Applying the Four Basic Principles of Medical Ethics to Artificial Intelligence

ALAN M. REZNIK, MD, MBA, FAAOS

EDITOR'S NOTE: This article is part one of a two-part series on ethics in artificial intelligence (AI). This article explores the underlying four basic principles of medical ethics as they apply to orthopaedic practice and AI. Part two will discuss the differences between shallow and deep AI as it applies to medical ethics. Previous articles published in *AAOS Now* in 2018 and earlier this year reviewed the basics of AI, how AI may be used to read X-rays, and natural language processing as it applies to medical diagnoses. Visit www.aaosnow.org to read more.

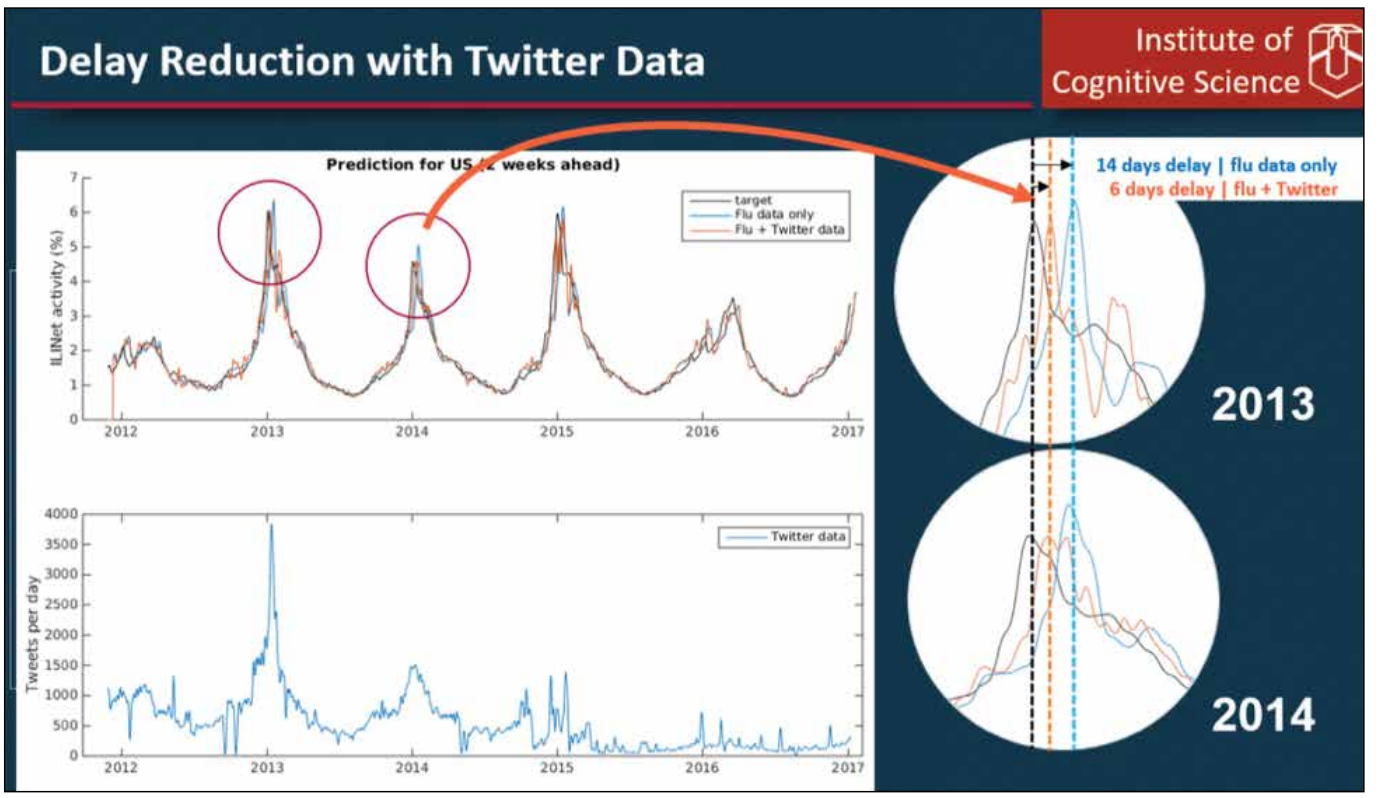

Fig. 1 The prediction of flu outbreaks improved by six days when artificial intelligence was used to analyze more social data.
COURTESY OF THE DEPARTMENT OF NEUROINFORMATICS AT THE INSTITUTE OF COGNITIVE SCIENCE AT THE UNIVERSITY OF OSNABRÜCK IN GERMANY

To understand the ethics of using AI in medical decision making, we must consider the four pillars of medical ethics: autonomy, justice, beneficence, and nonmaleficence. We equally need to understand something about data collected on actionable medical decisions, including "Should this patient have an X-ray, CT scan, MRI, mammogram, PET scan, or bone scan?" or "Is this X-ray, mammogram, or study cancer-free?" These sample questions and the data required to answer them make no ethical judgements on their own. If we do not attach an ethic to the data, we may expect that AI is ethically neutral. Yet, if we use AI to find medical themes in data, almost counterintuitively, the process seems to amplify the known and unknown biases contained in the data.

