

CASE AND COMMENTARY

How Should Clinicians Communicate With Patients About the Roles of Artificially Intelligent Team Members?

Daniel Schiff, MS and Jason Borenstein, PhD

Abstract

This commentary responds to a hypothetical case involving an assistive artificial intelligence (AI) surgical device and focuses on potential harms emerging from interactions between humans and AI systems. Informed consent and responsibility—specifically, how responsibility should be distributed among professionals, technology companies, and other stakeholders—for uses of AI in health care are discussed.

Case

Mr K is a 54-year-old man referred to Dr L's outpatient spine neurosurgery clinic because he has a 6-week history of left-sided lower back pain, left leg weakness, and shooting pain. Prior to Mr K's appointment, Dr L reviewed the MRI of Mr K's lumbar spine, noting herniation of disc between the fifth lumbar vertebra and the first sacral vertebra (L5-S1), which is compressing Mr K's sacral (S1) nerve root.

"What a classic case," Dr L murmurs to herself. She grabs her reflex hammer and walks down the hall to exam room 3. After performing a brief evaluation and reviewing Mr K's MRI with him, Dr L recommends surgery to relieve compression of the S1 nerve.

"Isn't that a dangerous procedure? Could I end up paralyzed?" Mr K asks.

"There are certain risks, but with the help of the *Mazor Robotics Renaissance*® Guidance System technology, the procedure is relatively safe." Dr L explains the surgical planning using the Mazor system: "It employs artificially intelligent software to analyze your images and plan placement of my surgical tools. I've been using this technology for about a year now, and I've done over 30 surgeries—just like the one I'm recommending for you—with this technology."

Mr K looks uncomfortable. "I don't want a robot doing my surgery. I want you to do it all."

Dr L wonders how to respond.

Commentary

In this commentary, we examine a hypothetical case involving an assistive surgical device that is in use today, the Mazor Robotics Renaissance Guidance System.¹ It can assist surgeons like Dr L in performing procedures such as spinal fixation.^{2,3} With a complex technology like the Renaissance System, a series of policies and procedures are important for ensuring its ethical use. These measures include well-designed clinical trials; creation and implementation of procedures before, during, and after surgery, especially concerning complications, errors, and robustness measures; training on the technology's characteristics, uses, and limitations; and how to inform patients about such information. Depending on the type of technology, approval by the US Food and Drug Administration or other regulatory entities might be required.

While these considerations might be relevant to any complex device, several more specific challenges emerge with respect to artificial intelligence (AI) technologies. AI can refer to a range of techniques including expert systems, neural networks, machine learning, and deep learning.⁴ Medical ethics has begun to highlight concerns about uses of AI and robotics in health care, including algorithmic bias, the opacity and lack of intelligibility of AI systems, patient-clinician relationships, potential dehumanization of health care, and erosion of physician skill.^{5,6} In response, members of the medical community and others have called for changes to ethical guidelines and policy and for additional [training requirements for AI devices](#).⁶

Given the potential of AI to augment human medical care, the proper role of health care professionals vis-à-vis their digital counterparts is particularly relevant. First, the “black-box” problem—the mystery of how the system derives its outputs—is an issue for any complex and opaque medical technology. It raises questions about how to communicate possible biases, risks, and error rates during the informed consent process.^{6,7} Second, as Mr K's concerns demonstrate, informed consent can be complex given uncertainties, fears, or even overconfidence about uses of AI. Finally, assigning responsibility and liability when errors occur is also complicated by the technical complexity and opacity of AI and the challenge of distributing responsibility across many parties. We address each of these ethical concerns below.

Informed Consent and the Black-Box Problem

One ethical challenge emerging from interactions between Mr K and Dr L in the case pertains to the difficulty of obtaining consent to use a novel AI device. As Appelbaum notes, “Valid informed consent is premised on the disclosure of appropriate information to a competent patient who is permitted to make a voluntary choice.”⁸ As is commonly known, relevant information includes the purpose of the treatment, its potential benefits and risks, and possible alternative treatment options. Yet the novelty and technical sophistication of an AI device places additional demands on the [informed consent process](#). When an AI device is used, the presentation of information can be complicated

by possible patient and physician fears, overconfidence, or confusion. Moreover, for an informed consent process to proceed appropriately, it requires physicians to be sufficiently knowledgeable to explain to patients how an AI device works, which is rendered difficult by the black-box problem.

