Ethics in Biology, Engineering & Medicine - An International Journal, 11(1):75–80 (2020)

Ethics in Artificial Intelligence: Hidden Dangers

Alan M. Reznik[a,b] & Fred R.T. Nelson[c,d,*]

[a]Connecticut Orthopaedics, Hamden, CT, USA; [b]Yale School of Medicine, New Haven, CT, USA; [c]Department of Orthopaedics, Henry Ford Hospital, Detroit, MI, USA; [d]Wayne State University School of Medicine, Detroit, MI, USA

*Address all correspondence to: Fred R.T. Nelson, MD, Henry Ford Fairlane, Orthopaedics, 19401 Hubbard Drive, Dearborn, MI 48126, USA, E-mail: fnelson1@hfhs.org

ABSTRACT: Artificial intelligence (AI) has been on the horizon for several decades and offers applications for the medical field, such as radiographic interpretation. Deep AI involves a neural network that is the equivalent of an electronic model of brain neurons. The network is a keystone for deep AI but involves no specific ethical reasoning. Given the spectrum of potential commercial developments, stakeholders must consider the ethics of decision making. The first of four basic elements of medical ethics is autonomy, which provides the patient with independence of thought and decision making and freedom from coercion or coaxing, leading to fully informed patient consent. The second is justice, in which new or experimental treatments share burdens and benefits that are distributed throughout all groups. The third is beneficence or the intent to do good for the patient. The fourth is nonmaleficence, where the goal is to do no harm to patient or society as a whole. AI is no longer the future; it is here and growing in orthopaedics-related applications. Knowledge of and consideration for medical ethics are critical when evaluating AI applications. In this article, we discuss the impact of the 2019 coronavirus disease epidemic and the international approach to ethical considerations.

KEY WORDS: artificial intelligence, ethics, justice, nonmaleficence, benefice, autonomy

I. ETHICS IN MEDICINE

Most medical schools in the United States include medical ethics in their curriculum, and the majority of graduates take the Hippocratic oath before graduation. The oath states, *I will abstain from all intentional wrong-doing and harm*. In the 1700s, the meaning of the oath morphed into the Latin "Primum non nocere" (First, do not harm). However, with the crush of clinical training and responsibilities, many orthopaedic surgeons are unable to recall all medical ethics elements as they apply to patient care. The American Academy of Orthopaedic Surgeons' Code of Medical Ethics and Professionalism for Orthopaedic Surgeons states that *The physician–patient relationship has a contractual basis and is based on confidentiality, trust, and honesty*.[1] It adds that *Orthopaedists shall not decline to accept patients solely on the basis of race, color, gender, sexual orientation, religion, or national origin or on any basis that would constitute illegal discrimination*.[1] Training and certification for human research requires ethics education. As in clinical practice, institutional review boards require deidentified data, data security, and patient consent.

The current code of ethics exists as a result of the many ethical errors and tragedies in medical research that occurred during the first half of the 20th century. Uninformed consent in the Tuskegee syphilis trials is a sad example. The unethical use of prisoners for research was most dramatic in German concentration camps during World War II. Medical experiments on women (known as "rabbits") at the Ravensbrück concentration camp was one of many human abuses committed during the war. These included surgical induction of lower-leg infections to determine whether infections could be treated using sulfanilamide.[2] As a result of these acts, the Nuremberg trials yielded the Nuremberg code, a set of ethical principles governing human research. The code was then tied to the Declaration of Geneva in 1948. Following many international conferences, the declaration was completed in 1964. The original declaration has been updated a number of times and it continues to be a touchstone for ethical guidance. Its principles are codified within the 1979 Belmont Report under the auspices of the Department of Health, Education, and Welfare. Three principles outlined in the Belmont Report are (1) respect for persons, (2) beneficence, and (3) justice. Within the context of justice, nonmaleficence (to do no harm) is best exemplified by the errors of the Tuskegee syphilis study. In light of this history, it is reasonable to question the ethics of any new medical treatment as well as the use of artificial intelligence (AI) in medical devices, data analysis, prediction, and guiding treatment allocation.[3] Ultimately, it seems clear that any computer aid to human analysis and medical thinking must meet our well-established ethical standards.