For example, in the first chess-playing AI programs, the AI engine determined that early queen sacrifice was associated with winning games without understanding why that was true for some masters' games, which caused many lost games. The engine made a correlation but had no understanding of the cause behind it or what preceded the event of interest—the queen sacrifice. Similarly, access to care, good nutrition, and general medical care may lead an AI engine to determine, based on poor outcomes, that care should be denied to those groups without understanding that if access was improved, outcomes might change.

In orthopaedics, even without the application of AI, we are trying to be more data-dependent or "evidenced-based." With the bias amplification observed in AI and our goal to be more evidence-based, there is a need to better understand how bias amplification occurs.

Neural networks

A neural network is the equivalent of an electronic model of brain neurons. There are many types of neurons using various mathematical expressions. Some are used to find local averages or maxima, or recall prior inputs as they process the next input (local memory). The "neurons" can be arranged in layers with interactions between them. The resulting network can look at input in a temporal sequence, randomly, or in a reverse direction. These properties are particularly helpful in language translation, as languages have different syntax.

In reaching a diagnosis, the output, AI can utilize a wide variety of potential inputs. For example, the network can learn the relationships between datasets from X-rays, other images, volumes of language (text or voice), and pathology slides; the network does not have any precondition or opinion on the information. It only tries to map a path to the best possible prediction of the desired output. Neural networks, by their nature, need and can handle large amounts of data to improve their function. In contrast, humans better predict outcomes with limited datasets and struggle with huge data sets. We see this contrast when we use Google's search function; we get millions of results from a search of billions of webpages in a fraction of a second. Google's neural network never gets tired of doing this, whereas humans get fatigued after reading and reviewing the first two pages of results.

Neural networks are the workhorse of deep AI. By design, they have no specific ethical reasoning. As a network uses existing examples to "learn" ways of predicting outcomes, answering questions, and making inferences, it will also find and reinforce all preexisting biases in the dataset being used to "train" it. The amplification of bias surprised the initial users of medical deep AI.

Understanding the ability of neural networks to process massive amounts of human data, it is not hard to imagine how AI must conform to the four basic ideals of medical ethics even in the simplest of medical applications. This understanding is critical to predict the future impact AI applications may have on medical care.

Autonomy in practice

After a careful history and physical exam, a healthcare provider has a set of possible test options, diagnoses, and then treatment options. He or she chooses how to present the data and communicate that information to the patient before offering treatment plans. The patient may ask questions and makes what we believe is an informed decision. This is not perfect autotomy because we are human and tend to favor what works well in our own hands. This may strain the concept of autonomy. However, there is the opposite example in AI: A machine could choose a treatment that may be technically "better." It may then turn out that that technique is not practical given the known resources and may equally strain our application of the third principle, beneficence.

With its ability to digest data faster than humans, AI may find many viable alternatives for some treatments; some may have never even been considered by the physician. It may find only one best option, with others that are vastly inferior. At each extreme, the contrast between the AI and human decision processes may eat away at both the physician and patient's autonomy. From personal experience, it seems that insurance carriers are beginning to use AI as negative screeners for preapproval of surgery. If training that particular AI engine yields fewer surgical approvals—saving carriers money, regardless

> SEE *MEDICAL ETHICS* ON PAGE 30

Basic principles of medical ethics

The four basic principles are as follows:

1. Autonomy is independence of thought and decision making, free of coercion or coaxing and requiring fully informed consent.
2. Justice is when new or experimental treatments share the burdens and benefits and are distributed throughout all groups.
3. Beneficence is the intent of doing good for the patient.
4. Nonmaleficence is the goal of doing no harm to the patient or society as a whole.

SOURCE: NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION

Medical Ethics

FROM PAGE 29

of the cost to individuals—it will give the carrier even more control of the decision process. Patient and physician autonomy will continue to be lost.

With AI in the equation, decisions will be far more data-dependent than we think. Data quality may vary greatly. AI will eventually tell us how long an operation should take, the complications expected by each surgeon, which hospital to use, and who should do what procedure and at what bundled price.

