



HHS Public Access

Author manuscript

Am J Bioeth. Author manuscript; available in PMC 2023 July 01.

Published in final edited form as:

Am J Bioeth. 2022 July ; 22(7): 30–33. doi:10.1080/15265161.2022.2075051.

Automating justice: an ethical responsibility of computational bioethics

Dr Vasiliki Rahimzadeh, PhD,

Stanford University, Stanford Center for Biomedical Ethics, Edwards Research Building, 300 Pasteur Dr, 2nd Floor, Stanford, 94305-6104 United States

Mr Jonathan Lawson,

Broad Institute, Cambridge, 02141-2023 United States

Miss Jinyoung Baek,

Broad Institute, Cambridge, 02141-2023 United States

Mr Edward S. Dove

University of Edinburgh, School of Law, Edinburgh, United Kingdom of Great Britain and Northern Ireland

In their proof-of-concept, Meier and colleagues describe the purpose and programming decisions underpinning Medical Ethics Advisor (METHAD), an automated decision support (ADS) system used to guide treatment interventions. The authors themselves note they are not the first to propose that machines could help humans make better, if not more ethical, decisions. Indeed, Alan Turing commented during a 1951 talk broadcast by the BBC that an “attempt to make a thinking machine will help us greatly in finding out how we think ourselves.” Like the authors, we are among a growing community of ‘computational bioethicists’ interested in how thinking machines may be leveraged to yield faster, more consistent, and potentially fairer decisions in clinical and research contexts. Unlike Meier and colleagues, however, we reject the notion that developers can sidestep justice and the impacts that competing justice claims have on decisional outcomes at the ADS design stage, pilot testing, or implementation. In this Open Peer Commentary, we engage with three central ideas about justice: (i) ADS should be used to support, not supplant human decision-making as a matter of algorithmic justice; (ii) gains in procedural justice are among the strongest rationales to pursue research and development of ADS for ethical decision-making, and finally the (iii) the values and priorities of stakeholders, e.g. patients, families and communities actually affected by decisions, should comprise the data inputs upon which ADS are ultimately trained. We draw on our collective work developing and implementing ADS for responsible data access management as members of the Global Alliance for Genomics and Health (GA4GH), and propose an agenda for future empirical work needed to advance the subdiscipline of computational bioethics with justice at the forefront.

(Corresponding Author) vrahim@stanford.edu.

Disclosure Statement

All authors are members of the Regulatory and Ethics Workstream of the Global Alliance for Genomics and Health.

Justice as a central consideration in ADS development and implementation

Meier and colleagues confirm “Like any ethical judgments taken by humans, ethical algorithmic decision-making must be rooted in a moral framework.” They adopt principlism for its generalizability across diverse cases and because it may be feasibly “translated into machine-readable values.” But Meier and colleagues opted not to incorporate the principle of justice into their algorithm for fear it would require making “specific, and possible unwarranted, health-political and socio-economical background assumptions.”

Though we understand the authors’ rationale, we consider this a missed opportunity. Any algorithmic proposal that omits consideration of substantive and procedural dimensions of justice is unethical *posteriori*. Put differently, socio-economic, political, and historical considerations are, in our view, inextricably intertwined with the other relevant set of decision factors that yield ethical judgments. Systemic barriers in access to critical care services, vaccines, and other necessary pandemic relief exemplify how distributive and social injustices compound in an inherently unequal healthcare system within which an algorithm like METHAD would expect to operate.

Stakeholder perspectives and experiences as necessary data inputs

We also worry that the authors’ discussion of the limitations of METHAD may understate the complexity of public perceptions around machine- versus human-generated decisions in healthcare, which some studies suggest correlate strongly with perceived level of trust in human systems (Lee and Rich 2021). How reliant will patients and providers become on decision outputs METHAD generates? There may be a particular risk that people will rely on ADS more than they should. Biased risk-assessment tools used in the criminal justice (Angwin et al. 2016) and health record systems in the U.S. (Gianfrancesco et al. 2018) offer cautionary lessons from which we must all learn moving forward in the era of intelligent machines. Given such wide variation in public perceptions, understanding why and how people would/not incorporate algorithmic advice into their decisions is a necessary precursor to broader adoption. Behavioral economists have argued, for instance, that “people draw on the very first piece of evidence at their disposal, however weak, when making subsequent decisions”, otherwise known as “the anchoring bias”, and there is some evidence this can apply to algorithms (Bessner and Guilhot 2019, 287).

Just as we consider matters of substantive and procedural justice to be necessary components of algorithmic development, so too should humans, or institutional leaders (Green 2022), remain the central actor from ADS design and training to implementation to evaluation. Meier and colleagues rightfully suggest that humans play a critical role when it comes to making decisions on the range of moral dilemmas in clinical care. Yet, we also wish to stress that ethical judgments made solely by humans or with facilitative algorithms may well entail a confluence of multiple ethics frameworks which yield internally or externally inconsistent outputs. Humans are, after all, irrational even when they are predictable. And as Green (2022) recently argues that incorrect assumptions about effective human oversight legitimizes the use of flawed and unaccountable algorithms in government, but perhaps in other sectors too. So, if quality controls for programs like METHAD rely on comparison

to decisions humans would make given the same inputs and contexts – which we broadly support– such programs would also need to account for multiplex, context-dependent factors relevant to making decisions in response to moral dilemmas.

For these reasons we disagree with Meier and colleagues that there will come a time when “machine intelligence has become efficient, accurate, and transparent enough to in fact *replace* human ethical-decision making in certain settings,” from normative as well as engineering standpoints. Even in “overwhelming emergency situations where greater numbers of morally relevant decisions must be taken than would be humanly possible,” we doubt the merit and effectiveness of a machine intelligence system to enhance justice without human input, much less oversight (Shaw 2022).

ADS strengths for enhancing substantive and procedural justice

Our position on the criticalness of justice in developing and implementing ADS is supported by our active investigations of ADS applications in managing access to genomic and related health data. By making data use terms and permissions machine readable, as the GA4GH Data Use Oversight System (DUOS) enables data contributors to code to do (Lawson et al. 2021) data stewards and access committees (DACs) of genomic and health-related data are able to adjudicate decision consistently and with the support of algorithms, as is possible with the DUOS (Cabili et al. 2021). In the future, we anticipate DUOS will offer DACs - who are comprised of humans decision-makers - the ability to configure automated decision-making given select inputs, while still placing the onus of oversight and control in human hands. Development of DUOS and similar automated workflows are a direct response to growing calls for alternative models of data stewardship to address decision inconsistencies and delays in data sharing.

ADS applied to ethical data governance also aligns with and, in some cases, activates rights afforded to individuals when appealing to procedural bodies in other legal contexts. These include the right to due process, a fair hearing, and the right to appeal. While algorithms may be calibrated differently based on the compliance standards and regulations they have been programmed to execute, the process points remain the same. For these reasons, we argue that the case for developing and implementing ADS in clinical and research settings is strongest when ethical decisions rest on procedural integrity or require regulatory compliance.

ADS frameworks implemented in tools such as DUOS enable at least three procedural justice goals. First, DUOS provides standard forms for submission and review of data access requests using standardized use terms that are machine readable. This ensures each applicant receives an equivalent review for the same standard set of inputs.

Second, DUOS actualizes the right to a fair hearing because it applies standard applicant validation and authentication. With these fields verified, the data access committee concentrates on ensuring proposed data uses comply with the machine-readable terms extracted from participant consents. This ADS-based process for reviewing data access

requests avoids discrepant interpretation of data use terms that lead to inconsistent data access decisions.

Third, the right to appeal is aided by DUOS' clarity in decision making. With uniform interpretation of data use terms using standard ontologies, investigators will be able to predict when and why their requests are likely to be denied even before a committee renders their decision. A data requestor is then able to appeal any decision with the data custodian or contact the DUOS administrators to inquire further.