II. AI

To better understand "thinking" machines, we must go back to some origins of the human need for such devices. The use of machines to aid human thinking dates back to da Vinci's "cart" (~ 1517), Babbage's calculating machine that was able to solve polynomial equations (~ 1830), and Lovelace's first computer algorithm, developed shortly after the invention of the computer.[4,5] No one doubts that the modern story of AI can be traced to 1950, when Alan Turing wrote *Computing Machinery and Intelligence* and laid the groundwork for today's AI.[6] But the cost of a megabyte was more than $2 million in 1960, and because of limited computing power, AI overpromised and underdelivered. Early AI efforts failed, and an AI winter followed. During the past 10 yr, with massive amounts of inexpensive storage available, as Turing predicted, the "imitation game" returned and AI was reborn. In 2020, with the combination of low-cost memory, the cloud, and high-performance computing, we are discovering new treatments, "reading" natural language to analyze medical papers, checking retinas for diabetic changes, analyzing X-rays at the pixel level, and predicting outcomes even before advanced treatments begin. In fact, the coronavirus disease (COVID) outbreak was predicted in December 2019 by natural language processing of medical reporting in multiple languages.[7]

Exploring AI applications in this regard, we can categorize AI or machine learning into two types: shallow and deep AI.[8,9]

A. Shallow AI

Shallow AI is embedded in many common electronic devices in medicine. In every office and all parts of the hospital can be found devices with sensors or inputs that use algorithms that are based on physics, chemistry, mathematics, and experimental knowledge to give reliable, reproducible results. These can include specific algorithms to solve a problem or strategies that learn (so-called heuristic programing) by storing data and referencing it by mathematical design (such as electrocardiogram computer readings). Once perfected, if we do not change the device's fixed function, the US Food and Drug Administration (FDA) can approve it for safe, general use. By fixed function, we mean that the algorithmic relationship of inputs to outputs cannot be altered by normally operating the software as a device. This would qualify the device to be predictable and safe, as opposed to a less predictable learning machine. Such well-tested reliability is very important to devices and their use.

B. Deep AI

Deep AI, neural networks, or deep learning systems are based on networks of perceptrons: digital models of a human neuron that can be arranged in layers, similar to a brain. Like the real brain, different types of neurons have distinct operative properties. Deep AI brains learn from structured or unstructured data to predict solutions to questions based on new data that are presented to the network. As with shallow AI technology, once perfected, if perceptron internal parameters do not change, the FDA can grant approval for safe, general use. Paradoxically, without allowing the machine to alter internal parameters (after training it and with FDA approval), it no longer can learn from new data. Naturally, with a software as medical device designation, any new training alters its function and invalidates its original FDA approval. Only retesting can revalidate the device. This is a challenge for AI that may not be at its best if bound by the software as a medical device requirement. It follows that newer standards for deep AI may be needed in the near future.

III. ETHICS IN AI

The use of AI in health care is expanding into diverse areas such as image recognition, risk prediction, patient-specific payment models, and clinical decision making.[9,10] The danger for AI in medicine is that four basic assumptions are needed to set up an AI task by training the computer's neural network (in deep AI). First, the training set must represent the real world in a reasonable distribution of actual patients. Second, the inherent characteristics of population on which the AI is used must be similar to the algorithm's original training population for the results to be valid. Third, population data that is used today must be representative of the population and its inherent traits that will exist in the foreseeable future. Finally, the AI application itself must not alter the patient population

as it acts on that population.[8] In other words, when AI provides treatment recommendations, if the future population differs from the current population, the prior training set may not represent the future population. This would violate the third AI data assumption.

