MEDICINE AND SOCIETY

Debra Malina, Ph.D., *Editor*

Algorithm-Aided Prediction of Patient Preferences — An Ethics Sneak Peek

Nikola Biller-Andorno, M.D., Ph.D., and Armin Biller, M.D.

The use of artificial intelligence (AI) is increasingly common in medicine, as it is in other fields. Just as AI systems using natural language processing answer legal questions, predict court decisions, and prepare contracts,[1] and as self-driving cars gather traffic information so they can make and act on decisions,[2] diagnostic algorithms already outperform radiologists and dermatologists,[3] estimate life expectancy,[4,5] and detect health risks.[6] Health care AI alone is expected to become a $20 billion market within the next 5 years.[7] Training algorithms to predict people's advance health care choices seem to be in keeping with the stream of emerging applications, and it would be surprising if we had long to wait before products based on such technology became commercially available. For instance, natural language processing of electronic health records (EHRs) can be used to identify patients with indications for palliative interventions almost as well as human classifiers can, but in a fraction of the time.[8]

Advance decisions about such matters as do-not-attempt-resuscitation (DNAR) status, organ donation, and curative versus palliative care are based on individual preferences and values and frequently reflect moral choices. They are notoriously difficult for others to make for the person in question, especially if no clear instructions have been provided through an advance directive or care plan. It is a well-known problem in clinical ethics that surrogates may make decisions that are inconsistent with the person's preferences and values.[9] Readily available intelligent decision support might improve the process, which must often take place in suboptimal conditions involving time constraints, stress, unclear advance directives, unavailability of surrogates, personal bias, or conflicts of interest.

We describe three potential applications in the boxes. An AI-based "DNAR Predictor" might be useful for physicians who have to make rapid decisions about patients they don't know. The hypothetical "Am I an Organ Donor?" app could foster reflection and decision making concerning personal choices such as organ donation. A "Be the Best Surrogate You Can Be" app might help a relative who is not well prepared for the role of surrogate. All these examples are thought experiments at this point, but we believe that similar applications will be developed in the near future. Should they prove to be useful, reliable, and convenient, they might easily become standard tools with widespread use. Advising surrogates and clinicians rather than replacing individual choice, they could be seen as enhancing patient autonomy and augmenting decision making.[10]

The DNAR Predictor

Ms. X. collapses on the street, and an ambulance transports her to the nearest hospital. In the emergency department, her condition remains unstable. Dr. A. looks up her file in the EHR used by all the town's health care institutions and discovers that she's in her late 70s, has several serious chronic conditions, and was diagnosed with cancer in the past year. He finds no DNAR order, Physician's Orders for Life-Sustaining Treatment (POLST) form, or similar document, no advance directive, and no information about a legal representative or a family doctor.

Dr. A. runs the DNAR Predictor program that was recently installed on the hospital system. It uses an algorithm that was trained on the data and DNAR preferences of a large number of patients and compares the available information about Ms. X. with the profiles of other patients to determine the likelihood that she would have wanted a resuscitation attempt. Dr. A. informs the team about the results of his search. A few moments later, Ms. X. goes into cardiac arrest.

Am I an Organ Donor?

In response to a challenge in his college sociology class, Mr. Y. is trying to figure out whether or not he wishes to be an organ donor, but he's having trouble reaching a conclusion. On his professor's advice, he downloads the "Am I an organ donor?" app on his smartphone, opens it, and answers a number of questions about his age, lifestyle, ethnicity, core values, and other topics. He allows the app to link to his Instagram account and his contact list. The app compares his profile with the profiles and donation decisions in its large database. After a few seconds, Mr. Y. is told that he's unlikely to want to donate his organs (the odds against it are 78%). Slightly taken aback, since he had considered himself an altruistic person, Mr. Y. ponders what the results mean for his class assignment and, more important, for his decision about obtaining a donor card.

Be the Best Surrogate You Can Be

Fifty-year-old Ms. Z. has had a massive stroke and is in the neurologic intensive care unit. The medical team has contacted her family to discuss how best to proceed, given that even if Ms. Z. survived with maximal therapy, she would probably be severely disabled for the rest of her life. The alternative is to withdraw life support and let her die.

