12-20-2018

Ethical Issues of Artificial Intelligence in Medicine

Tokio Matsuzaki

Recommended Citation

Matsuzaki, Tokio (2018) "Ethical Issues of Artificial Intelligence in Medicine," *California Western Law Review*: Vol. 55 : No. 1 , Article 7.
Available at: https://scholarlycommons.law.cwsl.edu/cwlr/vol55/iss1/7

ETHICAL ISSUES OF ARTIFICIAL INTELLIGENCE IN MEDICINE

DR. TOKIO MATSUZAKI*

TABLE OF CONTENTS

I. INTRODUCTION

Which will you choose for the future of medicine: Machine or human? For certain diseases, a fully automated surgery, performed using artificial intelligence ("AI") technology has a mortality rate of 1%

---

* Currently, Dr. Matsuzaki, an orthopedic surgeon by trade, works at the University of Tokyo as an Innovation Fellow for Japan Biodesign. Previously, he served as a visiting scholar at UC San Diego and as a postdoctoral fellow at Scripps Research. He is presently working on robotics for orthopedics, artificial muscle, machine learning, and anti-aging genes for the joint cartilage.

255

256 CALIFORNIA WESTERN LAW REVIEW [Vol. 55

due to mechanical error, while the same surgery performed by a human surgeon has a mortality rate of 10% due to human error.[1] In the wake of the recent technological advancements in medicine, the ethical implications of AI must be closely examined, especially since parts of the surgical process will soon be replaced by automation.

Recently, the development of "deep learning" or "deep structured learning" methods, in contrast to task-specific learning algorithms, has created an explosion in the rapid advancement of AI. In recent medical journals, many studies report high accuracy of imaging diagnosis using AI. Specifically, in some instances, AI devices have already exceeded the mean diagnostic rate of doctors, making AI technology an attractive and accurate diagnostic instrument.[2]

Robot automation has also been studied using machine learning algorithms.[3] Fully autonomous surgical robots will soon emerge given the inherent visual advantages provided by robots and the continued development of machine learning algorithms. Consequently, AI will become a crucial tool for assisting doctors and improving accuracy.

---

1. *See* Interpretable Machine Learning, *2v2 Debate: Caruana, Simard vs. Weinberger, LeCun*. Interpretable ML Symposium, NIPS 2017, YOUTUBE (Dec. 12, 2017), https://www.youtube.com/watch?v=2hW05ZfsUUo (noting the 10% mortality rate for surgeries performed by humans and highlighting that AI technology may be a safer way to perform surgery compared to the traditional human only method); s*ee also* Thierry Langanay et al., *Aortic Valve Replacement in the Elderly: The Real Life,* 93 ANNALS OF THORACIC SURGERY 70, 70–72 (2012) (observing that the mortality rate of patients ages 80–96 during aortic valve replacement was 6.9% when performed by a human surgeon).

2. *See generally* Andre Esteva et al., *Dermatologist-level Classification of Skin Cancer with Deep Neural Networks*, 542 NATURE 115, 117–18 (2017) (finding that AI technology exceeded the mean diagnostic rate of most human dermatologists when classifying and identifying skin cancer); *see also* Varun Gulshan et al., *Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs*, 316 JAMA 2402, 2403, 2407–08 (2016) (noting that automated and semi-automated evaluations of diabetic retinopathy "had high sensitivity and specificity for detecting diabetic retinopathy and macular edema" in pictures).

3. *See generally* Yohannes Kassahun et al., *Surgical Robotics Beyond Enhanced Dexterity Instrumentation: A Survey of Machine Learning Techniques and Their Role in Intelligent and Autonomous Surgical Actions*, 11 INT'L J. OF COMPUTER ASSISTED RADIOLOGY & SURGERY 553, 553 (2016) (detailing AI technology and machine learning used in surgery).

2018] ETHICAL ISSUES OF ARTIFICIAL INTELLIGENCE IN MEDICINE 257

AI is becoming more essential to the medical field due to the expansion in the amount of patient medical data collected, including patients' physiologic sensor data from wearable devices, genetic sequences, and race-dependent treatment options. As a physician, it is impossible to review all the data. Therefore, it is becoming clear that AI is necessary to analyze and prioritize what has become known as "big data," which is defined as "information assets

characterized by such a high volume, velocity, and variety to require specific technology and analytical methods for its transformation into value."[4] Additionally, a recent study indicated that in 21% of primary care patient cases the final diagnosis was completely different from the referral diagnosis, demonstrating the critical need for standardized diagnostic methods.[5]

Furthermore, the staggering number of deaths related to medical errors is another reason AI is a much-needed asset in the medical field. In the United States, medical error related deaths range from 210,000 to 400,000 annually.[6] These deaths are primarily caused by missed diagnoses and the lack of appropriate treatment.[7] The accuracy of AI technology can help reduce the number of deaths due to medical error. Surgical errors are also a big problem. According to a Mayo Clinic study published in 2017, 8.9% of surgeons reported making a medical error in the preceding three months.[8] This concerning statistic urges for a more stable and secure medical treatment system, which can be accomplished through supervised autonomous surgery or fully autonomous surgery.

---

4.   Andrea De Mauro et al., *What is Big Data? A Consensual Definition and a Review of Key Research Topics,* 4TH INTERNATIONAL CONFERENCE ON INTEGRATED INFORMATION 8 (Sep. 2014), https://www.researchgate.net/publication/265775800_What_is_Big_Data_A_Consensual_Definition_and_a_Review_of_Key_Research_Topics.