With these four assumptions regarding training data, AI makes decisions that will potentially alter medical care, whether by obtaining a diagnosis or decision making about coverage for care—all in the absence of AI informed consent. We assume that computers analyze accurate raw data that do not include opinions[11] (only facts). Hence, if we have adopted all four AI data assumptions, the four principles of medical ethics of autotomy, justice, benefice, and nonmaleficence that are explained below would have no role in AI.[12–14] As it turns out, this may not be the case.

A. Autonomy

Autotomy is involved in many preapproval processes for which surgical or other care is delayed or denied, based on an algorithm. The FDA does not oversee the insurance industry's use of AI. AI may interfere with the doctor–patient relationship, altering patient care and delaying interventions. It is clear that all four ethical considerations may come into play.

B. Justice

Justice can be affected by demographic data. An underserved population may remain underserved because zip code data confirms that the population performs poorly in response to a given treatment (e.g., a higher complication rate for total knee replacement). The group's general care must be improved before the data can change. In this case, AI would be correct but functionally and ethically wrong.[15]

C. Beneficence

A treatment may be undervalued based on skewed costs or lack of data for that specific new treatment in a given patient type. Use of prior AI data to identify current benefits may inhibit the use of new procedures and stifle innovation.

D. Nonmalifecence

Nonmaleficence may occur when the above principles collide with each other. Justice may increase cost above the societal beneficence. Autotomy may demand, for example, a marginal magnetic resonance imaging when the risk of being sued for a missed diagnosis is present, especially if personal cost to the physician is a potentially upset patient or time spent in court for a sound medical judgment that turned out to be wrong. Nonmaleficence may have two different meanings when we consider the whole society benefit verses the individual benefit.

E. Patient Privacy

AI could also violate the silent fifth ethical principle of patient privacy. AI interconnectivity could capture social media posts, search activity, location, credit card data, or Fitbit data (Fitbit, Inc.; San Francisco, CA) that could be used for health evaluation. Who owns one's personal data? Can personal data be fairly and ethically used to provide an equivalent benefit? Even the most recent COVID-19 outbreak predictions may take advantage of private, personal airline flight data without consent or significant oversight, for public good.[7,16] AI might amplify data bias, including real and artificial costs, and be blind to ethical and cultural implications.[12,17] To properly understand these concepts, a human being must be in the loop when we set up AI applications for health care.

IV. CONCLUSIONS

New data rights exist in the European Union that are not yet fully developed in the US.[18] In the US, the tort process may define who owns data, yet that process requires that harm occur first before legal action is taken. As physicians, we are trained to act to prevent or mitigate harm before it occurs, so this model is flawed from the start. If machines guide care, the idea of a medical license for AI to practice medicine may evolve into a necessity. The responsible entity for malpractice may shift to machines, programmers, data-set providers, or the machine manufacturers. We may cross a line from the doctor–patient relationship to a data–patient relationship. If that occurs, which set of ethics would apply? The risk here is that the benefit to society may outweigh or overrun the individual's rights.

Recently, the Vatican provided six principles for the use of AI: transparency, inclusion, responsibility, impartiality, reliability, and security/privacy.[19] Similarly, the US Department of Defense also officially adopted a series of ethical principles for the use of AI[20]; namely, AI in the military must be responsible, equitable, traceable, reliable, and governable. We can see how these fit well with medical ethics principles. One concern for AI application is whether the use of the three sets of ethics that are outlined here will be universally accepted or applied by all societies, private corporations, insurers, governments, and military forces.

Coming full circle, if we follow Hypocrites, we follow the oath to use our learned wisdom and judgment to do no harm, administer no poison, and divulge no person's secrets as we gain access to enter a home to apply our knowledge. Therefore, as we march toward the future, we are under the ethical obligation to protect our patients from any harm that could occur by creating machines that replace some of the essence of wisdom and judgment that we have been taught to use wisely.

REFERENCES

1. AAOS.org. Code of medical ethics and professionalism for orthopaedic surgeons [webpage on the Internet]. Rosemont, IL: American Academy of Orthopaedic Surgeons; 1988 [cited 2020 Apr 2]. Available from: https://aaos.org/about/bylaws-policies/ethics-and-professionalism/code/.