Ms. Z.'s relatives are at a loss. Should they let her go without having tried everything? But maybe keeping her alive is worse than death? Before the stroke, she was quite active, energetically pursuing many social projects in her community. None of her relatives have talked to her about what she would want in a situation like this. She had seemed healthy and fit and didn't believe in "pre-worrying." The family was not aware of an advance directive and didn't think one existed.

One of Ms. Z.'s nephews, a lawyer, recalled seeing a report on an app that was designed for exactly this situation. The report had drawn on empirical evidence in arguing that surrogates are frequently biased by their own preferences or make wrong assumptions about their relatives' values. AI could help eliminate bias and predict whether, under specified circumstances, the relative would have preferred palliative care over continued therapeutic efforts. The nephew suggests that the family feed information about Ms. Z. into the system and see what it recommends.

We also know, however, that the use of algorithms raises ethical worries, particularly when life-or-death decisions are concerned.[3] One objection relates to the individual nature of the decisions to be made. Resuscitation and organ-donor status seem like highly personal choices informed by one's life experiences and values. What import could an algorithm possibly have? It's unlikely that an AI-based prediction of a given person's choice will ever match ground truth completely, since unforeseeable considerations and inconsistencies may come into play. But even a prediction that has an 80% or 90% likelihood of being accurate might be valuable information. How good such predictions can get is an open question right now. The richer the individual-level data available for training, the better the prediction is likely to be. We may find that our supposedly highly individual responses depend in a fairly predictable way on our physical and psychological situation, socialization, previous experiences, and values.

BEYOND HUMAN LIMITS

It could be argued that algorithms trained on vast amounts of individual-level data are unwieldy or even superfluous. Who needs an algorithm to suggest the same decisions people would make themselves? Such a function might become critical, however, when choices have to be made, for instance, regarding continued life support for someone who can no longer make decisions.

Algorithms would not only be able to find patterns within our own past decision making but could also compare them to patterns and decisions of many other people. The "wisdom of crowds"[11] could thus be harnessed for individual decision making. This possibility raises a number of issues. For one, algorithms can become only as good as the data on which they're trained. If the available human decisions — on resuscitation status, for instance — are not well informed, the algorithm will perpetuate the results of bad decision making. Knowledge about the quality of the process and, even better, the outcomes of the decisions, as well as about the overall quality of the training data, will be important if we are to have a realistic view of the predictions' value.

A second, no less important, issue is the risk of bias.[12] Since most AI approaches require a huge amount of training data input to produce meaningful output, the preferences of "crowds" will be used to train programs. As a consequence, even as preferences are learned, human bias — certain attitudes — will be introduced into the training data. Then, when the AI program is applied to a population that holds different attitudes, it may make inaccurate predictions but do so with a high level of confidence. As a conse-

quence, there is a need to audit whether an algorithm is qualified to perform a specific task in a specific population. Discussion of the capabilities and limitations of detecting and managing algorithmic bias is ongoing in the AI world.[13,14]

But if training data were no longer needed, the specter of bias would disappear. Progress is being made on this front: Google's AutoML, for instance, taught itself to develop machine learning programs, and the code it generated scored higher in a task of localizing multiple objects than the code written by humans. DeepMind's algorithm AlphaGo Zero taught itself how to play the game Go, given only the basic game rules.[15] It scored better than its predecessor, AlphaGo, and defeated the world's best human player. David Silver, AlphaGo Zero's lead programmer, says: "By not using human data — by not using human expertise in any fashion — we've actually removed the constraints of human knowledge."[16]

TECHNICAL AND PRACTICAL QUESTIONS

Conceiving of AI as a substitute for human decision making is challenging from a technical point of view. Examining the relationship between AI and decision making, Jean-Charles Pomerol has delineated two major aspects of decision making: diagnosis and "look ahead."[17] Diagnosis involves pattern matching and is therefore perfectly amenable to AI. Look ahead requires both the ability to combine many actions and events and the ability to anticipate all possible reactions. Back in 1997, Pomerol had already predicted a brilliant future for machines that could perform both diagnostic and look-ahead functions. Today, IBM's Watson Trend predicts human preferences for use in recommendation systems.[18] DeepStack, a poker-playing algorithm for settings with imperfect information (as in poker, where not all players have identical information), handles each situation as it arises, using fast approximate estimates, which can be thought of as intuition, instead of in-depth computations.[19] Like human intuition, DeepStack's intuition needs to be trained, but then it can accurately predict the behavior of its antagonists and thereby act in ways that make its own behavior unpredictable.

Beyond general concerns about the wisdom and feasibility of using algorithms to predict personal choices, there are more practical questions regarding the generation and use of data.[20] Will data used for training algorithms be well protected against leaks or attacks? Will algorithms be designed for ethically appropriate purposes? We can imagine an algorithm-based clinical decision support tool that would — presumably on the basis of predicted patient preferences — recommend palliative care or the continuation of aggressive treatment primarily for economic reasons, and its recommendation could provide legitimation for action.

Additional questions emerge when we look beyond the use of a single algorithm. Will algorithm-generated information — for instance, about consumer choices, social life, health risks, and resuscitation status — be stored in some central database? Who will have access to it, and for what purposes? Clear governance and appropriate consumer-protection mechanisms need to be developed to address such challenges.[21,22]

MORALITY, TRANSPARENCY, HUMANITY

Even once rules have been defined for appropriate data handling, an issue remains that is particularly important with regard to moral decisions: algorithms are not always transparent. Concerns about trusting a "black box" have been expressed in relation to the entire field of machine learning, despite recent efforts to create AI that is explainable.[23] When it comes to moral decisions, ethicists have a long history of trying to spell out and probe arguments in a way that is comprehensible to others. On the other hand, we might argue, people are not necessarily following rational reasoning standards when defining their personal preferences — for instance, for or against organ donation. In fact, they sometimes find it hard to define what their preferences are when clinicians urgently need to know. Although some philosophers and computer scientists have argued in favor of harnessing machines as ethics advisors or even ethical agents,[24,25] outsourcing ethical considerations to machines still seems peculiar: we hold moral judgment dear as a function reserved for humans. Some ethicists, following Immanuel Kant, believe that human dignity rests in our species' capacity for moral judgment. Will we cede this special ability to

machines? What happens to human autonomy when we delegate important decisions to machines that we believe know better than we do?

On the other hand, since the advent of the global positioning system (GPS), we do not seem to have lost our capacity for spatial orientation, although we may bother less to mentally prepare a route or, through use of such devices, develop a sense of dependency on navigation assistance. Indeed, one might argue that a GPS allows us to train our orientation by providing constant feedback. We still tell the system where we want to go, and we choose the route. We can also depart from the system's suggestions at any time. We can turn it off. We can update it with new software. Yet future generations may find it quite unthinkable to do entirely without a GPS. Perhaps the role of AI-assisted ethical decision making will be similar.

The prospect that algorithms may compound the effects of evidence-based medicine, guidelines, and budget targets in limiting the scope available for individual clinical judgment is disconcerting to clinicians who believe that their professionalism is under threat.[26] The American Medical Association addresses this point by conceiving of AI not as artificial intelligence but as "augmented intelligence" that enhances rather than replaces physicians' expertise.[27] But even if algorithmic predictions were at first understood as voluntary decision support, they might eventually turn into perceived standards so that refraining from their use or departing from their recommendations would require justification.[28]

How would the patient–provider relationship be affected when physicians followed algorithms in making life-or-death decisions? And who would be accountable or liable for decisions that did not turn out well, whether physicians had used algorithm-based support tools or renounced their use?[29] Because of the insecurities reflected in such questions, an AI-based system might have low uptake by clinicians or low actual impact on decisions. On the other hand, if algorithms demonstrably helped clinicians, patients, and families to reach better decisions more efficiently, health care providers would probably be interested. Even if algorithms erred in a certain percentage of cases, their performance would need to be compared not with ideal conditions but with clinical reality, which can be messy and affected by time pressure, lack of crucial information, an overwhelmed family, unavailability of translators, and other factors. Algorithm-based preference prediction would be fast and easy to document — characteristics that might arouse concern that it would have too great an effect on decisions, with predictions taken at face value and immediately guiding and providing justification for clinical action. Defining the appropriate deployment of AI-based decision support will be vital for successful implementation.