5.   *See* Monica Van Such et al., *Extent of Diagnostic Agreement Among Medical Referrals,* J. OF EVALUATION IN CLINICAL PRACTICE, Feb. 2017, at 2.

6.   *See* John T. James, *A new, Evidence-based Estimate of Patient Harms Associated with Hospital Care,* 9 J. OF PATIENT SAFETY 122, 122, 126 (2013) (finding that by "[u]sing a weighted average of the 4 studies, a lower limit of 210,000 deaths per year was associated with preventable harm in hospitals.").

7.   *See id.* at 123, 126.

8.   Tait Shanafelt et al., *Preventable Deaths in American Hospitals*, NEJM CATALYST (Jan. 23, 2017), https://catalyst.nejm.org/medical-errors-preventable-deaths/.

258   CALIFORNIA WESTERN LAW REVIEW   [Vol. 55

This article describes the details of recent approaches to the application of AI in medicine, the limitations and potential problems of AI, as well as the accompanying ethical issues involved in implementing this new technology. Part II of this article provides the history of AI and the current function of AI in medicine. Part III discusses the third boom, deep learning, as well as another algorithm type, and describes medical applications using the latest algorithms. Finally, Part IV examines the problems related to AI and the ethical implications that should be considered when applying AI to medical care.

II. THE HISTORY OF ARTIFICIAL INTELLIGENCE

AI has a robust history spanning over seventy years. The mid 1950s is considered the first boom of AI and is often referred to as the era of exploring and reasoning. During the first boom, the concept of the perceptron was born, which gives one output from multiple inputs–the basis of deep learning.[9] In 1955, mathematics professor John McCarthy coined the term "AI" in his grant application seeking funds for the 1956 Dartmouth AI Conference.[10] In the same period, Alan Turing proposed the "Turing test" to evaluate a machine's capability by analyzing whether the intelligent behavior of a machine is equivalent to that of a human.[11] Then, in the early 1970s, AI systems were developed as diagnostic and therapeutic tools for the medical field.

The second boom, called the knowledge-based era, occurred around 1980, with the development of a system emulating the

---

9. *See generally* F. Rosenblatt, *The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain*, 65 PSYCHOL. REV. 386, 387, 399–401 (1958) (finding that a "theory has been developed for a hypothetical nervous system, or machine, called a perceptron" which is "designed to illustrate some of the fundamental properties of intelligent systems.").

10. *See generally* J. McCarthy et al., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence,* (Aug. 31, 1955), http://www.formal. stanford.edu/jmc/history/dartmouth/dartmouth.html (detailing the proposed two-month long study of "artificial intelligence," a newly coined term at the time, in order to obtain a deeper understanding of machine learning).

11. *See* A.M. Turing, *Computing Machinery and Intelligence*, 49 MIND 433, 433, 441 (1950) (explaining the evaluation methods required to determine whether machines can "think" and suggesting "at the end of the century . . . one will be able to speak of machines thinking without expecting to be contradicted.").

2018] ETHICAL ISSUES OF ARTIFICIAL INTELLIGENCE IN MEDICINE 259

knowledge of a human expert, called an expert system.[12] During this time, Stanford University developed a system that used AI technology to diagnose infectious blood diseases, called "MYCIN,"[13] while the University of Pittsburgh developed a computer-assisted diagnostic tool known as "INTERNIST-I."[14] However, complicated processes, such as the presence of comorbidity and clinical course, make it difficult for these systems to provide an accurate diagnosis and make treatment recommendations.[15]

Also, during the second boom, automatic diagnostic equipment, used to evaluate electrocardiograms, was developed.[16] Although diagnostic equipment serves as an effective screening test, the diagnostic rate of myocardial infarction, in which a specific signal rises, is poor.[17] On the other hand, more successful developments emerged in the second boom, such as algorithms like the multi-layered perceptron,[18] which is a basic architecture of deep learning. Also, backpropagation was developed, which allows for tuning the parameters.[19] Additionally, the basic algorithms of current AI, such as

---

12.   *See* PETER JACKSON, INTRODUCTION TO EXPERT SYSTEMS 1–2 (Addison-Wesley ed., 3d ed. 1998).

13.   Edward H. Shortliffe & Bruce G. Buchanan, *A Model of Inexact Reasoning in Medicine*, 23 MATHEMATICAL BIOSCIENCES 351, 353–54 (1975).

14.   *See* Randolph A. Miller et al., *Internist-1, An Experimental Computer-based Diagnostic Consultant for General Internal Medicine*, 307 NEW ENG. J. OF MEDICINE 468, 468 (1982).

15.   *See* Gaurav Sharma & Alexis Carter, *Artificial Intelligence and the Pathologist: Future Frenemies?*, 141 ARCHIVES OF PATHOLOGY & LABORATORY MEDICINE 622, 622–23 (2017) (noting that some limitations of using AI technology for diagnostic purposes).

16.   Mats-Erik Nygårds & Johan Hulting, *An Automated System for ECG Monitoring*, 12 COMPUTERS & BIOMEDICAL RES. 181, 181 (1978).

17.   Myocardial infarction occurs when the heart muscle stops or decreases blood flow to the heart, which can result in death. See Antoine Ayer & Christian Juhl Terkelsen, *Difficult ECGs in STEMI: Lessons Learned from Serial Sampling of Pre- and In-hospital ECGs*, 47 J. ELECTROCARDIOLOGY 448, 448, 456 (2014), for a discussion of certain types of electrocardiogram signals and patterns that are typically incorrectly identified by AI algorithms.

18.   *See generally* D. E. Rumelhart et al., *A General Framework for Parallel Distributed Processing*, *in* PARALLEL DISTRIBUTED PROCESSING: EXPLORATIONS IN THE MICROSTRUCTURE OF COGNITION, VOL. 1: FOUNDATIONS 45, 60 (MIT Press ed., 1986).

19.   *Id.* at 51.

260   CALIFORNIA WESTERN LAW REVIEW   [Vol. 55

recurrent neural networks, were popularized and developed during this period.[20] Since the 2000s, the neural network architecture, along with improvements in computer performance, has increased to multiple layers, thereby endowing machine learning with higher accuracy.

III. THE THIRD BOOM: DEEP LEARNING AND LEARNING STYLES, AND THE MEDICAL APPLICATION USING THE LATEST ALGORITHMS

In 2006, the beginning of the third boom was sparked when AI acquired a "vision" of deep learning through the work of Geoffrey Hinton, of Toronto University. Hinton proposed a new algorithm, called deep belief networks, and demonstrated its effectiveness on a deep learning algorithm.[21] Hinton's work won the ImageNet competition in 2012, which recognized the visual image ("ImageNet") for its image recognition accuracy.[22] Through the development of ImageNeT, a significant reduction in the error rate of image classification was achieved, which dramatically improved the function of the algorithm using deep learning.[23] As a result of this tremendous improvement in machine learning algorithms, many researchers began to utilize deep learning as a major algorithm.

Machine learning is an integral part of AI. Machine learning is defined as: (1) an algorithmic computing approach to making a determination or prediction, (2) acting without being explicitly programmed, and (3) automatically improving itself.[24] Machine learning has many kinds of algorithms. Deep learning is one

---

20.   *See* J.J. Hopfield, *Neural Networks and Physical Systems with Emergent Collective Computational Abilities*, 79 PROC. OF THE NAT'L ACAD. OF SCI. OF THE U.S. 2554, 2555–56 (1982) (explaining how neural networks used in AI technology work mathematically).

21.   Geoffrey Hinton et al., *A Fast Learning Algorithm for Deep Belief Nets*, 18 NEURAL COMPUTATION 1527, 1527–28, 1530–31 (2006).

22.   *See* Alex Krizhevsky, Ilya Sutskever & Geoffrey E. Hinton, *ImageNet Classification with Deep Convolutional Neural Networks*, 60 COMMUNICATIONS OF THE ACM 84, 84 (2017) (discussing the "deep convolutional neural network" Hinton and his team used to classify over 1.2 million images in the ImageNet competition).

23.   *See id.* at 90–91.

24.   *Cf.* Arthur L. Samuel, *Some Studies in Machine Learning Using the Game of Checkers*, 3 IBM J. OF RES. & DEV. 210, 210 (highlighting aspects of machine learning by studying a machine's ability to learn and to improve playing the game of checkers).

2018] ETHICAL ISSUES OF ARTIFICIAL INTELLIGENCE IN MEDICINE 261

algorithmic approach to implementing machine learning, by using multi-layer artificial neural networks that mimic human neural architecture.[25]   The development of high-performance computing systems, such as graphics processing units and field-programmable gate arrays, have also accelerated AI development. Additionally, open-source AI frameworks, which are software libraries for designing and deploying numerical computations, have contributed to strengthen AI algorithms. Examples of these frameworks include TensorFlow or Caffe, and GitHub, which is a platform of open-source projects. More recent achievements in AI are the use of high performance speech recognition and natural language processing by deep learning algorithms, or combinations of deep learning and other algorithms.

A. Learning Styles

There have been major learning styles developed for training machine learning algorithms and neural networks. These advancements in learning styles are capable of extending the potential of AI in the following ways: (1) Supervised learning, where the data is labeled and annotated; (2) Unsupervised learning, where data is not labeled or annotated; (3) Transfer learning, which incorporates the use of pre-trained neural networks for another task; (4) Reinforcement learning, which gets an agent to act on a task so as to maximize its rewards; and (5) Meta-learning, which is trained by exposure to a large number of tasks and then testing the machine's ability to learn the new task, also known as "learning to learn."[26]

For example, under the theory of supervised learning, the labeled medical data would be pathologic diagnosis, physician's diagnosis, or genomic data. A tremendous amount of time and effort is required to create this labeled data. Aside from the time commitment, this learning style requires the neural networks to be trained and involves a large number of pre-labeled images. In contrast, in unsupervised learning, unlabeled data is used to train neural networks and to generate useful

---

25. *See* Yann LeCun et al., *Deep Learning*, 521 NATURE 436, 436, 438 (2015).

26. *See generally* Chelsea Finn et al., *Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks*, 34TH INTERNATIONAL CONFERENCE ON MACHINE LEARNING 1–4 (Jul. 18, 2017), https://arxiv.org/abs/1703.03400 (describing an algorithm for meta-learning and providing an overview of common learning styles used for AI).

