

The promise of data science for health research in Africa

Received: 6 December 2021

Accepted: 15 September 2023

Published online: 29 September 2023

 Check for updates

Clement A. Adebamowo ^{1,2} ✉, Shawneequa Callier ^{3,4},
Simisola Akintola ^{2,5,6}, Oluchi Maduka ², Ayodele Jegede^{2,6,7},
Christopher Arima⁸, Temidayo Ogundiran ^{2,6,9}, Sally N. Adebamowo ^{1,2} &
BridgELSI Project as part of the DS-I Africa Consortium*

Data science health research promises tremendous benefits for African populations, but its implementation is fraught with substantial ethical governance risks that could thwart the delivery of these anticipated benefits. We discuss emerging efforts to build ethical governance frameworks for data science health research in Africa and the opportunities to advance these through investments by African governments and institutions, international funding organizations and collaborations for research and capacity development.

Data science is poised to revolutionize healthcare and research by enabling the development of novel interventions and groundbreaking strategies derived from high-quality and efficient analyses of the huge datasets derived from the activities of our multifaceted lives. Data science applications use high-performance computational infrastructure to process massive datasets from personal, public, and commercial sources including healthcare systems, smartphones, shopping records, social media postings, and wearable devices. Using novel, complex, and occasionally opaque algorithms, data scientists generate new insights and generalizable knowledge. Examples of data science applications include combination of diverse data streams to develop bio-preparedness, monitoring, and response strategies for infectious diseases outbreaks in human health and in agriculture¹. Other examples include the use of Geographical Information System (GIS) data to map spatial variations in the determinants, incidence, prevalence, and outcomes of disease, and the response of healthcare systems^{2,3}. Several countries also utilize data science to monitor and evaluate multi-sectoral progress towards meeting the UN Sustainable Development Goals (SDG), reduce fraud and corruption, identify fake pharmaceutical products, and improve supply chain management to prevent stock-outs^{4–11}. More recently, Large Language Models (LLM)

have captured public imagination with the release of tools like GPT-4. LLMs are rapidly being used to transform multiple industries and disciplines^{12–17}. With these capabilities, data science is rapidly transforming the landscape and touching vast areas of human endeavors.

Data science health research is the novel application of data science methods and technologies for systematic generation, collection, processing, storage, management, analyses, visualization, interpretation, and communication of health-related data to develop generalizable knowledge and generate actionable insights. Examples of data science health research include integration of data from multiple omics technologies to generate insights about biological mechanisms and diseases' pathways, and identify novel therapeutic and preventive opportunities¹⁸. Other examples include mining electronic health records for precision medicine¹⁹ and analysis of medical images for computational histopathology and radiology to improve diagnoses, among other uses^{20,21}. In public health, data scientists are transforming practice through the application of high-level computational methods to population-level datasets typically used in public health to advance precision public health practice and research²². Examples of public health data science research include mining social media data for early detection, analyses of trends, education, planning and implementation

¹Department of Epidemiology and Public Health, and Greenebaum Comprehensive Cancer Center, University of Maryland School of Medicine, Baltimore, MD, USA. ²Department of Research, Center for Bioethics and Research, Ibadan, Nigeria. ³Department of Clinical Research and Leadership, School of Medicine and Health Sciences, The George Washington University, Washington DC, USA. ⁴Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. ⁵Department of Business Law, Faculty of Law, University of Ibadan, Ibadan, Nigeria. ⁶Department of Bioethics and Medical Humanities, Faculty of Multidisciplinary Studies, University of Ibadan, Ibadan, Nigeria. ⁷Department of Sociology, University of Ibadan, Ibadan, Nigeria. ⁸Syracuse University College of Law, Syracuse, NY, USA. ⁹Department of Surgery, College of Medicine, University of Ibadan, Ibadan, Nigeria. *A list of authors and their affiliations appears at the end of the paper. ✉ e-mail: cadebamowo@som.umaryland.edu

of health systems responses to infectious and non-communicable diseases^{8,9,11,23}, and application of myriad novel technologies and algorithms for precision public health²². Data science health research therefore presents huge opportunities for the application of novel methods and transformative technologies that would solve many healthcare challenges facing African people today and enable wider availability of high quality and cost-efficient health services.

Africa, potentially, has the most to gain from implementation of data science for health care and research. With a population expected to reach 2.5 billion people or ~25% of the world's population by 2050, data science technologies would enable African countries to leapfrog legacy healthcare systems and technologies, and dramatically transform lives on the continent^{24,25}. Even though Africa currently constitutes 17% of the world's population, it bears 25% of the world's disease burden, has only 3% of the world's healthcare workers, and 2% of global health research output²⁶. This is due to limited infrastructure, lack of trained personnel, poor funding, economic and social instability which hinder access to clinical and preventive services²⁶. Global public health emergencies such as emerging and re-emerging infectious diseases epidemics and climate change pose more challenges to African countries than the rest of the world. African countries would therefore require innovative data science tools and strategies to overcome these challenges.

Substantial gaps exist in the representation of people from Africa in the datasets currently used to build data science models and applications²⁷. This underrepresentation renders data science models and algorithms unstable and potentially inaccurate in African populations²⁷. Without dedicated and focused efforts at remediation, persistence of this data science equity gap would worsen and this portends dire consequences for data science health research in African populations.

