



## AI Methods in Bioethics

Joshua August Skorburg, Walter Sinnott-Armstrong & Vincent Conitzer

To cite this article: Joshua August Skorburg, Walter Sinnott-Armstrong & Vincent Conitzer (2020) AI Methods in Bioethics, AJOB Empirical Bioethics, 11:1, 37-39, DOI: [10.1080/23294515.2019.1706206](https://doi.org/10.1080/23294515.2019.1706206)

To link to this article: <https://doi.org/10.1080/23294515.2019.1706206>



Published online: 25 Feb 2020.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

COMMENTARY



## AI Methods in Bioethics

Joshua August Skorburg, Walter Sinnott-Armstrong, and Vincent Conitzer

Duke University, Durham, North Carolina, USA

Seemingly every week, new artificial intelligence (AI) applications are being developed to assist and automate various forms of medical decision-making. Examples abound, but some notable instances include robot-assisted surgery, precision medicine, drug discovery and drug interactions, assessment of suicide risk from electronic health records or social media posts, or automated diagnosis on the basis of genetic sequencing, images, speech, text, and even mouse cursor movements.

In parallel, a new wave of scholarship in bioethics is exploring the ethical, legal, and social implications of these AI applications in medicine. This work focuses, among other areas, on how to protect patient privacy in the face of the large-scale data collection required to train AI systems; how AI applications can reproduce and exacerbate existing biases, and also create new forms of inequality; and how increasing reliance on AI technologies may render healthcare less patient-centered. Again, examples abound, but some high-profile work in this vein has revealed racial biases in neural networks trained to identify skin cancer (Adamson and Smith 2018). Related research raises ethical concerns about how new tools, such as automated speech-based diagnosis of psychopathology, may perform poorly on non-native English speakers. Similar work considers the potentially diminished role of patients' subjective experiences, in light of prolific data-mining of health records (Ruckenstein and Schüll 2017).

Crucial as these concerns are, far less attention has been paid to the role that AI applications might play in bioethics itself. Might AI improve bioethical inquiry? Could AI bolster methods in empirical bioethics? In this commentary, we will argue that the answers to these questions are a cautiously optimistic “yes” and that empirical bioethicists' engagement with AI need not be limited to ethical, legal, and social implications.

To see this, consider kidney transplants. There are regrettably not enough donors to supply kidneys to all of the patients in need. This raises a moral problem. Who, among many needy patients, should receive a kidney when one becomes available? These decisions are often based on features such as compatibility, age, and time on the waiting list. But there are longstanding debates in bioethics about whether other features should be considered, and which ethical principles ought to guide decisions about kidney allocation (Childress 1989).

To the extent that kidney allocation is increasingly determined by algorithms, these questions will demand answers in the form of design choices about what such algorithms are optimizing for, which features are included or excluded, and how they are weighted. These ethical issues regarding design choices are right in the empirical bioethicist's wheelhouse.

Indeed, we have developed a method (Freedman et al. 2018) for the case of kidney transplants that we think can also generalize to other issues in bioethics. In a forthcoming paper, we call this method Artificial Improved Democracy (AID) (Sinnott-Armstrong and Skorburg forthcoming).

The first step in this method is to ask experts and laypeople which features they think ought and ought not figure into decisions about kidney allocations. Then, after editing responses for clarity, redundancy, etc., we can use this curated list of features to construct forced-choice scenarios in which one (or a set of) valued feature(s) conflicts with another (or a set of) valued feature(s).

For example, our results suggest that most laypeople do think that age and time on the waiting list should be considered. Most people also think that race and religion should not be considered. But features such as mental health and criminal records are more controversial (Doyle et al. *in prep*).

In a controlled experimental setting, we can then ask, “who should get the kidney?” when Patient A is 34 years

old, with two dependents, and a history of consuming three alcoholic beverages a day and Patient B is 46 years old with one dependent and no history of alcohol consumption. After running hundreds and thousands of these choice trials, we can then use machine learning techniques to reveal which of the conflicting features *really do* seem to drive people's decisions about kidney allocations, as well as how different features interact with one another to produce this decision. On the basis of these findings, the third step in our method is to build models to predict individual-level and group-level decisions. In turn, these predictive models can be iteratively improved by being applied to new scenarios.

This method has several advantages. First, we can compare (in experts and laypeople alike) which features people think *should* be considered in allocating kidneys. Such information could reveal important differences between, for example, the values of doctors or hospital administrators on the one hand, and the values of patients or community stakeholders, on the other. To the extent that participatory research and patient-centered care (Department of Health 2009) are guiding ideals, the data collected in our studies can be used to inform policies to better align healthcare services with the values of the patients and communities utilizing them.

Second, we can compare which features people say *should* guide allocation decisions with which features *actually* seem to guide their (hypothetical) allocation decisions. When there is a gap between them, we can update the model to more closely approximate consensus values. For example, if people consistently say that race or religion should not be considered when deciding who gets a kidney, yet in their allocation decisions these features do seem to play a role (as reflected in our models of them), then we can update the models to no longer take these features into account in order to reflect the consensus.

Third, our method helps to shed light on important mechanisms of medical decision-making, yielding both descriptive and explanatory accounts. The more accurate such models become, the more we can understand the processes driving decisions about kidney allocations. And the more we understand about these processes, the better we will be able to guide design choices to build systems that instantiate both actual human values and more ideal decision procedures.

For these reasons, we think there is a strong case for using AI methods in empirical bioethics. The example of kidney allocation is meant to serve as proof of concept. We think the components of this method (first, design surveys to discern morally relevant features;

second, construct forced-choice conflict scenarios with these features; third, iterate predictive models on the basis of choice data) can generalize to other issues in empirical bioethics. This is clear enough for closely related cases, such as liver, lung, or heart transplant programs. But the method could also be used in cases involving other limited medical resources, such as scarce hospital beds or clinicians' time in critical care settings. We can imagine other applications in decisions involving life-sustaining treatments, experimental therapies, emergency medicine, or end-of-life issues.

To be sure, there are limitations, challenges, and difficulties which we have not addressed in this short commentary. But hopefully the foregoing has demonstrated that empirical bioethicists' engagement with AI technologies need not be limited to ethical, legal, and social implications which have tended to dominate the recent literature. Indeed, if our research program proves fruitful, then these emerging AI tools will open new modes of empirical and normative inquiry—tasks for empirical bioethicists if there ever were any.

### Author contributions

JAS, WSA, and VC all contributed to the conception, drafting, and editing of the manuscript.

### Conflicts of interest

None.

### Ethical approval

The studies referenced in this manuscript were approved by the institutional review board at Duke University.

### References

- Adamson, A. S., and A. Smith. 2018. Machine learning and health care disparities in dermatology. *JAMA Dermatology* 154 (11):1247–8. doi: 10.1001/jamadermatol.2018.2348.
- Childress, J. F. 1989. Ethical criteria for procuring and distributing organs for transplantation. *Journal of Health Politics, Policy and Law* 14 (1):87–113. doi: 10.1215/03616878-14-1-87.
- Department of Health. 2009. *Putting people at the heart of care*. Retrieved from [https://webarchive.nationalarchives.gov.uk/20130123200554/http://www.dh.gov.uk/en/Publicationandstatistics/Publications/PublicationsPolicyAndGuidance/DH\\_106038](https://webarchive.nationalarchives.gov.uk/20130123200554/http://www.dh.gov.uk/en/Publicationandstatistics/Publications/PublicationsPolicyAndGuidance/DH_106038)
- Doyle, et al. (in prep). Which features of patients should determine who gets a kidney?

- Freedman, R., J. S. Schaich Borg, W. Sinnott-Armstrong, J. P. Dickerson, and V. Conitzer. 2018. April. Adapting a kidney exchange algorithm to align with human values. In Thirty-Second AAAI Conference on Artificial Intelligence. doi: [10.1145/3278721.3278727](https://doi.org/10.1145/3278721.3278727).
- Ruckenstein, M., and N. D. Schüll. 2017. The datafication of health. *Annual Review of Anthropology* 46 (1):261–78. doi: [10.1146/annurev-anthro-102116-041244](https://doi.org/10.1146/annurev-anthro-102116-041244).
- Sinnott-Armstrong, W., and J.A. Skorburg. (forthcoming). How AI can AID bioethics. *Journal of Practical Ethics*.