262   CALIFORNIA WESTERN LAW REVIEW   [Vol. 55

clusters that have specific features and can be labeled data.[27] However, "unsupervised learning sometimes requires much larger training datasets" compared to supervised learning.[28] Transfer learning utilizes the pre-trained neural networks, which are trained by unsupervised learning or supervised learning.[29] Transfer learning usually requires smaller datasets than those used to initially train the model.[30] While supervised learning, unsupervised learning, and transfer learning are used for image classification, reinforcement learning is used mostly for taking action, such as in computer games, simulation, and surgical procedures. Reinforcement learning is a system that finds the most effective way to maximize given rewards, i.e., learning by trial and error. In 2017, this type of algorithm, developed by DeepMind Technologies and AlphaGo, was given media attention when it was reported to have outperformed the previously top-ranked human champions.[31]

Furthermore, reinforcement learning is expected to enhance personalized medicine by enabling physicians to optimize treatment strategies and treatment recommendations.[32] Also, Meta-learning has recently been developed in a manner which produces a learning algorithm which can essentially "learn" the algorithm itself.[33] The

---

27. Marc Kohli et al., *Implementing Machine Learning in Radiology Practice and Research*, 208 AM. J. ROENTGENOLY 754, 756 (2017).

28. *Id.*

29. *See id.* at 758–59 (providing an overview of transfer learning).

30. *See id.*

31. *See* Darrell Etherington, *Google's AlphaGo AI Defeats Team of Five Leading Go Players*, TECHCRUNCH (May 26, 2017), https://techcrunch.com/2017/05/26/googles-alphago-ai-defeats-team-of-five-leading-go-players/ (reporting that AI computer technology "defeated a team made up of five [human] Go champions in a demonstration match" playing a game called "Go").

32. *See* Yuxi Li, *Deep reinforcement learning: An overview*, 34 (Sept. 15, 2017), https://arxiv.org/pdf/1701.07274.pdf (discussing reinforcement learning in the context of healthcare and medicine); *see also* Niranjani Prasad et al., *A Reinforcement Learning Approach to Weaning of Mechanical Ventilation in Intensive Care Units*, 1 (Apr. 20, 2017), https://arxiv.org/pdf/1704.06300.pdf (discussing the possibility of creating a AI technology-based "decision support tool that leverages available patient information in the data-rich ICU setting to alert clinicians when a patient is ready for initiation of weaning, and to recommend a personalized treatment protocol.").

33. *See* Finn et al., *supra* note 26, at 1 (noting "the goal of meta-learning is to train a model on a variety of learning tasks, such that it can solve new learning tasks using only a small number of training samples.").

2018] ETHICAL ISSUES OF ARTIFICIAL INTELLIGENCE IN MEDICINE 263

Meta-learning algorithm is trained by exposure to a certain task and then undergoes testing in this new task.[34] Meta-learning streamlines the learning process because it is faster and more efficient to learn a new task, than to learn by reinforcement.[35] This algorithm can be used for industrial robots and, most importantly, has the potential to perform medical robotic tasks.

B. AI in Medical Image Diagnosis

In the past two years, a few highly influential papers, concerning AI and medical diagnostics, have been published. The first paper was published by Google and describes automated diagnosis of diabetic retinopathy using retinal images.[36] The authors of the Google article used a deep learning algorithm, coupled with transfer learning, to diagnose diabetic retinopathy.[37] A few months after Google's paper was published, Stanford University published a paper about skin cancer diagnosis using a similar method.[38] Aside from these two significant publications, around this same time, numerous countries competed in the Cancer Metastases in Lymph Nodes Challenge 2016 ("CAMELYON16") competition, which was organized to study algorithms used to detect metastases in lymph nodes, and resulted in the publication a study.[39] The CAMELYON16 study serves as a good example of an international initiative to improve diagnostic accuracy using AI technology. Collectively, all three publications proved that the deep learning algorithm outperformed human doctors; in fact, AI was more accurate than doctors.

In February 2018, the US Food and Drug Administration ("FDA") announced marketing clearance for Viz, which is AI's Contact

---

34.   *See id.* at 1, 5–6.

35.   *See id.* at 1, 4–6.

36.   *See* Gulshan et al., *supra* note 2, at 2402, for the Google sponsored study.

37.   *Id*. at 1.

38.   *See* Esteva et al., *supra* note 2, at 115, for the study conducted by Stanford University concerning skin cancer classifications using neural networks.

39.   See Babak Ehteshami Bejnordi et al., *Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women with Breast Cancer*, 318 JAMA 2199, 2199, 2208 (2017), for the CAMELYON16 study finding deep learning algorithms can be helpful for pathological diagnosis notwithstanding certain practical limitations.

264   CALIFORNIA WESTERN LAW REVIEW   [Vol. 55

application. Viz is the first AI-based clinical decision support system and can analyze computed tomography images and alert healthcare providers to a possible stroke in patients.[40] Two months after the announcement of Viz, the FDA permitted the marketing of the first medical device using AI to detect diabetic retinopathy.[41]

C. AI in Robotic Surgery

Machine learning and deep learning are now studied for robotic automation, marking the first step on the path toward fully autonomous surgery. In response to this progression, some commercial products are currently available, such as: (1) Direct control: da Vinci surgical system (Intuitive Surgical, USA); (2) Shared control: Mako Robotic-Arm Assisted Surgery (Stryker, USA); (3) Supervised autonomy: CyberKnife (Accuray, USA); and (4) Fully autonomous, which is not yet available.[42]

The widely used da Vinci robotic surgical system, approved in 2000 by the US FDA, is a direct control surgical system with a master console and controller on the surgeon's side and moving robotic arms on the patient's side.[43] Aside from a direct control surgical system, there is also what is known as a shared control system. An example of a shared control system is the Mako Robotic-Arm Assisted Surgery that is designed for knee or hip replacement.[44] The shared control system allows the doctor to guide the robotic arm to achieve the preplanned removal of the bone and cartilage. In these types of robotic surgeries,

---

40. *FDA Permits Marketing of Clinical Decision Support Software for Alerting Providers of a Potential Stroke in Patients*, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm596575.htm (last updated Feb. 13, 2018).