Justice in data collection and training neural networks

In epidemiology, we ask a network to learn about a disease incidence in an area and use the data to study an outbreak. We frequently use zip codes to identify locations. Unknown to the network, zip codes are full of unknown and known bias. In particular, the housing market is not uniform, and zip codes have some inherent bias by socioeconomic status, race, religion, and political preference. Healthcare utilization measures such as vaccination rates, nutritional status, number of doctors available, and proximity to medical facilities are tied tightly to socioeconomic issues and are indirectly biased by such factors. The bias is tracked all too well by addresses and zip codes, so using a patient's home location automatically may have an impact on the outcome of the analysis. The data may therefore yield racially or socioeconomically biased results. Our goal may have been an unbiased epidemiological analysis, but because AI uses all the data to find relationships, the results may still reflect inherent zip code bias. Yet, for data to have epidemiological value, geographic location can be very important. The same is true when zip code data are included in nonepidemiologic studies, such as patient satisfaction after joint replacement or outcomes of knee ligament reconstructions. Hence, the ethics of using AI with addresses or zip codes may need to be justified and checked for unethical bias each time they are used.

Social media may also be used to assist in medical care as it has been shown to help identify the start of a flu outbreak (Fig. 1). Similarly, demographic data may be gleamed from online social activity. Private information, such as education level, sexual orientation, and even race, may be embedded in the data associated with social media. We must be concerned since social media has also been shown to inherent bias. It is clear that application of social media in the medical setting may lead to a conflict with the principle of justice.

Beneficence and data bias

Those of us who trained in orthopaedic surgery over the past 50 years were indoctrinated on standard pediatric growth charts. The charts were also tied to a standard atlas of bone age. Surgical interventions for growth plate arrests were based on those predictions. Studies of a small European population found that the data were not as universal as once thought. This resulted in bone age rules that were less valuable in other populations.

The issue is that ethnicity as a factor has an uncomfortable feel. Yet, to avoid the natural bias of some datasets, it may be necessary to include such information. AI has the advantage that it can take even more factors into consideration. We must ask ourselves whether

> *"I believe medical ethics will be pushed to the forefront as artificial intelligence plays a bigger role in directing care."*
>
> **— Alan M. Reznik, MD, MBA, FAAOS**

including race or genetic makeup in data will help individuals in a more accurate way. This is in opposition to the objection we often see in just asking racial questions. Patients may be put off by such requests and have the right to refuse to answer. The idea of creating ubiquitous racially biased datasets may be unacceptable in our current society, and AI may indirectly yield that result.

This is a medical ethics dilemma where data collection must include location, race, and/or genetic origins in order to protect "beneficence" for individuals. However, justice in medical ethics may dictate the decreased use of addresses, zip codes, socioeconomic factors, and genetic profiles in data collection and application.

Nonmaleficence for patients and society

Nonmaleficence requires that we do no harm to the patient or society as a whole. In the future, AI may be used to drive disabled patients to appointments using autonomous cars. What if a car carrying an older patient has to make the choice between hitting a child who ran into the street or flipping the car over and possibly killing the passenger? This dilemma for autonomous vehicles is a frequently discussed example of the internal conflict of autonomous driving. What would a human driver do? How does the car decide? What would society expect? What does not harm the patient versus society as a whole?

Although that scenario is outside the realm of medical AI considerations, it serves as a good example of issues that can arise. The answers to such questions vary by cultural norms. Within the purview of nonmaleficence, we must ask how AI will balance individual needs versus society and differing cultures in daily medical care. Like the car driving and deciding what to do, if AI is programmed to consider cost of health outcomes, then cost may have a huge impact on decision making.

In orthopaedics, we have developed appropriate use criteria and clinical practice guidelines, with the goal of finding the best evidence and working toward best practices. The data in the AAOS Surgical Risk Reduction Toolkit indicate that poorly controlled diabetes, smoking, hypertension, poor nutrition, and obesity are risk factors for increased medical complications and costs.

Like zip codes, such data also may be linked to socioeconomics and/or race. Understanding that any individual should have the right to autonomy devoid of coercion but that individuals with the aforementioned conditions are at higher risk of complications, what choices do we offer our patients? However, complications require more resources and may indirectly consume medical options and treatment "bandwidth" for other members of society. Furthermore, because race and socioeconomic factors are potentially tied to those risk factors, we bump directly into the ethical principles of autotomy and justice.

The future of ethical AI

We are driven by outcomes analysis and alternative payment methods; as a result, we face ethical challenges. We know that disadvantaged groups already have less access to care, lower rates of some orthopaedic interventions, and higher complication rates. With the overarching goal of better allocation of medical resources and cost reduction, how do we still address all four ethical principles?

In the near future, I believe medical ethics will be pushed to the forefront as AI plays a bigger role in directing care, and we will be forced to routinely examine potential conflicts in applying the principles.

References for the studies cited can be found in the online version of this article, available at www.aaosnow.org.

Alan M. Reznik, MD, MBA, FAAOS, specializes in sports medicine and arthroscopic surgery and serves on the *AAOS Now* Editorial Board. Dr. Reznik holds six patents, including one on search enhancement, and is currently developing an AI search application for smartphones. Dr. Reznik is chief medical officer of Connecticut Orthopaedic Specialists, associate professor of orthopaedics at Yale University School of Medicine, and a consultant.