DEVELOPMENT CONSIDERATIONS

It seems inevitable that machine learning will make its way into the realm of moral decisions. To protect patients and clinicians, we will need to ensure that the safety, validity, reproducibility, usability, and reliability of algorithm-based decision aids are established with the required scientific rigor.[30]

Getting good training data is not easy. Digital platforms that not only contain health records but also help patients generate and document well-considered preferences regarding future health care choices, providing interfaces for exchange with care providers as needed, could be a great resource for the training of algorithms. Ideally, the decisions could be assessed retrospectively by surrogates, care providers, or both.

Transparency and comprehensibility — achieved, for example, by revealing an algorithm's decision trees and how the results relate to human considerations — will also be important features of ethical algorithm-aided decision making.

Performance benchmarks will need to be developed by comparing algorithmic predictions of preferences with the current standard — the projections of surrogates or the courses that care providers would have taken. Given that trying to surmise another person's preferences is a tricky and imperfect business,[31] we would not be surprised to find that algorithms outperform even close relatives.

Quality control and monitoring of algorithm-based support systems are crucial and will have to incorporate vigilance for built-in bias and discrimination. The structures, standards, and processes that the pharmaceutical regulators use in evaluating drugs for approval might serve as a

model. In fact, the Food and Drug Administration has over the past few years started approving algorithms used in health care for functions such as early detection of atrial fibrillation, diagnosis using ultrasound imaging, and prediction of seizures[32] and is currently further developing and refining its procedures.[33]

Clear and enforceable rules regarding data protection and privacy as well as robust mechanisms for consumer protection are necessary preconditions for trustworthy AI-based preference predictors.

CONCLUSIONS

In order to harness the potential of algorithms for improving decision making, we will need to remain aware of their limits. It is well known, for instance, that even well-performing algorithms can be unreliable in individual cases. Context and explanations are still hard for algorithms to grasp. A single algorithm may have shortcomings that necessitate a combination of approaches. From an ethical perspective, it is of prime importance not to rush into commercial exploitation of potentially attractive decision-support apps, but instead to carefully probe what is really feasible. At this point, algorithms may be able to provide likelihoods of someone's preferences on matters such as resuscitation, organ donation, or palliative care, but such outputs do not and should not amount to outsourcing and automating moral judgment. The ethical decision and the responsibility that comes with it still rest with the human agent.

Algorithms may prompt us to revisit some questions that ethicists have long puzzled over, such as how we can know what a good ethical decision is. They also raise new ones: Will algorithms end up making better, more reliable, and more consistent moral choices than humans do? What can we learn from algorithms to improve our ethical reasoning and decision-making skills? How can we create the most effective synergies between human and artificial intelligence? The algorithm-aided prediction of individual health care preferences may make for an interesting test case as we explore these questions.

Disclosure forms provided by the authors are available at NEJM.org.

From the Institute of Biomedical Ethics and History of Medicine, University of Zurich, and the Collegium Helveticum — both in Zurich, Switzerland (N.B.-A.); and the Department of Neuroradiology, University of Heidelberg, Heidelberg, Germany (A.B.).