The black-box problem emerges for at least a subset of AI systems, including neural networks, which are trained on massive data sets to produce multiple layers of input-output connections.⁹ The result can be a system largely unintelligible to humans beyond its most basic inputs and outputs.¹⁰ In other words, those interacting with an AI system might not understand to any appreciable degree how it works (ie, its functioning seems like a black box). This challenge pertains not only to neural networks but also to any informationally or technically complex system that may be opaque to those who interact with it, such as Mazor's advanced and proprietary image recognition algorithms.³

The opacity of an AI system can make it difficult for health care professionals to ascertain how the system arrived at a decision and how an error might occur. For instance, can physicians or others understand why the AI system made the prediction or decision that led to an error, or is the answer buried under unintelligible layers of complexity? Will physicians be able to assess whether the AI system was trained on a data set that is representative of a particular patient population? And will physicians have information about comparative predictive accuracy and error rates of the AI system across patient subgroups? In short, if physicians do not fully understand (yet) how to explain an AI system's predictions or errors, how could this knowledge deficit impact the quality of an informed consent process and medical care more generally?

Ongoing conversations within many professional communities will be needed to grapple with these issues, but recommendations are already emerging. For example, Char et al. state,

Physicians who use machine-learning systems can become more educated about their construction, the data sets they are built on, and their limitations. Remaining ignorant about the construction of machine-learning systems or allowing them to be constructed as black boxes could lead to ethically problematic outcomes.⁶

Moreover, professional societies are recommending that AI systems be "transparent."¹¹

Assuming Dr L is well informed about the Renaissance Guidance System, she should seek to explain to Mr K the core technologies used, such as the basic nature of the image recognition algorithm. She should clearly distinguish between the roles human caregivers will play during each part of the procedure and the roles the AI/robotic system or device will play. For example, she should explain that she is responsible for the preoperative plan, whereas the Renaissance Guidance System will manually guide

placement of tools or implants.³ Also, Dr L should clearly state the potential harms that might result from either human or robotic missteps.

Patient Perceptions of AI

Interconnected with lack of knowledge about AI systems—including how errors could occur—are varied perceptions patients and health care professionals have about AI technology. Computing experts offer wide-ranging visions of where AI is going, from utopian views in which humanity's problems are largely solved to dystopian scenarios of human extinction.¹² These visions can influence whether patients, such as Mr K in the case, and physicians embrace AI (perhaps too quickly) or fear it (even though it might improve health outcomes). For example, a 2016 survey of 12 000 people across 12 European, Middle-Eastern, and African countries found that only 47% of respondents would be willing to have a "robot perform a minor, non-invasive surgery instead of a doctor," with that number dropping to 37% for major, invasive surgeries.^{12,13} These findings indicate that a sizeable proportion of the public has uneasiness about medical AI.

How should a physician respond to patients like Mr K who express concerns about the use of AI? In addition to delineating the role of the AI system, the physician can address the patient's fears or overconfidence by describing the risks and potential novel benefits attributable to the AI system. For example, beyond merely sharing that she has used this procedure in the past, Dr L should describe studies comparing the Renaissance Guidance System to human surgeons.² In this way, the patient's inaccurate perceptions of AI can be countered with a professional assessment of the benefits and risks involved in a specific procedure. While these 2 recommendations are important for proper informed consent, understanding and responding to patients' fears is also essential to good patient engagement and medical care. These 2 recommendations are not intended to be an exhaustive list; rather, they are a starting point for addressing sources of serious clinical and ethical concern about AI.

Medical Errors, AI, and the Problem of Many Hands

Suppose that Dr L uses the AI device to treat Mr K and a medical error occurs. How might one begin to assign responsibility for the error? Determining who is morally responsible and perhaps [legally liable](#) for a medical error involving use of a sophisticated technology is often complicated by the "problem of many hands."¹⁴ This problem refers to the challenge of attributing moral responsibility when the cause of a harm is distributed among multiple persons—and perhaps organizations—in a way that obfuscates blame attribution. As Harris et al. state, individuals might use a many hands argument in an attempt "to evade personal responsibility for wrongdoing."¹⁵ Given that many parties are involved in the design, sale, procurement, and use of AI systems in health care, identifying the primary locus of responsibility for a medical error can be difficult.¹⁶ Moreover, the opacity of some AI systems compounds this challenge in new ways. Yet

transparency and clarity about roles and responsibilities can help ensure that the responsibility net is cast neither too narrowly nor too broadly.

A first step towards assigning responsibility for medical errors (thus hopefully minimizing them in the future) is to disentangle which people and professional responsibilities might have been involved in committing or preventing the errors. In the context of health care and AI, we suggest the following as a subset of the actors who could in principle be held ethically responsible for a medical error.

- *Coders and designers.* Coders and designers should be responsible for documenting what they created and, insofar as possible, implementing strategies for making explainable the technology and its underlying processes, such as how the AI is learning from training data.
- *Medical device companies.* Companies should clearly articulate prerequisites for successful application of an AI technology, such as the quality of diagnostics, imaging, and preparation for surgical procedures. Moreover, given black box concerns with AI systems, physicians might require additional information and training. Companies should therefore detail types of errors and side effects, their likelihood and severity, and differences in predictive accuracy and error rates across demographic subgroups, health conditions, and patient histories. Given uncertainties and risks surrounding complex, novel AI technologies in health care, companies should be responsible for providing meaningful information to hospitals and physicians, even if doing so surpasses what the law strictly requires.
- *Physicians and other health care professionals.* Physicians should be responsible for acquiring basic understanding of the AI devices they use and the types and likelihood of errors across subgroups, insofar as this information is available. Physicians should also be responsible for communicating relevant information to patients and health care teams and for adhering to use standards provided by device companies. Thus, if a medical error occurs because instructions for using an AI device were not followed, the primary responsibility could lie with the physician (or team); however, if a medical error occurs because adequate instructions or training were not provided by the company, the primary responsibility could lie elsewhere.
- *Hospitals and health care systems.* Hospitals are key to ensuring proper development, implementation, and monitoring of protocols and best practices for use of AI systems in health care. This organizational responsibility includes providing training, protocols, and best practices related to AI use and properly informing patients about the technology. Hospitals should also be involved in

developing robustness measures (including simultaneous diagnosis and crosschecking by physicians and AI). Best practice standards are also needed for error assessment and mitigation in cases of complications and for quality improvement.

Other actors, including regulators, insurance companies, pharmaceutical companies, and medical schools, also have important responsibilities. Each actor can take steps to ensure safe, ethical use of AI systems and encourage others to do so, too. These actions can help promote coordination among the various stakeholders about the use of AI in health care and contribute to a clearer sense of how to assign responsibility for successes as well as errors.

Challenges of AI in Health Care

While the challenges of integrating AI into the health care arena involve variations of familiar ethical issues, AI nevertheless presents new possibilities and concerns that deserve renewed attention. We suggest that companies provide detailed information about AI systems, which can help ensure that physicians—and subsequently their patients—are well informed. By explaining to patients the specific roles of health care professionals and of AI and robotic systems as well as the potential risks and benefits of these new systems, physicians can help improve the informed consent process and begin to address major sources of uncertainty about AI. Hopefully, the health care community will collectively meet these goals by encouraging open and robust dialogue about evaluating new AI technologies and integrating them into training and patient care.