Advancing the subdiscipline of computational bioethics

Our reliance on raw data to better understand the biology of human health and disease, as well as on computing systems to help guide policy decisions that improve population-level health outcomes, motivated the info-computational turn in bioethics (Vic 2019; Schneider et al. 2021). Computational bioethics borrows epistemological commitments that ground fields such as computational sociology and computational law. It is a growing subdiscipline dedicated to the theory and practice of applying fundamental principles of computer science to automate ethical reasoning in clinical care and research. Building on the contribution from Meier and colleagues, we propose that cross-cutting research is critically as sophisticated algorithms and machine learning applications are developed. Specifically, future research will need to investigate organizational factors that mediate responsible implementation of ADS from the perspectives of target end users and diverse stakeholders whom ADS-informed ethical decisions will principally affect. Interdisciplinary scholars, e.g. implementation scientists, software programmers, lawyers and ethicists, as well as empirical approaches are needed to achieve the above empirical research aims if ADS are to be deployed responsibly and incorporate justice as a core principle from the outset.

Funding details

This work was supported by the National Human Genome Research Institute under Grant 3U24HG010262-04S2.

References

- Angwin J, Larson J, Mattu S, and Kirchner L 2016. Machine Bias. ProPublica, May 23.
- Bessner D, and Guilhot N 2019. The Decisionist Imagination: Sovereignty, Social Science and Democracy in the 20th Century. Berghahn Books.
- Cabili MN, Lawson J, Saltzman A, Rushton G, O'Rourke P, Wilbanks J, Rodriguez LL, Nyronen T, Courtot M, Donnelly S, and Philippakis AA 2021. Empirical validation of an automated approach to data use oversight. *Cell Genomics* 1(2): 100031. doi: 10.1016/j.xgen.2021.100031.
- Gianfrancesco MA, Tamang S, Yazdany J, and Schmajuk G 2018. Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. *JAMA Internal Medicine* 178(11): 1544–1547. doi: 10.1001/jamainternmed.2018.3763. [PubMed: 30128552]
- Green B 2022. The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review* 45: 105681. doi: 10.1016/j.clsr.2022.105681.
- Lawson J, Cabili MN, Kerry G, Boughtwood T, Thorogood A, Alper P, Bowers SR, Boyles RR, Brookes AJ, Brush M, Burdett T, Clissold H, Donnelly S, Dyke SOM, Freeberg MA, Haendel MA, Hata C, Holub P, Jeanson F, Jene A, Kawashima M, Kawashima S, Konopko M, Kyomugisha I, Li H, Linden M, Rodriguez LL, Morita M, Mulder N, Muller J, Nagaie S, Nasir J, Ogishima S,

Ota Wang V, Paglione LD, Pandya RN, Parkinson H, Philippakis AA, Prasser F, Rambla J, Reinold K, Rushton GA, Saltzman A, Saunders G, Sofia HJ, Spalding JD, Swertz MA, Tulchinsky I, van Enkevort EJ, Varma S, Voisin C, Yamamoto N, Yamasaki C, Zass L, Guidry Auvil JM, Nyrönen TH, and Courtot M 2021. The Data Use Ontology to streamline responsible access to human biomedical datasets. *Cell Genomics* 1(2): 100028. doi: 10.1016/j.xgen.2021.100028.

Lee MK, and Rich K 2021. Who Is Included in Human Perceptions of AI?: Trust and Perceived Fairness around Healthcare AI and Cultural Mistrust. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–14. CHI '21. New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3411764.3445570.

Schneider M, Vayena E, and Blasimme A 2021. Digital bioethics: introducing new methods for the study of bioethical issues. *Journal of Medical Ethics* 0: 1–8.

Shaw J 2022. Emerging Paradigms for Ethical Review of Research Using Artificial Intelligence. *American Journal of Bioethics*(January 20).

Vic C 2019. The Info-Computational Turn in Bioethics. In *Contemporary Debates in Bioethics: European Perspectives*, ed. Mihailov E, Wangmo T, Federiuc V, and Elger B, 108–120. De Gruyter Open. doi: 10.2478/9783110571219-011.