41. *See FDA Permits Marketing of Artificial Intelligence-Based Device to Detect Certain Diabetes-Related Eye Problems,* U.S. FOOD & DRUG ADMIN., https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm604357.htm (last updated Apr. 12, 2018) [hereinafter *FDA Diabetes Detection Device*].

42. Micheal Yip & Nikhil Das, *Robot Autonomy for Surgery*, 2–4 (July 10, 2017), https://arxiv.org/pdf/1707.03080.pdf.

43. *See generally* DA VINCI SURGERY: MINIMALLY INVASIVE SURGERY, http://www.davincisurgery.com/ (last visited Dec. 3, 2018).

44. *See generally Mako Robotic-Arm Assisted Technology,* STRYKER, https://patients.stryker.com/mako-robotic-arm-assisted-technology (last visited Dec. 3, 2018).

the surgeon is able to get haptic feedback to prevent bone resection outside the preplanned area. The shared control of the doctor and the robotic arm allows for a more accurate surgery compared to the conventional method.[45]

The next level of robotic surgery is supervised autonomy. An example of a supervised autonomy device is the CyberKnife. The CyberKnife is a noninvasive radiotherapy device for tumor removal which requires the doctor to define the area of radiation, and then automated focused radiation is applied.[46] Currently, there are no fully autonomous devices available in the United States. However, in September 2017, a Chinese hospital reported having a fully autonomous dental implant surgery device.[47] Although, this fully autonomous dental implant device is not commercially available yet.

Further, researchers at The Sheikh Zayed Institute for Pediatric Surgical Innovation in Washington D.C., published a paper regarding the performance of automated surgery. The paper demonstrated that a supervised autonomous surgical robot and an autonomous robot each outperformed open surgery, laparoscopic surgery, and robot-assisted surgery, which is similar to the da Vinci device used to suture a pig intestine.[48] Although, it will be a while before full automation is available for other types of surgery. In the meantime, some portions of the surgical process are likely to be replaced by automation.

---

45. *See* Eli Kamara et al., *Adoption of Robotic vs Fluoroscopic Guidance in Total Hip Arthroplasty: Is Acetabular Positioning Improved in the Learning Curve?*, 32 THE J. OF ARTHROPLASTY 125, 126–130 (2017) (concluding "a new haptic robotic technique for acetabular bone preparation and cup insertion" was a precise and accurate surgical method); *see also* Stuart W. Bell et al., *Improved Accuracy of Component Positioning with Robotic-Assisted Unicompartmental Knee Arthroplasty: Data from a Prospective, Randomized Controlled Study*, 98 THE J. OF BONE & JOINT SURGERY AM. 627, 627–28, 630–33 (2016) (finding that "[r]obotic-assisted surgical procedures . . . lead to improved accuracy of implant positioning compared with conventional unicompartmental knee arthroplasty surgical techniques.").

46. *See generally General Information*, CYBERKNIFE ACCURACY, http://www.cyberknife.com/treatment/radiation-therapy/general (last visited Dec. 3, 2018).

47. *See Chinese Robot Dentist is First to Fit Implants in Patient's Mouth without any Human Involvement*, SOUTH CHINA MORNING POST, http://www.scmp.com/news/china/article/2112197/chinese-robot-dentist-first-fit-implants-patients-mouth-without-any-human (last updated Sep. 22, 2018).

48. *See* Azad Shademan et al., *Supervised Autonomous Robotic Soft Tissue Surgery*, SCI. TRANSLATIONAL MED., May 4, 2016, at 1, 4–6.

266   CALIFORNIA WESTERN LAW REVIEW   [Vol. 55

IV. LIMITATIONS AND THE POTENTIAL PROBLEMS OF AI

Over the past few years, AI has become the dominant choice for image diagnosis in medicine. Although still in the experimental phase, autonomous AI systems will be used in the future for completing surgery, recommending treatments, predicting medical conditions, analyzing health data, analyzing genomic data, and generally, as robotic diagnostic tools. However, before these systems become fully integrated in the medicine, the limitations of the algorithm, the liability of the diagnosis produced by AI, security, and privacy in interacting with patient data, and morality of medical actions should be explored.

A. The Need for Massive Data Training and Issues with Reliability

Massive amounts of data are needed to properly train the neural networks and to ensure an optimal neural network algorithm. Without enough image data, robust and accurate systems cannot be developed. In turn, this can lead to overfitting,[49] which does not generalize well to new data. Also, it may be difficult to establish AI in certain countries because they do not have enough data needed to train the neural networks. One example of this data disparity is in Chile, where the total number of computed tomography and magnetic resonance imaging examinations is approximately 4 to 10 times smaller than in the United States. Additionally, deep learning relies heavily on labeled data to ensure the quality of the output of the algorithm. Consequently, these datasets require many experts and data analysts. Open-source data is available for some rare diseases.[50] However, for more common diseases, data is not made publicly available.

For common diseases, it should be considered whether AI technology will be cost-effective. Some factors to be explored are the

---

49. Overfitting occurs when the data is trained too well, which leads to the inability to fit the evaluation data and causes poor performance. *See* Krizhevsky, Sutskever & Hinton, *supra* note 22, at 88–89 (describing methods to reduce overfitting such as the "dropout" method which "consists of setting to zero the output of each hidden neuron with probability 0.5" resulting in "neurons which are 'dropped out'" and "do not contribute to the forward pass and do not participate in backpropagation.").

50. *See* Mathew I. Bellgard et al., *Second Generation Registry Framework*, SOURCE CODE FOR BIOLOGY & MED., June 20, 2014, at 4 (noting that there is a "rare disease registry using the second generation RDRF. . . available.").

2018] ETHICAL ISSUES OF ARTIFICIAL INTELLIGENCE IN MEDICINE 267

amount of time it takes for experts to label the data and the source of the data. Another consideration is the potential lack of information in the dataset, i.e., a missing value. Although structured data is easier to input and generally more complete, most biomedical data is not designed for AI training datasets. However, to combat these data set requirements, some scientists have shown that deep learning can be easily fooled by an extremely similar image that is far from the actual image. For example, deep learning systems have mistaken blue and orange wavy lines for starfish.[51] However, currently this is not a completely viable solution for the lack of data because the output of the result is not always correct. The overall concern is the reliability of the data, which indicates that at this stage of development, a human doctor should make the final decision.

B. Lack of Transparency and Lack of Evidence

The "hidden layer," is the middle layer of the neural networks in the deep learning algorithm and includes millions of parameters.[52] Although these parameters are optimized by automatic calculations, it is very difficult to determine how these parameters work and interact with each other. This inability to see how the parameters work is known as the "black box." Some studies have tried to shed light on this phenomenon, but the issue of transparency remains unsettled.[53] While deep learning has a "black box," "Explainable AI" is an available transparent algorithm containing decision trees and random forests. However, the performance of the transparent algorithm in image recognition, is far from acceptable.[54] Another area of inquiry is whether

---

51. Anh Nguyen et al., *Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images*, 1, 4–5 (Apr. 2, 2015), https://arxiv.org/pdf/1412.1897.pdf.

52. See Krizhevsky, Sutskever & Hinton, *supra* note 22, at 84, 91, for a study utilizing deep conventional neural networks, which encompasses this "hidden layer" of neural networks.

53. *See generally* Davide Castelvecchi, *The Black Box of AI*, 538 NATURE 20, 21, 23 (2016) (detailing attempts to solve the issue of transparency and how the issue has caused scientists to have to create AI methods that do not use deep learning). The explanation for transparency is very difficult to express in this paper. A more comprehensive example would be in the form of a mathmatical expression.

54. *See generally* David Gunning, *Explainable Artificial Intelligence (XAI)*, DEFENSE ADVANCED RESEARCH PROJECTS AGENCY, https://www.darpa.mil/program/

268   CALIFORNIA WESTERN LAW REVIEW   [Vol. 55

it is necessary to be able to explain how automated decisions are made, given that human explanations of decisions and predictions are usually wrong.[55] A recent algorithm study found AI is useful for diagnosing images, predicting a medical condition, and analyzing big health data.[56] However, this finding has not been backed up by other researchers to confirm its validity.

Countries should independently establish the evidence used in AI systems. Studies have found that computer-aided diagnosis for imaging have outperformed human-only diagnosis, proving that human doctors combined with AI will have a better outcome than humans in medical image diagnosis.[57] However, randomized clinical trials are required to confirm this hypothesis. Additionally, other applications of AI, including predictive medicine, treatment recommendations, and analysis of big health data and genomic data, should be explored by a randomized control study.

C. Liability, Security, and Privacy

Who is going to take responsibility for the AI system? This debate mirrors the growing concerns around autonomous vehicles. Criminal liability, the tort of negligence, and breach of warranty must be discussed before utilizing AI in medicine.[58] Recently, the FDA

---

explainable-artificial-intelligence (last visited Dec. 3, 2018) (explaining that the "effectiveness of these [AI] systems is limited by the machine's current inability to explain their decisions and actions to human users.").

55. Clint Boulton, *The Hidden Risk of Blind Trust in AI's 'Black Box'*, CIO (July 6, 2017, 3:00 AM), https://www.cio.com/article/3204114/artificial-intelligence/the-hidden-risk-of-blind-trust-in-ai-s-black-box.html.

56. *See* Andrew L. Beam & Isaac S. Kohane, *Big Data and Machine Learning in Health Care*, 319 JAMA 1317, 1317–18 (2018) (explaining the positive impact big data and machine learning has on the health care field).

57. *See e.g.,* Keith J. Dreyer & J. Raymond Geis, *When Machines Think: Radiology's Next Frontier*, 285 RADIOLOGY 713, 713, 715 (2017) (explaining mammograms, radiographs, and computed tomography as examples of areas suited for computer-aided diagnosis).

58. *See* J.K.C Kingston, *Artificial Intelligence and Legal Liability*, 2016 INT'L CONF. ON INNOVATIVE TECH. AND APPLICATIONS OF ARTIFICIAL INTELLIGENCE *in* RESEARCH AND DEVELOPMENT IN INTELLIGENT SYSTEMS XXXIII: INCORPORATING APPLICATIONS AND INNOVATIONS IN INTELLIGENT SYSTEMS XXIV 269, 269 (Bramer et al. eds., 2016) (using the example of self-driving cars to discuss "whether criminal

2018] ETHICAL ISSUES OF ARTIFICIAL INTELLIGENCE IN MEDICINE 269

approved the first AI medical device for detecting diabetic retinopathy with an accuracy of 87% to 90%.[59] Therefore, one solution to the concern of liability is if the automated product explains the limitations of the AI system, it might not be found liable. When applying this framework, it should be clarified whether the programmer made a correct algorithm, the program was used properly, or if the quality and accuracy of the data were appropriate. On the doctor's side, the diagnosis and treatment should be determined by various aspects, including history-taking, physical examination, laboratory tests, and imaging. Therefore, the majority of the responsibility will fall on doctors until the whole diagnostic AI system is established.

Another concern with the implementation of AI in medicine is security. Security issues are not limited to AI systems, other areas of AI will give rise to security concerns. Therefore, new solutions and protections will continue to be required to ensure security. In the future, when automation of diagnosis and surgery is established, security flaws will have the potential to cause serious human bodily harm.

One concern is that AI decision-making, especially deep learning, often has no transparency. This means that doctors and patients are not able to know how the AI system reached the decision. The European Union's General Data Protection Regulation ("GDPR") is a regulation addressing data protection and privacy and recently went into enforcement May 25, 2018.[60] Under the GDPR, people have the right to obtain "meaningful information about the logic involved" in automated decisions.[61] This regulation will be difficult to comply with using deep learning. Although developers must comply with the GDPR, this restrictive regulation may prevent further expansion of AI. A separate regulation or exception under the GDPR, which recognizes

---

liability could ever apply" to AI programs, and whether product liability or negligence should govern civil law with regards to AI programs).

59. *See FDA Diabetes Detection Device, supra* note 41 (finding an "[AI medical device] was able to correctly identify the presence of more than mild diabetic retinopathy 87.4% of the time and was able to correctly identify those patients who did not have more than mild diabetic retinopathy 89.5% of the time.").

60. *GDPR Portal: Site Overview,* EUGDPR, https://www.eugdpr.org/ (last visited Dec. 3, 2018).

61. Andrew D. Selbst & Julia Powles, *Meaningful Information and the Right to Explanation*, 7 INT'L DATA PRIVACY L. 233, 234 (2017).

270   CALIFORNIA WESTERN LAW REVIEW   [Vol. 55

the nuances of AI, data collection, and security would be preferable.[62] In addition, to being able to create a medical database for constructing AI systems, the database should be prioritized and supported by the government or large organizations.[63] This is because data collected for AI systems is not just a registry of data from medical associations, it includes many types of integrated and sensitive data.

D. Morality in the Future

When analyzing the moral implications of AI, it is first important to discuss the patient benefits of direct-control robotic surgery using state-of-art technology. The advantages of robotic surgery include "greater surgical precision, increased range of motion, improved dexterity, enhanced visualization, and improved access."[64] However, a recent large-scale study demonstrated that robotic surgery is associated with a prolonged operating time, doubling surgeries by over 4 hours, and higher hospital costs, about a 20% increase, compared with laparoscopic surgery.[65] Maintenance costs are also high.[66] In addition, in a robotic prostatectomy surgery, for the treatment of prostate cancer,

---

62. *C.f.* NAT'L SCI. AND TECH. COUNCIL COMM. ON TECH., EXEC. OFFICE OF THE PRESIDENT, *Preparing for the future of artificial intelligence: AI and Regulation* at 17 (2016) ("The general consensus of the RFI commenters was that broad regulation of AI research or practice would be inadvisable at this time. Instead, commenters . . . called for existing regulation to be adapted as necessary to account for the effects of AI.").

63. *See id. at* 40–41 (recommending federal agencies "prioritize open training data and open data standards in AI" and that the government "prioritize basic and longterm AI research.").

64. Da Vinci Surgery: Frequently Asked Questions, DA VINCI SURGERY, http://www.davincisurgery.com/da-vinci-gynecology/da-vinci-surgery/frequently-asked-questions.php (last visited Dec. 3, 2018).

65. Laparoscopic surgery is an abdominal surgery using a long instrument through small incisions with a tiny video camera called a laparoscope. *See* In Gab Jeong, et. al., *Association of Robotic-Assisted vs Laparoscopic Radical Nephrectomy with Perioperative Outcomes and Health Care Costs, 2003 to 2015*, 318 JAMA 1561, 1561–62 (2017) (discussing robotic-assisted surgery in patients receiving a radical nephrectomy).

66. *See* Chelsea Hill et al., *Robotic Joint Replacement Surgery: Does Technology Improve Outcomes?*, 34 THE HEALTH CARE MANAGER 128, 128 (2015) (discussing how the high costs of maintenance and upkeep for robotic equipment may hinder the use of robotics in hospitals).