African researchers, institutions, governments, and the private sector are already using data science for research, discoveries, and preventive and clinical care^{28,29}. Most of these uses involve novel applications or extensions of current healthcare expertise and technologies. Examples of data science applications already in use in Africa include teleradiology and telepathology, patients' navigation and clinical decision support (CDS) tools, integration of genomics data into public health and clinical care, and cancer screening^{29–33}. However, most of these applications were designed, developed, tested, and validated outside Africa. They may not have been adequately evaluated in African populations and may be insensitive to local contexts and health priorities^{28,34,35}. They may therefore primarily benefit people outside the continent. Given the novel opportunities being created by data science applications, it is critical to develop and implement technologies that are relevant and adapted to the contexts in which they would be used in Africa.

Investment in data science health research infrastructure in Africa

Several initiatives are being implemented in Africa to develop data science health research capacity, build infrastructure, implement training programs, organize scientific conferences, and engage in international collaborations that would empower African institutions to generate datasets, develop and apply data science models, and close the data science gap between Africa and high-income countries (HIC). In 2022, the NIH Common Fund awarded 20 grants worth \$74.5 million in the "Harnessing Data Science for Health Discovery and Innovation in Africa (DS-I Africa)" program to accelerate data science health research in Africa. The projects being implemented by the DS-I Africa program include a Coordinating Center, an Open Data Science platform, seven training programs, four ethical, legal, and social implications (ELSI) projects, and seven research projects. (Table 1)³⁶.

The DS-I Africa initiative builds on the infrastructure previously developed by programs such as the \$176 million Wellcome Trust and

NIH-funded Human Heredity & Health in Africa (H3Africa) program³⁷. H3Africa built new collaborations among scientists, developed genomics research infrastructure, and created publicly available governance and ethics policies for the African genomics research ecosystem^{38–40}. Data science conferences and training programs are also proliferating in Africa including the Data Science Africa—an AI and Data Science Research Group at Makerere University, Uganda, the multi-country African ML and AI organization—Deep Learning Indaba, the School for Data Science and Computational Thinking at Stellenbosch University in South Africa, the African Institute for Mathematical Sciences Centre of Excellence in Cameroon and the African Center of Excellence in Data Science in the University of Rwanda. These programs are critical to generating data that will close the data science gap in Africa and enrich global resources for data science health research.

Developing a comprehensive framework for the governance of data science health research across Africa

In contrast with other disciplines where data science is also rapidly advancing, health research already has established frameworks and infrastructure for ethical governance. Substantial investments by the US National Institutes of Health (NIH), UK Wellcome Trust, the European Union through the European-Developing Countries Clinical Trials Partnership (EDCTP), African governments and institutions have significantly expanded African health research ethics infrastructure in the past few decades^{41–45}. Despite these investments, there remain many unresolved challenges including concerns about quality of informed consent, data ownership, data sharing, benefit-sharing, privacy, autonomy, exploitation, and weak governance^{46–49}. Recent examples of these challenges include controversies about community benefit and data sharing during COVID-19 pandemic research and unauthorized use of DNA samples from African populations to develop a DNA genotyping microarray chip^{38,50}. Given the methods and technologies used in data science health research, its potential to exacerbate preexisting health research ethics problems and generate new ones are quite substantial.

Research consortia like H3Africa developed policies on samples and data sharing, biorepositories, publications, collaborations, and commercialization³⁸. They also provided training for researchers and ethics committees. For example, H3Africa's publication policy gives African researchers protected time to analyze and publish before their data becomes publicly available. This protection, which is designed to accommodate the infrastructural and personnel challenges faced by African researchers, enables them to frame the narrative about their research and advance their research priorities. Other policies require that the funded studies should focus on African health priorities, be led by African researchers, and that African institutions should be the primary recipients of the research grants, even when they collaborate with international institutions. These are meaningful practices that the emerging data science health research programs should emulate and further develop.

The rapid evolution of data science methods, utilization of complex algorithms, and huge datasets obtained from a variety of sources under uncertain consenting procedures particularly challenges the current model of ethical review of health research^{51,52}. When data science health research projects are conducted within single institutions, e.g., computational histopathology of diagnostic biopsies, sufficient ethical oversight can be provided by the institution's health research ethics committee. However, even in these situations, institutional health research ethics committees may lack sufficient knowledge, expertise, and experience in evaluating the ethical dilemmas that may accompany these studies and struggle to provide adequate ethical review and oversight⁵³. In addition, the methods and technologies of data science often run counter to the established principles and practices of ethics review of health research thereby creating

Table 1 | Harnessing data science for health discovery and innovation in Africa (DS-I Africa) projects, collaborating institutions and study aims

No.	Project title	Grant type	Collaborating institutions	Objectives
1.	Advancing Public Health Research in Eastern Africa through Data Science Training (APHREA-DST)	Training program	Columbia University, USA Addis Ababa University, Ethiopia University of Nairobi, Kenya	Develop new context-specific MS in public health data science. Mentor faculty to build and strengthen capacity in health data science. Conduct short-term training programs.
2.	Computational Omics and Biomedical Informatics Program (COBIP)	Training program	Oregon Health & Science University (OHSU), USA University of Cape Town, South Africa.	Develop data science training program focused on the health priorities of Africa. Implement multi-disciplinary in biomedical data science. Establish an international center of excellence in computational omics and biomedical informatics in Africa.
3.	Data Science for Child Health Now in Ghana (DS-CHANGE)	Training program	KNUST, Ghana University of Washington, Seattle Children's Hospital and Research Institute Smile Train	Deliver a comprehensive mentored interdisciplinary training program in data science health research. Increase capacity in biomedical data science. Develop proficiency in effective team science methods for interdisciplinary teams.
4.	Makerere University Data Science Research Training to Strengthen Evidence-Based Health Innovation, Intervention and Policy (MakDARTA)	Training program	Makerere University, Uganda Johns Hopkins University, USA	Establish graduate degree training in data science at MakCHS. Provide Postdoctoral training. Train MakCHS faculty to enhance their capabilities for data science research and mentorship of data science trainees.
5.	NYU-Moi Data Science for Social Determinants Training Program (DSSD)	Training program	New York University, USA Moi University, Kenya Brown University, USA	Create intellectual meeting spaces for a data science and health. Propel and sustainably advance data science capacity in Kenyan institutions.
6.	Research Training on Harnessing Data Science for Global Health Priorities in Africa	Training program	Harvard University, USA Heidelberg University, Germany University of KwaZulu-Natal, South Africa	Train mid-level and senior researchers at UKZN. Build a critical mass of junior public health and medical professionals across South Africa, Ghana, Nigeria, Tanzania, and Uganda. Implement sustainable master's degree program in health data science at the UKZN.
7.	Research Training in Data Science for Health in Rwanda	Training program	Washington University in St. Louis, USA University of Rwanda, Rwanda African Institute for Mathematical Sciences, Rwanda	Identify competencies in major scientific areas. Foster innovative team science transdisciplinary approach to research. Build institutional capacity to support long-term sustainability.
8.	Combating Antimicrobial Resistance (AMR) in Africa Using Data Science (CAMRA)	Collaborative research center	Redeemer's University, Nigeria J. Craig Venter Institute, Inc., USA University of Nebraska Medical Center, USA	Comparative phenotypic and genotypic studies of clinical isolates to inform trends in AMR and dynamics of transmission. Develop portable screening tool for clinical care. Explore benefit of an aminoglycoside conjugated to an antimicrobial peptide for enhanced bactericidal activity.
9.	Developing data science solutions to mitigate the health impacts of climate change in Africa: the HEZAT Center	Collaborative research center	University of the Witwatersrand, South Africa Liverpool School of Tropical Medicine (LSTM), UK IBM Research Africa, South Africa University of Basel, Switzerland University of Washington, USA	Prepare a data science ecosystem and platform to support analysis of existing environmental, biomedical, and other data. Increase knowledge and develop sustainable solutions to address climate change and health. Advance capacity for data science research in climate change and health in sub-Saharan Africa. Serve as a continental resource and center of excellence for policy and practice.
10.	Harnessing Data Science to Promote Equity in Injury and Surgery for Africa (D-SINE)	Collaborative research center	University of Buea, Cameroon University of California, Los Angeles and Berkeley, USA The Cameroonian Ministry of Public Health African Institute for Mathematical Sciences, Cameroon University of Cape Town, South Africa.	Decrease the burden of injuries and surgical diseases through improved surveillance, prevention, and treatment. Improve access to quality surgical care in Cameroon and other SSA countries.
11.	Multimorbidity in Africa: Digital innovation, visualization, and application (MADIVA)	Collaborative research center	University of the Witwatersrand, South Africa African Population and Health Research Center, Kenya IBM Research Africa	Link integrated data sets and create dashboard that will allow researchers to find data and plan projects. Create a clinic dashboard to help clinic managers and health officials monitor and plan effectively.

Table 1 (continued) | Harnessing data science for health discovery and innovation in Africa (DS-I Africa) projects, collaborating institutions and study aims

No.	Project title	Grant type	Collaborating institutions	Objectives
			South African Population Research Infrastructure Network Vanderbilt University Medical Center, South Africa	Develop robust methods for risk profiling of individuals and communities using heterogeneous data and pioneer use of polygenic risk scores in Africa.
12.	MUST Data Science Research Hub (MUDSRh)	Collaborative research center	College of Ophthalmologists of Eastern, Central and Southern Africa Kwame Nkrumah University of Science and Technology, Ghana Massachusetts Institute of Technology, USA Massachusetts General Hospital, USA	Use data science to analyze medical images for advancing clinical care. Integrate data science with implementation science to promote clinical impact.
13.	Role of Data Streams In Informing Infection Dynamics in Africa- INFORM Africa	Collaborative research center	Institute of Human Virology Nigeria Centre for the AIDS Programme of Research In South Africa	Establish data streams from public and private sectors to understand multilayer interactions that explain the dynamics and impact of COVID-19 pandemic. Develop geospatial tools for pandemic surveillance and response to improve preparedness. Expand data science research opportunities and capacity. Maintain sustained engagement with the policy makers and governments.
14.	Utilizing health Information for Meaningful impact in East Africa through Data Science (UZIMA-DS)	Collaborative research center	Aga Khan University, Kenya University of Michigan, USA	Harmonization of multimodal data sources for meaningful use and analyses. Leveraging temporal patterns of data to identify trajectories through prediction modeling using AI/ML-based methods. Engaging with key stakeholders to identify pathways for dissemination and sustainability of these models into target communities.
15.	Bridging Gaps in the ELSI of Data Science Health Research in Nigeria (BridgELSI)	ELSI project	Center for Bioethics and Research (GBR), Nigeria George Washington University, USA University of Maryland School of Medicine, USA	Evaluate current legal frameworks for ethical oversight of data science health research. Develop new and innovative governance frameworks to support data science health research in Nigeria. Prospectively evaluate the knowledge, attitude, and practices to current and emerging ELSI of data science research in Nigeria.
16.	DS-I Africa - Law	ELSI project	University of Kwazulu-Natal, South Africa	Investigate the legal themes associated with modes of informed consent to the use of data; the nature and content of individual and community rights in genomic data; the use of individuals' geospatial data for public health surveillance; the cross-border sharing of data; and the use of data as basis for Artificial Intelligence (AI).
17.	Public Understanding of Big data in Genomics Medicine in Africa (PUBGEM-Africa)	ELSI project	University of Cape Town, South Africa	Investigate models of public engagement and preparedness for big data use in health. Explore the roles and responsibilities of stakeholders regarding intellectual property and commercial use of genomics big data in health. Investigate public perceptions of governance of big data in health. Develop data governance frameworks for innovation and health in Africa.
18.	Research for Ethical Data Science in Southern Africa (REDSSA)	ELSI project	Stellenbosch University, South Africa University of North Carolina, USA	Obtain key stakeholder views regarding the governance of data science in DSI-Africa Research Hubs in Southern Africa. Develop guidance documents. Create ELSI networks and communication channels on data science in Southern Africa.
19.	DS-I Africa Coordinating Center	Coordinating Center	University of Cape Town, South Africa	Develop an efficient CC addressing all the joint administrative, collaborative, and logistical needs of the consortium.
20.	Open Data Science Platform	Open Data Science Platform	University of Cape Town, South Africa	Establish the ODSP gateway and its elwazi platform to facilitate novel discoveries for health.

situations that may be beyond the capacity of individual ethics committees to resolve^{54–56}. In such cases and in others where data science health research is being conducted in multiple institutions within the same country, collaborations between the ethics committees or centralization of ethics review, for example, by national health research ethics committees may be required to provide ethical oversight⁵⁷. National health research ethics committees can constitute standing review committees that, in addition to chartered members, may also include local and international experts as ad-hoc members who can provide ethical oversight for complex data science health research within national boundaries⁵⁷. This centralization of ethical review enables efficient utilization of scarce data science health research ethics expertise and improves the efficiency of ethical review^{57,58}. Despite these innovations, even national health research ethics committees are susceptible to some of the problems affecting institutional health research ethics committees including lack of resources, lack of independence, poor funding and lack of efficacy, albeit to a lesser degree^{44,45,53}.

What should be next on the agenda for data science health research ethics in Africa?

Improve institutional and national health research ethics governance infrastructure

Despite the tremendous investments in recent decades, the capacity, resources, and infrastructure for ethical oversight of health research in Africa remains weak and poorly resourced. A surge of data science health research projects would significantly strain and may overwhelm the system. The major responsibility for building and maintaining national and institutional health research ethics infrastructure rests with African governments and local institutions. Information showing how research significantly boosts the intellectual and economic capital of institutions and countries, and are engines for growing local and national economies may encourage more investment in research infrastructure, including ethical review in Africa. Many African institutions built their current health research ethics programs to support local investigators involved in collaborative international research projects. Research sponsors should incentivize development of local data science health research ethics capacity by linking progress in this domain with new research funding⁵⁹. This would be highly impactful and motivate significant institutional response.

Research ethicists should engage with their local research and data science communities to better understand data science health research methods and projects, and jointly develop ethical governance frameworks that build on existing research ethics oversight infrastructure. Well-funded, well-designed, and sufficiently long training programs that have enabled African countries to avoid widespread egregious harm to research participants despite recent growth in the volume and sophistication of health research on the continent, are also needed for data science health research ethics⁴². These would improve the knowledge of research ethicists about data science health research and that of data scientists about health research ethics, build local capacity that would enable local ownership and sustainment of training programs, and support the conduct of research into contextual data science health research ethics in Africa while contributing to the global health research ethics discourse.

Develop culturally and resource-level appropriate national laws, guidelines, and regulations, and the infrastructure for enforcement

Many African governments are rushing to enact laws similar to the European Union's General Data Protection Regulation (GDPR) and modifying them for their environments^{60,61}. The major challenges with GDPR and similar data protection laws include lack of sufficiently explicit frameworks for enforcement, complexity of certain provisions, a focus that is often insufficient for the nuances of data science

health research^{62,63}. Other suggestions for ethical governance of data science include giving participants ownership of their digital selves or using blockchain technologies to protect digital privacy and securely share data^{64,65}. These approaches are highly technological, expensive and are not resource-level appropriate in the African health research setting⁶⁶.

Research and training consortia are developing novel policies, ideas, and implementation strategies for ethical regulation of data science health research^{67,68}. These consortia must engage frontline stakeholders in different research environments and serve as petri-dishes for experiments into ethical regulation of data science health research. Governments and their agencies also have major roles to play in engendering and maintaining public trust, accountability, and support that are required to sustain public engagement in and support for data science health research.

Existing health research ethics regulations already have the essential elements for the ethical oversight of all types of health research. While data science health research includes novel methods and technologies, these do not abrogate the foundations, principles, and practices of modern health research ethics. African countries can quickly introduce sufficient oversight of data science health research by adding to or modifying existing regulations.

Develop and implement enforceable multinational regulations

Most data science health research sponsors, principal investigators, and scientists belong to international or commercial organizations that may not have local offices in Africa and may not be subject to national laws, guidelines, and regulations⁵⁸. This poses significant problems for oversight and accountability. Multilateral agencies including the United Nations and its organs, governments, advocates, bioethicists, and researchers have conducted multiple consultations and stakeholders' meetings leading to issuance of guidelines on the use of data science in healthcare, research, and policy^{69,70}. These guidelines call for development of multinational frameworks for data science health research to prevent egregious harm and maximize data science's benefits to global populations⁵². To ensure relevance and implementation, these multinational agencies should work with African institutions to develop the mechanisms for enforcement of these principles, model laws, guidelines, and regulations for ethical oversight of data science health research across national borders. They should rigorously engage a broad range of stakeholders including those whose voices are typically drowned out in global discourse. Innovations in virtual meeting technologies should enable cost-effective, frequent, and sustained global engagement of stakeholders.

Reduce digital inequity and increase volume and diversity of African datasets

Data science relies on large repositories of data generated by individuals as they engage with the healthcare system, during activities of daily living, and participation in research⁷¹. Healthcare data may derive from electronic health records, surveillance data, diseases' registries, etc., while other datasets may be derived from wearable devices and other digital footprints. Large-scale genomic, transcriptomic, proteomic, and other omics research projects are also generating huge amounts of data for data science health research. Despite interventions like H3Africa and DS-I Africa, more of these data are still being generated in HIC compared to LMIC^{27,72}. The resulting digital data inequity is pervasive and growing worse⁷³. Digital health innovations also contribute to the widening inequities because of the "inverse care law" which postulates that well-resourced individuals are more likely to be aware of and utilize digital health interventions⁷⁴. Unaddressed, these inequities will lead to severe and adverse health outcomes for majority of the world's population⁷⁵. Urgent, sustained, large-scale efforts are required to reverse this trend for the sake of equity and justice.

Multi-level interventions guided by frameworks for digital health equity such as the digital determinants of health (DDOH) would be useful for identifying the barriers and facilitators, and guide meaningful interventions to increase the volume of digital health data generated in Africa⁷³. General investments in healthcare systems, implementation of electronic health records, improvements in diseases' registries, and broader utilization of digital systems will increase the amount of digital data generated by African populations. Additional systematic interventions that are similar to but substantially larger than programs like H3Africa and DS-I Africa are needed to ensure that African countries, at a minimum, keep up with the high volume of omics and other data types being generated in HIC for data science.

Reduce and eliminate algorithmic bias, data colonization, and extractive research

Data science technologies produce algorithmic bias by replicating and reinforcing societal biases that benefit or disadvantage certain individuals or groups. This results in structural, racial and ethnic biases in the HIC where most data science technologies are developed⁷⁶. These algorithmic biases coupled with the lack of equity and diversity in the foundational datasets used to develop, train, and validate data science algorithms lead to algorithmic deprivation, discrimination, and distortion^{77–81}. Other concerns, particularly with respect to data science in Africa, are data colonization and extractive research⁸². Biased and inequitable algorithms lead to ethically, socially, politically, and economically undesirable outcomes in health research and health care, and can negatively affect perceptions of fairness, acceptability, and trust in applications derived from data science health research. These have the potential of denying populations that are most in need, the benefits of data science health research⁷⁶. These harms are unpredictable and may not be remediable post hoc, they therefore require vigorous and robust attention a priori^{51,52,76}.

Many approaches have been recommended to reduce or eliminate algorithmic bias in data science health research^{79,81,83–85}. These include improving the diversity of data scientists through targeted capacity building programs, creation and implementation of guidelines and policies, implementation of programs to detect and rectify algorithmic bias, training data scientists on health research ethics^{52,85–87}. These interventions require long-term commitments that go beyond the typical duration of many HIC research grant award mechanisms. They also require strong commitment by national governments, local institutions, and research sponsors. Novel strategies for supporting the development of personnel, resources, and infrastructure for data science health research in Africa that are aligned with clear goals and objectives, rather than just utilizing frameworks developed and used in the substantially different health research ecosystems of HIC are desperately needed.

Given the scope of data science health research, its potential to improve health outcomes, engender more equitable research participation, reduce marginalization, and utilize heterogeneous data types, all stakeholders must be urgently engaged in development of the most efficacious governance frameworks for it.

References

- Keshavamurthy, R., Dixon, S., Pazdernik, K. T. & Charles, L. E. Predicting infectious disease for biopreparedness and response: a systematic review of machine learning and deep learning approaches. *One Health* **15**, 100439 (2022).
- Tanser, F. C. & le Sueur, D. The application of geographical information systems to important public health problems in Africa. *Int. J. Health Geogr.* **1**, 4 (2002).
- Stewart, K. et al. Modeling spatial access to cervical cancer screening services in Ondo State, Nigeria. *Int. J. Health Geogr.* **19**, 28 (2020).
- Georgakopoulos, S. V., Gallos, P. & Plagianakos, V. P. Using Big Data Analytics to Detect Fraud in Healthcare Provision in 2020 *IEEE 5th Middle East and Africa Conference on Biomedical Engineering (MECBME 2020)*. (IEEE, New Jersey, 2020).
- Gebremeskel, G. B., Yi, C., He, Z. & Haile, D. Combined data mining techniques based patient data outlier detection for healthcare safety. *Int. J. Intell. Comput. Cybern.* **9**, 42–68 (2016).
- Kumar, A., Choudhary, D., Raju, M. S., Chaudhary, D. K. & Sagar, R. K. Combating counterfeit drugs: a quantitative analysis on cracking down the fake drug industry by using Blockchain technology in *Proc. 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence 2019)*. (IEEE, Noida, India, 2019).
- United Nations. *Big data for sustainable development*, <https://www.un.org/en/global-issues/big-data-for-sustainable-development> (2023).
- Arslan, J. & Benke, K. K. Artificial Intelligence and Telehealth may provide early warning of epidemics. *Front. Artif. Intell.* **4**, 556848 (2021).
- Judson, S. D. et al. COVID-19 data reporting systems in Africa reveal insights for future pandemics. *Epidemiol. Infect.* **150**, e119 (2022).
- Shaw, N. & McGuire, S. Understanding the use of geographical information systems (GISs) in health informatics research: a review. *BMJ Health Care Inform.* **24**, 940 (2017).
- Kimera, R. et al. in *Leveraging Data Science for Global Health* (eds Celi, L. et al.) 329–350 (Springer International Publishing, 2020).
- Waisberg, E. et al. GPT-4: a new era of artificial intelligence in medicine. *Irish J. Med. Sci.* <https://doi.org/10.1007/s11845-023-03377-8> (2023).
- Akbasli, I. T. & Bayrakci, B. Balancing caution and innovation: exploring the potential of large language models in critical decision-making. *Crit. Care* **27**, 172 (2023).
- Bair, H. & Norden, J. Large Language Models and Their Implications on Medical Education. *Acad. Med.* <https://doi.org/10.1097/ACM.0000000000005265> (2023).
- Perera Molligoda Arachchige, A. S. Large language models (LLM) and ChatGPT: a medical student perspective. *Eur. J. Nucl. Med. Mol. Imaging* <https://doi.org/10.1007/s00259-023-06227-y> (2023).
- Qureshi, R. et al. Are ChatGPT and large language models “the answer” to bringing us closer to systematic review automation? *Syst. Rev.* **12**, 72 (2023).
- Sorin, V., Barash, Y., Konen, E. & Klang, E. Large language models for oncological applications. *J. Cancer Res. Clin. Oncol.* <https://doi.org/10.1007/s00432-023-04824-w> (2023).
- Dar, M. A. et al. Multiomics technologies: role in disease biomarker discoveries and therapeutics. *Brief. Funct. Genomics* **22**, 76–96 (2023).
- Martínez-García, M. & Hernández-Lemus, E. Data integration challenges for machine learning in precision medicine. *Front. Med.* **8**, 784455 (2021).
- Shmatko, A., Ghaffari Laleh, N., Gerstung, M. & Kather, J. N. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. *Nat. Cancer* **3**, 1026–1038 (2022).
- Milam, M. E. & Koo, C. W. The current status and future of FDA-approved artificial intelligence tools in chest radiology in the United States. *Clin. Radiol.* **78**, 115–122 (2023).
- Goldsmith, J. et al. The emergence and future of public health data science. *Public Health Rev.* <https://doi.org/10.3389/phrs.2021.1604023> (2021).
- Cohen, I. G., Lynch, H. F., Vayena, E. & Gasser, U. *Big Data, Health Law, and Bioethics* (Cambridge University Press, 2018).
- World Health Organization. *Atlas of African health statistics 2022: health situation analysis of the WHO African Region*. (World Health Organization. Regional Office for Africa, 2022).
- Anon. in *The Economist* (The Economist Group, London, 2020).

26. World Health Organization. *World health statistics 2022: monitoring health for the SDGs, sustainable development goals*, <https://www.who.int/publications-detail-redirect/9789240051157> (2022).
27. Gao, Y., Sharma, T. & Cui, Y. Addressing the challenge of biomedical data inequality: an Artificial Intelligence Perspective. *Annu. Rev. Biomed. Data Sci.* <https://doi.org/10.1146/annurev-biodatasci-020722-020704> (2023).
28. Owoyemi, A., Owoyemi, J., Osiyemi, A. & Boyd, A. Artificial intelligence for healthcare in Africa. *Front. Digital Health* **2**, 6 (2020).
29. Ezugwu, A. E., Oyelade, O. N., Ikotun, A. M., Agushaka, J. O. & Ho, Y.-S. Machine learning research trends in Africa: a 30 years overview with bibliometric analysis review. *Arch. Comput. Methods Eng.* **30**, 4177–4207 (2023).
30. Bellemo, V. et al. Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study. *Lancet Digital Health* **1**, e35–e44 (2019).
31. Botwe, B. O. et al. The integration of artificial intelligence in medical imaging practice: Perspectives of African radiographers. *Radiography* **27**, 861–866 (2021).
32. Guo, J. & Li, B. The application of medical artificial intelligence technology in rural areas of developing countries. *Health Equity* **2**, 174–181 (2018).
33. Wani, S. U. D. et al. Utilization of Artificial Intelligence in disease prevention: diagnosis, treatment, and implications for the healthcare workforce. *Healthcare* **10**, 608 (2022).
34. Zou, J. & Schiebinger, L. Ensuring that biomedical AI benefits diverse populations. *EBioMedicine* **67**, 103358 (2021).
35. Kaushal, A., Altman, R. & Langlotz, C. Health care AI systems are biased. *Sci. Am.* **11**, 17 (2020).
36. Data Science for Health Discovery and Innovation in Africa. *Our Projects* <https://dsi-africa.org/projects> (2023).
37. Consortium, H. A. et al. Research capacity. Enabling the genomic revolution in Africa. *Science* **344**, 1346–1348 (2014).
38. de Vries, J. et al. Addressing ethical issues in H3Africa research—the views of research ethics committee members. *HUGO J.* **9**, 1–4 (2015).
39. Bentley, A. R., Callier, S. L. & Rotimi, C. N. Evaluating the promise of inclusion of African ancestry populations in genomics. *NPJ Genom. Med.* **5**, 5 (2020).
40. Ozulumba, T. *Sustaining breakthroughs in health research in Africa*, <https://www.nature.com/articles/d44148-021-00124-y> (2021).
41. Millum, J., Grady, C., Keusch, G. & Sina, B. Introduction: the Fogarty International Research Ethics Education and Curriculum Development Program in historical context. *J. Empir. Res. Hum. Res. Ethics* **8**, 3–16 (2013).
42. Millum, J., Sina, B. & Glass, R. International research ethics education. *JAMA* **313**, 461–462 (2015).
43. Ndebele, P. et al. Research ethics capacity building in Sub-Saharan Africa: a review of NIH Fogarty-funded programs 2000–2012. *J. Empir. Res. Hum. Res. Ethics* **9**, 24–40 (2014).
44. Hummel, P., Adam, T., Reis, A. & Littler, K. Taking stock of the availability and functions of National Ethics Committees worldwide. *BMC Med. Ethics* **22**, 56 (2021).
45. Kohler, J., Reis, A. A. & Saxena, A. A survey of national ethics and bioethics committees. *Bull. World Health Organ.* **99**, 138–147 (2021).
46. Bedeker, A. et al. A framework for the promotion of ethical benefit sharing in health research. *BMJ Glob. Health* **7**, e008096 (2022).
47. Staunton, C. & de Vries, J. The governance of genomic biobank research in Africa: reframing the regulatory tilt. *J. Law Biosci.* **7**, lsz018 (2020).
48. Chaudhry, I. et al. Strengthening ethics committees for health-related research in sub-Saharan Africa: a scoping review. *BMJ Open* **12**, e062847 (2022).
49. Sudoi, A., De Vries, J. & Kamuya, D. A scoping review of considerations and practices for benefit sharing in biobanking. *BMC Med. Ethics* **22**, 102 (2021).
50. Moodley, K. et al. Ethics and governance challenges related to genomic data sharing in southern Africa: the case of SARS-CoV-2. *Lancet Glob. Health* **10**, e1855–e1859 (2022).
51. Metcalf, J. & Crawford, K. Where are human subjects in Big Data research? The emerging ethics divide. *Big Data Soc.* **3**, 2053951716650211 (2016).
52. Ada Lovelace Institute. Looking before we leap: Ethical review processes for AI and data science research. (Ada Lovelace Institute, Exeter, UK, 2022).
53. Ferretti, A., Ienca, M., Velarde, M. R., Hurst, S. & Vayena, E. The Challenges of Big Data for Research Ethics Committees: A Qualitative Swiss Study. *J. Empir. Res. Hum. Res. Ethics* **17**, 129–143 (2022).
54. Mahomed, S. & Labuschaigne, M. L. The evolving role of research ethics committees in the era of open data. *S. Afr. J. Bioethics Law* <https://doi.org/10.7196/SAJBL.2022.v15i3.XX> (2023).
55. Silaigwana, B. & Wassenaar, D. Biomedical Research Ethics Committees in Sub-Saharan Africa: A Collective Review of Their Structure, Functioning, and Outcomes. *J. Empir. Res. Hum. Res. Ethics* **10**, 169–184 (2015).
56. Bain, L. E., Ebuenyi, I. D., Ekukwe, N. C. & Awah, P. K. Rethinking research ethics committees in low- and medium-income countries. *Res. Ethics* **14**, 1–7 (2018).
57. National Health Research Ethics Committee of Nigeria. (Federal Ministry of Health of Nigeria, Abuja, 2007).
58. Leonelli, S. Locating ethics in data science: responsibility and accountability in global and distributed knowledge production systems. *Philos. Trans. R. Soc. A: Math., Phys. Eng. Sci.* **374**, 20160122 (2016).
59. Terry, R., Littler, K. & Olliaro, P. Sharing health research data? the role of funders in improving the impact [version 2; peer review: 3 approved]. *F1000Research* <https://doi.org/10.12688/f1000research.16523.2> (2018).
60. Greenleaf, G. & Cottier, B. Data privacy laws and Bills: Growth in Africa, GDPR influence. *GDPR Influence* **152**, 11–13 (2018).
61. Daigle, B. Data protection laws in Africa: a pan-African survey and noted trends. *J. Int'l Com. Econ.* **1**, 1–27 (2021).
62. McCall, B. What does the GDPR mean for the medical community? *Lancet* **391**, 1249–1250 (2018).
63. Irish Council for Civil Liberties. 5 years: GDPR's crisis point, <https://www.iccl.ie/wp-content/uploads/2023/05/5-years-GDPR-crisis.pdf> (2023).
64. Mann, S. P., Savulescu, J., Ravaud, P. & Benchoufi, M. Blockchain, consent and present for medical research. *J. Med. Ethics* **47**, 244–250 (2021).
65. Morley, J. et al. The ethics of AI in health care: a mapping review. *Soc. Sci. Med.* **260**, 113172 (2020).
66. Kalenzi, C. Artificial Intelligence and Blockchain: How Should Emerging Technologies Be Governed? *Front. Res. Metrics Analytics* <https://doi.org/10.3389/frma.2022.801549> (2022).
67. Global Health Network. *Global Health Data Science*, <https://globalhealthdatascience.tghn.org/> (2023).
68. Bridging Gaps in the ELSI of Data Science Research (BridgELSI) Project. 21 May, <https://bioethicscenter.net/bridgelsi-project/> (2023).
69. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. (2021).
70. World Health Organization. *Health Data as a global public good – a call for Health Data Governance* 30 September, <https://www.who.int/news-room/articles-detail/health-data-as-a-global-public-good-a-call-for-health-data-governance-30-september> (2021).

71. Edwards, T. L. et al. Challenges and Opportunities for Data Science in Women's Health. *Annu. Rev. Biomed. Data Sci.* <https://doi.org/10.1146/annurev-biodatasci-020722-105958> (2023).
72. Mills, M. C. & Rahal, C. The GWAS Diversity Monitor tracks diversity by disease in real time. *Nat. Genet.* **52**, 242–243 (2020).
73. Richardson, S., Lawrence, K., Schoenthaler, A. M. & Mann, D. A framework for digital health equity. *npj Digital Med.* **5**, 119 (2022).
74. Lyles, C. R., Nguyen, O. K., Khoong, E. C., Aguilera, A. & Sarkar, U. Multilevel determinants of digital health equity: a literature synthesis to advance the field. *Annu. Rev. Public Health* **44**, 383–405 (2023).
75. Brewer, L. C. et al. Back to the future: achieving health equity through health informatics and digital health. *JMIR Mhealth Uhealth* **8**, e14512 (2020).
76. Kordzadeh, N. & Ghasemaghaei, M. Algorithmic bias: review, synthesis, and future research directions. *Eur. J. Inf. Syst.* **31**, 388–409 (2022).
77. Yu, P. K. The algorithmic divide and equality in the age of artificial intelligence. *Fla. L. Rev.* **72**, 331 (2020).
78. Panch, T., Mattie, H. & Atun, R. Artificial intelligence and algorithmic bias: implications for health systems. *J. Glob. Health* **9**, 010318 (2019).
79. Thomasian, N. M., Eickhoff, C. & Adashi, E. Y. Advancing health equity with artificial intelligence. *J. Public Health Policy* **42**, 602–611 (2021).
80. de Hond, A. A., van Buchem, M. M. & Hernandez-Boussard, T. Picture a data scientist: a call to action for increasing diversity, equity, and inclusion in the age of AI. *J. Am. Med. Inform. Assoc.* **29**, 2178–2181 (2022).
81. Leavy, S. in *Proc. 1st International Workshop on Gender Equality in Software Engineering*. (eds E. Abraham, E. Di Nitto, & R. Mirandola) 14–16 (Association for Computing Machinery, 2018).
82. Gwagwa, A., Kraemer-Mbula, E., Rizk, N., Rutenberg, I. & De Beer, J. Artificial Intelligence (AI) deployments in Africa: benefits, challenges and policy dimensions. *Afr. J. Inf. Commun.* **26**, 1–28 (2020).
83. Park, Y. et al. in *Healthcare Information Management Systems: Cases, Strategies, and Solutions* (eds Joan M. Kiel, George R. Kim, & Marion J. Ball) 223–234 (Springer International Publishing, 2022).
84. Dankwa-Mullan, I. & Weeraratne, D. Artificial intelligence and machine learning technologies in cancer care: addressing disparities, bias, and data diversity. *Cancer Discov.* **12**, 1423–1427 (2022).
85. United Nations Sustainable Development Group. Data Privacy, Ethics and Protection. Guidance Note on Big Data for Achievement of the 2030 Agenda. 19 (United Nations, Geneva, Switzerland, 2017).
86. Katell, M. et al. in *FAT*20: Proceedings of the 2020 conference on fairness, accountability, and transparency*. (eds M. Hildebrandt et al.) 45–55 (Association for Computing Machinery, New York, NY, 2020).
87. Embi, P. J. Algorithmic vigilance—advancing methods to analyze and monitor artificial intelligence-driven health care for effectiveness and equity. *JAMA Netw. Open* **4**, e214622–e214622 (2021).

Acknowledgements

This work is supported by the following grants: 1. Bridging Gaps in the ELSI of Data Science Health Research in Nigeria (BridgELSI, U01 MH127693) to CAA and TO. 2. The Maryland Department of Health's Cigarette Restitution Fund Program (CH-649-CRF) to CAA and SNA. 3. The University of Maryland Greenebaum Comprehensive Cancer Center Support Grant (P30CA134274) to CAA and SNA. The funding agencies did not play any role in the publication.

Author contributions

C.A.A. conceived the study, led the co-authors in writing and revising the manuscript, read all papers and documents referenced in the manuscript to ensure correct interpretation and finalized the version submitted for publication. S.C. co-wrote the paper, revised the manuscript, and approved the final version. S.A., O.M., A.J., C.A., T.O., and S.N.A. co-wrote, read and critiqued several drafts of the manuscript and contributed to development and refinement of the ideas in this paper. They reviewed and approved the final draft of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Clement A. Adebamowo.

Peer review information *Nature Communications* thanks Robert Terry, and the other anonymous, reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

BridgELSI Project as part of the DS-I Africa Consortium

Clement A. Adebamowo ^{1,2}✉, Shawneequa Callier ^{3,4}, Simisola Akintola ^{2,5,6}, Oluchi Maduka ², Ayodele Jegede^{2,6,7}, Christopher Arima⁸, Temidayo Ogundiran ^{2,6,9} & Sally N. Adebamowo ^{1,2}