1. Aletras N, Tsarapatsanis D, Preotiuc-Pietro D, Lampos V. Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective. Peer J Comput Sci 2016;2:e93 (https://peerj.com/articles/cs-93).
2. Poczter S, Jankovic LM. The Google Car: driving toward a better future? J Bus Case Stud 2013;10:7-14.
3. Rajpurkar P, Irvin J, Ball RL, et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med 2018;15(11):e1002686.
4. Avati A, Jung K, Harman S, Downing L, Ng A, Shah NH. Improving palliative care with deep learning. BMC Med Inform Decis Mak 2018;18:122.
5. Beck M. Can a death-predicting algorithm improve care? Wall Street Journal. December 2, 2016 (https://www.wsj.com/articles/can-a-death-predicting-algorithm-improve-care-1480702261).
6. Bloch-Budzier S. NHS using Google technology to treat patients. BBC News. November 22, 2016 (https://www.bbc.com/news/health-38055509).
7. Insights Team. Rethinking medical ethics. Forbes Insights. February 11, 2019 (https://www.forbes.com/sites/insights-intelai/2019/02/11/rethinking-medical-ethics).
8. Lindvall C, Lilley EJ, Zupanc SN, et al. Natural language processing to assess end-of-life quality indicators in cancer patients receiving palliative surgery. J Palliat Med 2019;22:183-7.
9. Abdoler E, Wendler D. Using data to improve surrogate consent for clinical research with incapacitated adults. J Empir Res Hum Res Ethics 2012;7:37-50.
10. Lamanna C, Byrne L. Should artificial intelligence augment medical decision making? The case for an autonomy algorithm. AMA J Ethics 2018;20:E902-E910.
11. Surowiecki J. The wisdom of crowds. New York: Random House, 2005.
12. Courtland R. Bias detectives: the researchers striving to make algorithms fair. Nature 2018;558:357-60.
13. Choi H, Jang E, Alemi AA. WAIC, but why? Generative ensembles for robust anomaly detection. Ithaca, NY: Cornell University, 2019 (arXiv:1810.01392 [stat.ML]).
14. Nalisnick E, Matsukawa A, Teh YW, Gorur D, Lakshminarayanan B. Do deep denerative models know what they don't know? Ithaca, NY: Cornell University, 2019 (arXiv:1810.09136 [stat.ML]).
15. Silver D, Hubert T, Schrittwieser J, et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science 2018;362:1140-4.
16. Vincent J. DeepMind's Go-playing AI doesn't need human help to beat us anymore. The Verge. October 18, 2017 (www.theverge.com/2017/10/18/16495548/deepmind-ai-go-alphago-zero-self-taught).
17. Pomerol J-C. Artificial intelligence and human decision making. Eur J Oper Res 1997;99:3-25.
18. Godoy-Lorite A, Guimerà R, Moore C, Sales-Pardo M. Accurate and scalable social recommendation using mixed-membership stochastic block models. Proc Natl Acad Sci U S A 2016;113:14207-12.
19. Moravčík M, Schmid M, Burch N, et al. DeepStack: expert-level artificial intelligence in no-limit poker. Ithaca, NY: Cornell University, 2017 (arXiv:1701.01724v1).
20. Char DS, Shah NH, Magnus D. Implementing machine learning in health care — addressing ethical challenges. N Engl J Med 2018;378:981-3.

21. Montreal Declaration for a Responsible Development of Artificial Intelligence. 2018 (https://www.montrealdeclaration-responsibleai.com/the-declaration).
22. Food and Drug Administration. Digital health innovative action plan. 2019. https://www.fda.gov/downloads/medicaldevices/digitalhealth/ucm568735.pdf.
23. Hutson M. Has artificial intelligence become alchemy? Science 2018;360:478.
24. Anderson M, Anderson SL, eds. Machine ethics. Cambridge, United Kingdom: Cambridge University Press, 2011.
25. Giubilini A, Savulescu J. The artificial moral advisor: the "ideal observer" meets artificial intelligence. Philos Technol 2018;31:169-88.
26. Could you sue diagnostic algorithms or medical robots in the future? The Medical Futurist. June 30, 2018 (https://medicalfuturist.com/could-you-sue-diagnostic-algorithms-or-medical-robots-in-the-future).
27. American Medical Association. Augmented intelligence in health care: policy report. 2018 (https://www.ama-assn.org/system/files/2019-01/augmented-intelligence-policy-report.pdf).
28. De Witte B. Is it unethical to refrain from using Algorithms when diagnosing patients? Linked in. 2018 (www.linkedin.com/pulse/unethical-use-algorithms-diagnose-patients-bart-de-witte).
29. Goldhahn J, Rampton V, Spinas GA. Could artificial intelligence make doctors obsolete? BMJ 2018;363:k4563.
30. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. JAMA 2018;320:2199-200.
31. Zikmund-Fisher BJ, Sarr B, Fagerlin A, Ubel PA. A matter of perspective: choosing for others differs from choosing for yourself in making treatment decisions. J Gen Intern Med 2006;21:618-22.
32. Topol E. Deep medicine: how artificial intelligence can make healthcare human again. New York: Basic Books, 2019.
33. Food and Drug Administration. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) — discussion paper and request for feedback. April 2, 2019 (https://www.regulations.gov/docket?D=FDA-2019-N-1185).

DOI: 10.1056/NEJMms1904869