References

1. Garrity M. Who is Mazor Robotics' biggest competitor in the spine market? *Becker's Spine Review*. November 27, 2017. <https://www.beckersspine.com/orthopedic-a-spine-device-a-implant-news/item/39026-who-is-mazor-robotics-biggest-competitor-in-the-spine-market.html>. Accessed June 6, 2018.
2. Joseph JR, Smith BW, Liu X, Park P. Current applications of robotics in spine surgery: a systematic review of the literature. *Neurosurg Focus*. 2017;42(5):e2.
3. Mazor Robotics. Renaissance: how it works. <https://mazorrobotics.com/en-us/product-portfolio/mazor-x/mazorx-how-it-works>. Published May 29, 2018. Accessed May 31, 2018.
4. Brookfield Institute for Innovation + Entrepreneurship; Policy Innovation Hub; Ontario. Intro to AI for policymakers: understanding the shift. https://brookfieldinstitute.ca/wp-content/uploads/AI_Intro-Policymakers_ONLINE.pdf. Published March 2018. Accessed July 13, 2018.
5. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. *N Engl J Med*. 2016;375(13):1216-1219.

6. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. *N Engl J Med*. 2018;378(11):981-983.
7. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA*. 2017;318(6):517-518.
8. Appelbaum PS. Clinical practice. Assessment of patients' competence to consent to treatment. *N Engl J Med*. 2007;357(18):1834-1840.
9. Castelvechi D. Can we open the black box of AI? *Nature*. 2016;538(7623):20-23.
10. Knight W. The dark secret at the heart of AI. *MIT Technology Review*. April 11, 2017. <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/>. Accessed June 20, 2018.
11. American Medical Association. Augmented intelligence in health care H-480.940. <https://policysearch.ama-assn.org/policyfinder/detail/augmented%20intelligence?uri=%2FAMADoc%2FHOD.xml-H-480.940.xml>. Modified 2018. Accessed December 7, 2018.
12. Müller VC, Bostrom N. Future progress in artificial intelligence: a survey of expert opinion. In: Müller VC, ed. *Fundamental Issues of Artificial Intelligence*. Cham, Switzerland: Springer; 2016:555-572.
13. PricewaterhouseCoopers. What doctor? Why AI and robotics will define new health. <https://www.pwc.com/gx/en/industries/healthcare/publications/ai-robotics-new-health/ai-robotics-new-health.pdf>. Published April 2017. Updated June 2017. Accessed October 15, 2018.
14. PricewaterhouseCoopers. What doctor? Why AI and robotics will define new health: data explorer. <https://www.pwc.com/gx/en/industries/healthcare/publications/ai-robotics-new-health/data-explorer.html#!/D/18/stackedbars?cut=Territory&Tecf=0>. Accessed December 6, 2018.
15. Thompson DF. Moral responsibility of public officials: the problem of many hands. *Am Polit Sci Rev*. 1980;74(4):905-916.
16. Harris CE, Pritchard MS, Rabins MJ. *Engineering Ethics: Concepts and Cases*. 4th ed. Belmont, CA: Wadsworth Cengage Learning; 2009.
17. Dixon-Woods M, Pronovost PJ. Patient safety and the problem of many hands. *BMJ Qual Saf*. 2016;25(7):485-488.

Daniel Schiff, MS is a PhD student at the Georgia Institute of Technology in Atlanta, where he studies artificial intelligence and its intersection with social policy. He earned an AB in philosophy from Princeton University and an MS in social policy from the University of Pennsylvania.

Jason Borenstein, PhD is the director of Graduate Research Ethics Programs and the associate director of the Center for Ethics and Technology at the Georgia Institute of Technology in Atlanta. His appointment is divided between the School of Public Policy and the Office of Graduate Studies.

Editor's Note

The case to which this commentary is a response was developed by the editorial staff.

Citation

AMA J Ethics. 2019;21(2):E138-145.

DOI

10.1001/amajethics.2019.138.

Conflict of Interest Disclosure

The author(s) had no conflicts of interest to disclose.

The people and events in this case are fictional. Resemblance to real events or to names of people, living or dead, is entirely coincidental. The viewpoints expressed in this article are those of the author(s) and do not necessarily reflect the views and policies of the AMA.