Ethics of AI in global health research

Cape Town, 29&30 November 2022

Global Forum on Bioethics in Research

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.

This paper was prepared for GFBR 2022. Further details on the meeting are available at www.gfbr.global.