2. Bagatur E. Nazi medicine. Part 1: Musculoskeletal experimentation on concentration camp prisoners during World War II. Clin Orthop Relat Res. 2018 Oct;476(10):1899–905.
3. Parasidis E, Pike E, McGraw D. A Belmont Report for health data. N Engl J Med. 2019 Apr 18;380(16):1493–5.
4. Collier B, MacLachlan J. Charles Babbage and the engines of perfection. New York, NY: Oxford University Press; 1998.
5. Bromley AG. Difference engine. In: Reilly ED, editor. Concise encyclopedia of computer science. Hoboken, NJ: Wiley; 2004. p. 253–4.
6. Turing AM. Computing machinery and intelligence. Mind. 1950;LIX(236):433–60.
7. Niiler E. An AI epidemiologist sent the first warnings of the Wuhan virus [monograph on the Internet]. Boone, IA: Wired, Conde Nast; 2020 [cited 2020 Apr 2]. Available from: https://www.wired.com/story/ai-epidemiologist-wuhan-public-health-warnings/.
8. Reznik AM. Evolution versus revolution in AI: Why the distinction? Am Acad Orthop Surg Now. 2019 Dec;2019(Dec):22–3.
9. Myers TG, Ramkumar PN, Ricciardi BF, Urish KL, Kipper J, Ketonis C. Artificial intelligence and orthopaedics: An introduction for clinicians. J Bone Joint Surg Am. 2020 May 6;102(9):830–40.
10. Wang VM, Cheung CA, Kozar AJ, Huang B. Machine learning applications in orthopaedic imaging. J Am Acad Orthop Surg. 2020 May 15;28(10):e415–7.
11. Bronson WH, Fewer M, Godlewski K, Slover JD, Caplan A, Iorio R, Bosco J. The ethics of patient risk modification prior to elective joint replacement surgery. J Bone Joint Surg Am. 2014 Jul 2;96(13):e113.
12. Reznik AM, Urish K. Applying the four basic principles of medical ethics to artificial intelligence. Am Acad Orthop Surg Now. 2019;2019(July):29–39.
13. Gillon R. Medical ethics: Four principles plus attention to scope. Br Med J. 1994 Jul 16;309(6948):184–8.
14. McCormick TR. Principles of bioethics [monograph on the Internet]. Seattle, WA: University of Washington; 2018 [cited 2020 Apr 2]. Available from: https://depts.washington.edu/bhdept/ethics-medicine/bioethics-topics/articles/principles-bioethics.
15. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019 Oct 25;366(6464):447–53.
16. Bogoch II, Brady OJ, Kraemer MUG, German M, Creatore MI, Kulkarni MA, Brownstein JS, Mekaru SR, Hay SI, Groot E, Watts A, Khan K. Anticipating the international spread of Zika virus from Brazil. Lancet. 2016 Jan 23;387(10016):335–6.
17. Maxmen A. Self-driving car dilemmas reveal that moral choices are not universal. Nature. 2018 Oct;562(7728):469–70.
18. Horgan D, Romao M, Morre SA, Kalra D. Artificial intelligence: Power for civilisation - and for better healthcare. Public Health Genomics. 2019;22(5-6):145–61.
19. Pontificial Academy for Life, Microsoft, IBM, FAO, The Italian Ministry of Innovation. Rome Call for AI ethics [monograph on the Internet]. Rome, Italy: Rome Call; 2020 [cited 2020 Mar 26]. Available from: http://www.academyforlife.va/content/dam/pav/documenti%20pdf/2020/CALL%2028%20febbraio/AI%20Rome%20Call%20x%20firma_DEF_DEF_.pdf.
20. Defense Innovation Board, Department of Defense. AI principles: Recommendations on the ethical use of artificial intelligence by the Department of Defense [monograph on the Internet]. Arlington, VA: US Department of Defense; 2019 [cited 2020 Apr 2]. Available from: https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF.