Swarms: The next frontier for cancer nanomedicine,” in Cancer, Complexity, Computation., 1st ed., vol. 46, I. Balaz and A. Adamatzky, Eds. Emergence, Complexity, Computation Vol 46. : Springer, 2022.
4. 4 S. Hauert and S. N. Bhatia, “Mechanisms of cooperation in cancer nanomedicine: towards systems nanotechnology,” Trends Biotechnol, vol. 32, no. 9, pp. 448–455, Sep. 2014, doi: 10.1016/J.TIBTECH.2014.06.010.

5. International AI research: the issue of moral pluralism

Serene Ong, Centre for Biomedical Ethics, Yong Loo Lin School of Medicine, National University of Singapore

Context: Ethical AI frameworks regulating transnational and cross-sectoral healthcare research

Commentary

Transnational approach AI research poses a conflict between harmonisation of principles and moral pluralism[1–3]. AI research requires large datasets for greater decision making accuracy; [4,5] however, concerns around data colonialism may be worsened by uneven regulatory frameworks between sectors and countries,[6,7] and standards may be compromised in a 'race to the bottom'. International cooperation based on a set of common principles for responsible AI could help focus AI research, and build trust across transnational boundaries[1,8]. Commonalities in the ethical principles that underpin published frameworks suggest that a core set of principles is feasible[2]. However, much of the international discussion has emanated from high-income countries (HICs). Very few ethical frameworks applicable to the specific context of AI research in low- and middle-income countries (LMICs) have been published,[9] which is problematic for two reasons.

First, it is well-recognised that ethical frameworks and AI research projects should be developed in tandem with AI and digitisation initiatives; [4,10] however, most LMICs lack the resources to carry out AI research[11,12]. While open sharing of expertise and resources from HICs can aid LMICs in the development of their AI initiatives and ethical frameworks, structural constraints preclude the straightforward transposition of frameworks from other countries[4,13]. Second, cultural differences[14] invite us to think about the place of plural ethical values in the development of overarching ethical frameworks for AI research, particularly with notable gaps in representation.

Conclusion and recommendations

As we move towards a globally interconnected landscape of AI research in healthcare, there is an increasing call for transnational ethical frameworks to regulate AI research. I caution against the current trend, putting forward the view that there is value in the diversity of different ethical frameworks, especially in research. Distinct perspectives can contribute innovative and novel ways of approaching problems and discovering solutions. To achieve this, we need to be respectful of multiple perspectives and recognise the possibility of engagement across differences. The challenge is to gain consensus around shared value commitments in ways that can accommodate and respect the pluralism across transnational frameworks, and to do so in ways that share research ownership and investment across countries.

I suggest a way forward through a two-level framework, with a core statement regarding the existence of any common or shared values, as well as a secondary procedural layer to guide decision-making that can accommodate both shared and plural values in a consistent process for practical regulation and decision-making. Identified core values provide a continuity between different countries and organisations on which trust and a practical framework can be layered upon. The secondary procedural layer should be configured as a flexible space for accommodating contextual features, local nuances and reasonable disagreement.

Rather than focus on the formulation of a universal set of values or principles, harmonisation here ought instead to be directed towards procedural engagement in decision-making. It is more practically feasible and ethically defensible to agree on practical processes for addressing ethical disagreement within a research project without addressing that disagreement through enforcing a single set of universal ethical values. It is just as important that global conversations on AI ethics (and the development of frameworks and guidelines in particular) are not dominated by a small set of actors. Dialogues between different countries and organisations will be necessary to build respectful engagement, mutual understanding and clarity.

Acknowledgements: This work is supported by the Singapore Ministry of Health's National Medical Research Council under its Science Health, and Policy Relevant Ethics, Singapore (SHAPES) Programme (MOH-000951).

References

1. 10.1007/s13347-020-00402-x
2. 10.1038/s42256-019-0088-2
3. 10.1186/s12910-021-00577-8
4. 10.1634/theoncologist.2018-0255
5. 10.7861/futurehosp.6-2-94
6. 10.1080/03080188.2020.1840225
7. 10.1148/radiol.2020201434
8. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization (2021)
9. 10.1007/s43681-022-00218-9
10. 10.1177/1460458220938535
11. https://tinyurl.com/uz8p5vyu
12. 10.1136/bmjgh-2018-000798
13. 10.1186/s12992-020-00584-1
14. 10.1007/s00146-022-01499-8

6. A shift to openness: open consent and open science in AI health research in South Africa

Amy Gooden & Meshandren Naidoo (presenter), University of KwaZulu-Natal, South Africa

Artificial Intelligence (AI) is an invaluable tool for health research and product development in South Africa – however it requires a vast amount of health and genomic data – which is tricky to access! In South Africa, open science has become a focus in the country, as reflected in the Department of Science and Innovation (DSI) *Draft National Open Science Policy* (2022), which aims to incorporate open science into South Africa's national strategy and foster a legal environment that is more open to innovation and data-sharing. Furthermore, given that the Academy of Science of South Africa (ASSAf) has recently published the *draft Code of Conduct for Research* in terms of the *Protection of Personal Information Act 4 of 2013* (POPIA), now is the time to examine this – Keeping with the open science framework, how can we best harness the potential of AI? Open consent may be a solution! But, is it feasible?

A central aspect of health research is consent. In South Africa, various modes of consent have been developed and utilised in the context of health research – including broad consent, specific consent, and blanket consent. Whilst the Department of Health *Ethics in Health Research: Principles, Processes and Structures* (2015) permits broad consent, it does not recommend the use of blanket consent. In contrast, POPIA (and the *draft Code of Conduct for Research*) requires consent to be *specific*, voluntary, and informed. A mode of consent that has not yet been explored in South Africa's research context, and specifically in AI health research, is open consent. Open consent – developed by the Personal Genome Project (PGP) – entails that individuals openly donate and share their data for research.[1] It recognises that privacy cannot be guaranteed, and therefore makes no assurances regarding the anonymity, privacy, or confidentiality of data.[2] Further, to ensure that consent is informed, open consent requires potential participants to take (and pass) an examination to test their understanding of the research and its processes. Thus, open consent can be perceived as an attempt to marry *open science* – and the benefits associated with it – with *informed consent*. Not only does open consent pose potential solutions to the provision of samples and data for health research, and the uptake of AI in this area, but it also promotes open science.

This presentation will consider whether open consent aligns with POPIA (and *draft Code of Conduct for Research*) and the type of consent that it proposes, namely specific consent, as well as whether open consent aligns with the *Ethics in Health Research: Principles, Processes and Structures* recommendation of broad (but not blanket) consent for health research – thereby determining whether open consent may be a viable (and unexplored) solution to consent for AI health research in South Africa. This presentation will also contemplate how the *Draft National Open Science Policy* promotes open science in terms of AI health research (and other applications of AI).

References

1. Fida K Dankar, Marton Gergely & Samar K Dankar 'Informed consent in biomedical research' (2019) 17 *Computational and Structural Biotechnology Journal* 470.
2. Jeantine E Lunshof, Ruth Chadwick, Daniel B Vorhaus et al 'From genetic privacy to open consent' (2008) *Nature Reviews Genetics* 5.

7. A regulatory framework for AI-health research in the Caribbean

Derrick Aarons, The Caribbean Public Health Agency (CARPHA), Jamaica

Introduction

The convergence of Artificial Intelligence (AI), big data methods, and microsystems engineering makes AI-based algorithms for computational neuroscience one of the fastest growing fields of neuro-medical research[1], however, ethical issues such as incidental findings and privacy concerns, transparency and bias, and algorithm discrimination arise[2].

To protect the participants in such research endeavours, robust and appropriate regulations for research involving AI should be implemented across all Caribbean states, which would be in keeping with the objectives of the research protections proposal approved within the Caribbean Community and Commons Market (CARICOM) in 2015 [3].

Commentary

In 2015, the Caribbean Public Health Agency (CARPHA), the regional public health institution in the Caribbean with the responsibility for providing strategic direction in analyzing, defining, and responding to the public health priorities of the 24 member states across the Caribbean, presented the Ministers of Health of the Caribbean Community (The COHSOD) with a 'green paper' proposal to regulate the conduct of research with human participants to provide 'best practices'; ensure consistency and harmonization throughout the Caribbean; and through legislation protect the inhabitants of countries in the region from 'ethics dumping' and harmful exploitative research activities[3]. Currently, no legislation for research with human participants exists in any of these countries, except for Guyana and the Bahamas.

The COHSOD was requested to establish a regional regulatory framework for research involving human participants by approving model legislation for CARICOM countries to regulate research along with regulations for sanctions for non-compliance. CARPHA's proposal was put to the vote and approved unanimously, with the approved green paper being sent to the CARICOM headquarters in Guyana for the legislative draughtsmen to prepare the legislation. However, since then, nothing further has been heard.

With the need for conducting ethical reviews of health research protocols across the Caribbean at a high standard[4], and the current accelerated use of AI in the collecting and processing of health data for research, the legislation and regulations which should have been prepared at the CARICOM Headquarters will need updating to address the new ethical issues posed by AI in health research.

Recommendations

The scope of the CARICOM draft regulations on research should be expanded and updated to provide a more comprehensive research ethics framework that would enable all Caribbean countries to address the new realities and challenges posed by the use of AI in health-related research.

References

1. International Bioethics Committee (IBC). Report of the International Bioethics Committee of UNESCO (IBC) on the Ethical Issues of Neurotechnology. Paris; December 2021: SHS/BIO/IBC-28/2021/3 Rev.
2. Social and Human Sciences Commission (SHSC) of UNESCO. Recommendation on the Ethics of Artificial Intelligence. Report of the SHS, 41st UNESCO General Conference, 2021: pp. 14-39.
3. Aarons D. The historical development of research ethics in the English-speaking Caribbean. *CANREC Bull*.2019; 1(1): 3-7.
4. Aarons D.E. Research in epidemic and emergency situations: A model for collaboration and expediting ethics review in two Caribbean countries. *Developing World Bioethics Journal* Jul.2017; 00:1-10. https://doi.org/10.1111/dewb.12157.

8. How to translate universal principles to local realities: the Chilean experience in AI

Sofía P. Salas, Center for Bioethics, School of Medicine, Clínica Alemana Universidad del Desarrollo, Santiago, Chile

An online experiment, aimed to explore the moral dilemmas faced by autonomous vehicles, which obtained responses from millions of participants from all over the world, evidenced important "cross-cultural ethical variation" in their decisions[1]. The rapidly growing field of AI in research and for clinical uses, have prompted several organizations and governments to launch guidelines and frameworks, and even "AI principles" intended to be used globally. There is consensus regarding some important principles, such as transparency, justice and fairness, non-maleficence, responsibility and privacy; however, the way how these principles are implemented in different settings need to be addressed locally, according to the culture and context of each country or region where AI is deployed. In different areas of research, it has been documented that researchers from the "global South" are under-represented, even when the focus is based on a country or region in this region[2]. Likewise, most of the present guidelines and draft policy documents on AI are written by specialists from more developed countries and organizations. Not surprisingly, there are few guidelines from Africa or South America.

In the present report, I will briefly present the Chilean initiative aimed to have local standards on the development and use of AI, in order to position the country as a world hub for data science, entrepreneurship and innovation ecosystem, and inclusive technological revolution for the 21st century. To elaborate this national policy, the Chilean National Council of Innovation for Development organized an interdisciplinary and diverse group of 12 experts from academia, the productive sector and civil society, and convened in a broad and open manner more than 1,300 people who participated in workshops, 400 who participated in meetings in each of the country's region, and more than 5,300 people who attended 15 meetings in which AI was examined from multiple perspectives and disciplines[3]. The draft document generated by all these initiatives was then submitted to a public consultation process, in which more than 200 people participated. After a thorough analysis, with further meetings with national experts, the first Chilean AI policy was launched in 2020. This AI Policy is based on four principles: AI with focus on the well-being of people, respect for human rights and security; AI for sustainable development; inclusive AI; and globalized AI. In order to achieve an inclusive AI, the actions will place special emphasis on the attributes of integrity and quality of the data to guarantee that their biases are known and adequately treated. This national policy was structured in three axes: 1st: talent development, technology infrastructure, and data management; 2nd: it includes basic and applied research, technology transfer, innovation, entrepreneurship, improvement of public services, economic development based on technology, among others; and 3rd: includes ethics, normative aspects and socioeconomics effects.

To become a useful instrument for people, it is of special importance that AI is developed with gender and sexual diversity perspective, including groups that have been historically relegated such as native people, people with special disabilities, or the most vulnerable sectors of the economy. Likewise, AI must be developed with special consideration for children and adolescents from a perspective of protection, provision and participation.

A special characteristic of this policy is to consider that every action related to AI must be approached in an interdisciplinary manner, enhancing the contribution of the various areas of knowledge. In its own words, "it is impossible to address it from the exclusive view of experts".

References

1. Awad E, Dsouza S, Kim R, Schulz J, Henrich J, Shariff A, et al. The Moral Machine experiment. *Nature*. 2018;563(7729):59-64.
2. Liverpool L. Researchers from global south under-represented in development research. *Nature*. 2021.
3. Política Nacional de Inteligencia Artificial. Chile IA. Ministerio de Ciencia, Tecnología, Conocimiento e Innovación. Gobierno de Chile 2020.

9. Developing a governance framework for data science health research in Nigeria

Oluchi Maduka (presenter), Simisola Akintola, Shawneequa Callier, Ayodele Jegede, Clement Adebamowo
Center for Bioethics and Research, Ibadan, Nigeria

Introduction
Data Science Health Research (DSHR) has enormous potential for discovery and optimization of healthcare. Data science models that incorporate Artificial Intelligence (AI) and Machine learning technologies are examples of data science related innovations that are yielding transformative changes for DSHR in Nigeria. Following COVID-19, the use of AI in health research have become common, with institutions such as the Nigeria Center for Disease Control (NCDC), the Nigerian Institute of Medical Research (NIMR), and the Institute of Human Virology Nigeria (IHVN) utilizing AI for surveillance, health research, and infection dynamics. Through '*ubenwa'* and '*helpmum'*, the private sector's involvement has gained global attention. Data science can be used to aggregate huge amounts of data from multiple levels of the health care system and other spheres of human activities to make discoveries and inferences, however, it raises substantial ethical, legal, and social issues, such as questions about the content and quality of the consent given by individuals, privacy, ownership of data, and benefits and harms to individuals participating in DSHR. These issues are urgent global concerns, particularly in low- and middle-income countries like Nigeria, where regulation is evolving. To effectively respond to these challenges, in the Bridging Gaps in the Ethical, Legal, and Social Implications of Data Science Health Research (BridgELSI) project, we reviewed existing legal and ethical oversight for DSHR in Nigeria as a foundation for development of novel ethical oversight of DSHR in the country.

Methods
Legal research and analysis approaches, including text analysis and case law research, were used to review and assess legal rules such as statutes, guidelines, regulations, and policies applicable to DSHR. We conducted manual and electronic searches using the index of Laws of Federation of Nigeria 2010 and its Annual Supplements, Nigerian Weekly Law Report, LawPavillion Electronic Law Report, and WestLaw. Other secondary sources were accessed using Google Scholar, Jstor, PubMed, and HeinOnline.

Result
Our review reveals that Nigeria has a robust health research ethics framework for health research whereas data privacy and protection fall under several laws and regulations. The Federal Constitution, the National Health Act, the Nigeria Data Protection Regulation, the National Code for Health Research Ethics, and the Code of Medical Ethics, provide guidance for DSHR, particularly in relation to data subject rights, privacy and confidentiality, consent of research participants, protection of cultures, groups and communities, and protection from discrimination. However, there are gaps in the laws as it relates to use of anonymized data, de-identified data, and publicly available data for research.

Conclusion
There is a need for the development of new governance frameworks for DSHR that builds on existing laws and includes broad stakeholders' engagement. This ensures the protection of participants, researchers, and their products, and enhances trust and buy-in by the general population.

Keywords: Data science, health research, data protection, artificial intelligence, BridgELSI

10. Adaptability of India's Health Data Regulations

Rupanjali Karthink, Duke Law School, Duke University, USA

Globally, regulatory policies in general are developed using one-time decision-making processes, using limited ex-ante assessments and ad-hoc reviews or revisions. However, in the real world, changes are constantly occurring across different domains - social, economic, technological, political, and cultural. To be meaningful and effective, the regulations need to keep pace with these changes. Adaptive regulations offer an array of mechanisms where decision-making is not a one-time process, instead, it is iterative and planned, based on new information and changing circumstances.

The complexity and rise of health data have resulted in increasing applications of artificial intelligence (AI) in healthcare, patient engagement, and administrative activities. While there are numerous benefits of electronic access and exchange of health data, there are increasing risks of privacy and security breaches. Further, with the advent of big data and advanced analytics, a lot of non-health data is being collected and traded online. This non-health data could be a better predictor of an individual's health than his/her health records. Thus, new and emerging technologies are changing the nature of privacy and how it could be protected. In India, laws and policies regarding AI and health data are still evolving. Therefore, there is an opportunity to apply theoretical principles of adaptive regulation to India's AI and health data policies and propose recommendations to design adaptive policies and laws.

Based on the review of literature, 6 broad features of adaptive regulation are synthesized from the perspective of a learning-oriented decision-making process. These are: (i) assessing risk and uncertainties, (ii) broader and fuller impact assessment, (iii) monitoring and evaluation, (iv) iterative decision-making and policy adjustment, (v) public participation, and (vi) adaptive governance structures. These six features are embedded in the form of an adaptive regulatory cycle with three stages of pre-implementation, implementation, and post-implementation. In this context, India's four health data policies and one legislative bill are analyzed and 10 key stakeholders interviewed. The analysis is anchored on the application of the adaptive regulatory cycle with six adaptive features.

In the health data sector, India's regulatory cycle in the pre-implementation stage (assessing risks and uncertainties, and broader impact assessments) indicates low adaptiveness on the books and moderate adaptiveness in practice. In the implementation stage (monitoring and evaluation), it indicates high adaptiveness on the books and moderate adaptiveness in practice. And in the post-implementation stage (iterative decision-making), it indicates high adaptiveness both on the books as well as in practice. Regarding the two overarching adaptive features of public participation and adaptive governance structures, the former shows high presence both on the books and in practice while the latter shows low presence both on the books and in practice.

Combined document and interview analysis indicate a gap in monitoring and evaluation (M&E) in practice whereas it indicates high prevalence of iterative decision-making in both theory and practice. However, the interview analysis also suggests that these iterations and policy revisions are not informed by formal policy evaluations. Therefore, this finding connects with the limited effectiveness of M&E in practice. To address these gaps in India's health data sector, it is recommended to introduce structured decision-making processes (e.g. risk assessment, regulatory impact assessment, etc., using simplified and flexible methodologies), focus on formal policy evaluations, and strengthen inter-agency coordination. Further, considering this sector is dynamic and nascent in law and policymaking, more built-in provisions of periodic reviews are recommended.

Overall, most of India's analyzed health data policies are still evolving. Therefore, before India begins to use AI in health research, it should create a more 'adaptive framework' for health policy so that it can best learn and improve how AI is working, and address any ethical issues that arise with using AI.

Website and social media

Website

The GFBR website contains details of all previous GFBR meetings, including meeting reports and presentations where available. Presentations from this meeting will be posted on the site shortly after the event. www.gfbr.global

Social media

At this Forum meeting we encourage the use of social media to engage in conversation and to spread the discussions to those unable to attend the meeting itself.

If you use Twitter, please use the hashtag #gfbr to tag your tweets about the meeting and follow us @ForumBioethics

Social media etiquette

Pease follow these guidelines to ensure that social media is used in a positive way that benefits the meeting and its participants:

1. Be polite and constructive

   If you are going to tweet during a presentation or discussion, make sure you do so on a positive note. Share what you learned from the session or pose an interesting key question that would warrant further discussion. If the presenter has a social media profile, tag them in your post, and use the conference hashtag #gfbr.

2. Respect presenters' requests for no social media

   Some topics discussed may be sensitive or present early findings from research that has not yet been published. The chair should indicate at the beginning of a session if the presenter would prefer their talk not to be tweeted.

3. Engage with others

   The meeting is an opportunity to learn from others, to hear about their experiences and perspectives. Social media is a useful tool for these interactions but do take the opportunity to talk to people in person during the breaks as well!