The American Journal of Bioethics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uajb20

Automating Justice: An Ethical Responsibility of Computational Bioethics

Vasiliki Rahimzadeh, Jonathan Lawson, Jinyoung Baek & Edward S. Dove

To cite this article: Vasiliki Rahimzadeh, Jonathan Lawson, Jinyoung Baek & Edward S. Dove (2022) Automating Justice: An Ethical Responsibility of Computational Bioethics, The American Journal of Bioethics, 22:7, 30-33, DOI: 10.1080/15265161.2022.2075051

To link to this article: https://doi.org/10.1080/15265161.2022.2075051

Published online: 23 Jun 2022.

they did not explain why an AI driven tool is the right approach to aid ethical decision making.

It seems that METHAD is a classic example of technical sweetness. The authors have created a tool without appropriate prior reflection, and we are left wondering what benefit this tool could have (in the best-case scenario) or what harm such a tool could cause (in the worst-case scenario).

FUNDING

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

Aurelia Sauerbrei http://orcid.org/0000-0002-2831-1834

REFERENCES

AI HLEG. 2019. *Ethics guidelines for trustworthy AI*. Brussels: European Commission.

Anderson, M., S. L. Anderson, and C. Armen. 2006. An approach to computing ethics. IEEE Intelligent Systems 21 (4):56–63. doi:10.1109/MIS.2006.64.

Douglas, H. 2017. The bitter aftertaste of technical sweetness. In *Frankenstein*, ed. M. W. Shelley, D. H. Guston, E. Finn, and J. S. Robert, 247. Cambridge, MA: MIT.

McDougall, R. J. 2019. Computer knows best? The need for value-flexibility in medical AI. *Journal of Medical Ethics* 45 (3):156–60. doi:10.1136/medethics-2018-105118.

Meier, L. J., A. Hein, K. Diepold, and A. Buyx. 2022. Algorithms for ethical decision-making in the clinic: A proof of concept. *The American Journal of Bioethics* 22 (7):4–20. doi:10.1080/15265161.2022.2040647.

Szasz, F. M. 1984. *The day the sun rose twice: The story of the trinity site nuclear explosion, July 16, 1945*. 1st ed. Albuquerque: University of New Mexico.

Topol, E. J. 2019. High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine* 25 (1):44–56. doi:10.1038/s41591-018-0300-7.

Tucker, A. 2002. Obituary: Victor Weisskopf. *The Guardian*. https://www.theguardian.com/news/2002/apr/26/guardianobituaries.obituaries

World Health Organization. 2021. *Ethics and governance of artificial intelligence for health*. Geneva: WHO.

THE AMERICAN JOURNAL OF BIOETHICS
2022, VOL. 22, NO. 7, 30–33
https://doi.org/10.1080/15265161.2022.2075051

Taylor & Francis
Taylor & Francis Group

OPEN PEER COMMENTARIES

Automating Justice: An Ethical Responsibility of Computational Bioethics

Vasiliki Rahimzadeh[a], Jonathan Lawson[b], Jinyoung Baek[b], and Edward S. Dove[c]

[a]Stanford University; [b]Broad Institute; [c]University of Edinburgh

In their proof-of-concept, Meier and colleagues (2022) 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 (iii) the values and priorities of stakeholders, e.g., patients, families, and communities actually

CONTACT Vasiliki Rahimzadeh vrahim@stanford.edu Stanford Center for Biomedical Ethics, Stanford University, Stanford, CA 94305-6104, USA

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.

JUSTICE AS A 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 other relevant decisions 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 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 stress that ethical judgments humans make alone 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, 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 an intelligent machine or related ADS 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 such as the GA4GH Data Use Oversight System (DUOS) (Lawson et al. 2021), data access committees (DACs) and other institutional data stewards can more consistently adjudicate access decisions with the support of ADS and other computational tools (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 strengthened when ethical decisions rest on procedural integrity or require regulatory compliance.

DUOS and similar ADS tools used to manage human genomic and health-related data advance 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 (Vică 2019; Schneider, Vayena, and Blasimme 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 critical as sophisticated algorithms and machine learning applications are developed. Specifically, future research should focus on investigating 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, will achieve the above research aims if ADS are to be deployed responsibly and incorporate justice as a core principle from the outset.

DISCLOSURE STATEMENT

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

FUNDING

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

ORCID

Vasiliki Rahimzadeh http://orcid.org/0000-0003-3537-7601
Jonathan Lawson http://orcid.org/0000-0003-3593-1289
Jinyoung Baek http://orcid.org/0000-0002-8756-9931
Edward S. Dove http://orcid.org/0000-0003-4095-3677

REFERENCES

Angwin, J., J. Larson, S. Mattu, and L. Kirchner. 2016. Machine bias. ProPublica, May 23.

Bessner, D., and N. Guilhot. 2019. *The decisionist imagination: Sovereignty, social science and democracy in the 20th century*. New York; Oxford: Berghahn Books.

Cabili, M. N., J. Lawson, A. Saltzman, G. Rushton, P. O'Rourke, J. Wilbanks, L. L. Rodriguez, T. Nyronen, M. Courtot, S. Donnelly, et al. 2021. Empirical validation of an automated approach to data use oversight. *Cell Genomics* 1 (2):100031. doi:10.1016/j.xgen.2021.100031.

Gianfrancesco, M. A., S. Tamang, J. Yazdany, and G. Schmajuk. 2018. Potential biases in machine learning algorithms using electronic health record data. *JAMA Internal Medicine* 178 (11):1544–7. doi:10.1001/jamainternmed.2018.3763.

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., M. N. Cabili, G. Kerry, T. Boughtwood, A. Thorogood, P. Alper, S. R. Bowers, R. R. Boyles, A. J. Brookes, M. Brush, et al. 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, M. K., and K. Rich. 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*, CHI '21, 1–14. New York, NY: Association for Computing Machinery. doi:10.1145/3411764.3445570.

Meier, L. J., A. Hein, K. Diepold, and A. Buyx. 2022. Algorithms for ethical decision-making in the clinic: A proof of concept. *The American Journal of Bioethics* 22 (7):4–20. doi:10.1080/15265161.2022.2040647.

Schneider, M., E. Vayena, and A. Blasimme. 2021. Digital bioethics: Introducing new methods for the study of bioethical issues. *Journal of Medical Ethics* 0: Medethics-2021-107387–8. doi:10.1136/medethics-2021-107387.

Shaw, J. 2022. Emerging paradigms for ethical review of research using artificial intelligence. *The American Journal of Bioethics*.

Vică, C. 2019. The info-computational turn in bioethics. In *Contemporary debates in bioethics: European perspectives*, ed. E. Mihailov, T. Wangmo, V. Federiuc, and B. Elger, 108–120. Warschau/Berlin: De Gruyter Open. doi:10.2478/9783110571219-011.

THE AMERICAN JOURNAL OF BIOETHICS
2022, VOL. 22, NO. 7, 33–35
https://doi.org/10.1080/15265161.2022.2075969

OPEN PEER COMMENTARIES

An All-Too-Human Enterprise

Tod Chambers

Northwestern University Feinberg School of Medicine

On reading "Algorithms for Ethical Decision-Making in the Clinical: A Proof of Concept," (Meier et al. 2022) I imagined that for some the fundamental problem with the authors' approach is the very idea that a machine could reproduce the workings of the human mind. There are those like Hubert Dreyfus and Stuart Dreyfus who argue that while a computer can beat a human in a game of chess, the computer is not "thinking" when playing in the same way a human does; a form of intelligence has been created, they grant, but that form is not the same as *human* intelligence (1986). Others might counter that, as with a medical diagnosis, what matters in the end is not how one reached the diagnosis or triumphed over the opponent but simply that the correct answer was found and the game was won. Do I really care what kind of intelligence was able to figure out the cause of a series of mysterious symptoms?

Similarly, does it really matter how one resolves a moral problem, or what kind of intelligence has come to conclude which moral principle has greater weight? There are a number of virtues in having artificial intelligence "take over" spheres of our lives. Often the computer can produce answers in a way that is faster and more consistent than human intelligence, and in their conclusion the authors identify precisely this virtue when they describe the likely applications of METHAD. But even if we assume that their algorithm

CONTACT Tod Chambers t-chambers@northwestern.edu Feinberg School of Medicine, Northwestern University, Chicago, IL, USA.