2018] ETHICAL ISSUES OF ARTIFICIAL INTELLIGENCE IN MEDICINE 271

there was no improvement in the primary outcome–survival rate and quality of life–or major complications.[67] However, there were lower levels of blood loss and improvements in the functional recovery.[68] However even with robotic assistance, human error still exists.[69] Even in direct controlled robotic surgery, the incidence of robotic malfunctions is still reported to be 3.7% during surgery.[70] Although several challenges remain in direct-control robotic surgery, surgical errors made by autonomous robotic surgical devices will likely be one of the biggest legal issues in the future. The surgical errors made by autonomous robotic surgical devices may have been avoided by a human doctor, but these devices may outgrow the need for humans in the future. This raises these questions: should we continue developing these devices until the surgical error rate caused by the robot becomes zero and continue to allow patient injury caused by human error until we perfect the system? Or should we recommend autonomous robotic surgery once satisfactory results are obtained at the sacrifice of a few patients? Additional education is needed in the field of AI technology to answer these questions. Also, there need to be meaningful discussions of the advantages, disadvantages, and ethics of autonomous robotic surgery.

---

67.   *See* Dragan Ilic et al., *Laparoscopic and Robotic-Assisted versus Open Radical Prostatectomy for the Treatment of Localised Prostate Cancer*, COCHRANE DATABASE OF SYSTEMATIC REVIEWS, Sept. 2017; *see also* Hyunsuk Frank Roh et al., *Robot-Assisted Laparoscopic Surgery Versus Conventional Laparoscopic Surgery in Randomized Controlled Trials: A Systematic Review and Meta-Analysis*, PLoS ONE, Jan. 23, 2018, at 1 (finding that "treatment outcomes between robotassisted laparoscopic surgery (RLS) and conventional laparoscopic surgery (CLS)" did not result in a "statistically better. . . . . outcome[], with the exception of lower estimated blood loss.").

68.   Roh et al., *supra* note 67, at 1 (noting the improvement in blood loss using robot-assisted laparoscopic surgery); *see also* Ilic et al., *supra* note 67 (finding that robot-assisted laparoscopic radical prostatectomy was better at preserving erectile function than a human only surgery).

69.   *See* Homa Alemzadeh et al., *Adverse Events in Robotic Surgery: A Retrospective Study of 14 Years of FDA Data*, PLoS ONE, Apr. 20, 2016, at 1, 3, 10 (discussing various device malfunctions and errors that can occur when using robotic assistance in medical surgeries and procedures).

70.   *See* Emad Rajih et al., *Error Reporting from the Da Vinci Surgical System in Robotic Surgery: A Canadian Multispecialty Experience at a Single Academic Centre*, 11 CANADIAN UROLOGICAL ASS'N J. 197, 199 (discussing an error rate of 3.7% within the third generation of the da Vinci surgical system).

In the case of future treatments, an AI system may indicate that eligibility for an operation or treatment is met. However, in practice, a doctor, considering the patient's background or behavior, might consider that eligibility is not met empirically. Which one, doctor or AI system, should have legal and moral authority regarding treatment in the face of inconclusive evidence? Given the current state of technology, it should be the doctor. However, if the doctor explains both treatment options, including the one by the AI system, while obtaining informed consent from the patient, the patient might be confused by the different treatment strategies. The transparency of explanations provided by an AI system might affect the ethical relationship between doctor and patient. In these instances, adequate information and education about AI decisions should be provided to the patient before a decision is required.

V. THE ESSENTIALS OF MEDICAL ETHICS

The American Medical Association provides several principles that define the essentials of honorable behavior of physicians. The Code of Medical Ethics provides:

    I.   A physician shall be dedicated to providing competent medical care, with <u>compassion</u> and <u>respect for human</u> dignity and rights.

    II.   A physician shall uphold the *standards of professionalism*, be honest in all professional interactions, and strive to report physicians deficient in character or competence, or engaging in fraud or deception, to appropriate entities.

    III.   A physician shall <u>respect the law</u> and also recognize a <u>responsibility</u> to seek changes in those requirements which are contrary to the best interests of the patient.

    IV.   A physician shall <u>respect the rights</u> of patients, colleagues, and other health professionals, and shall <u>safeguard</u> patient confidences and <u>privacy</u> within the constraints of the <u>law</u>.

    V.   A physician shall continue to study, apply, and *advance scientific knowledge*, maintain a commitment to medical education, make relevant information available to patients, colleagues, and the public, obtain consultation, and use the talents of other health professionals when indicated.

    VI.   A physician shall, in the provision of <u>appropriate</u> patient care, except in emergencies, be free to choose whom to

2018] ETHICAL ISSUES OF ARTIFICIAL INTELLIGENCE IN MEDICINE 273

> serve, with whom to associate, and the environment in which to provide medical care.
>
> VII. A physician shall recognize a <u>responsibility</u> to participate in activities contributing to the improvement of the community and the betterment of public health.
>
> VIII. A physician shall, while caring for a patient, regard <u>responsibility</u> to the patient as paramount.
>
> IX. A physician shall <u>support</u> access to medical care for all people.[71]

In principle, the italicized words could be replaced or complemented by AI. To preserve their professionalism, physicians should use AI to respond to big data and make fewer misdiagnoses. Also, to keep up with advancing scientific knowledge, doctors should learn to leverage AI technology. The roles exemplified by the underlined words in the above code of ethics, however, i.e., those tasks specifically reflecting the humanity and legality of medicine, should not be performed by AI technology in its present state-of-art. In the current state of technology, human doctors should maintain these roles.

VI. CONCLUSION

AI has a long history with various applications in medicine. However, recent advancements in deep learning will greatly affect diagnostic methods and treatments based on AI systems. The algorithms of AI improve every year and will adapt to developments in areas like image diagnosis and robotic surgery in the future. However, there are many issues that must be addressed before completing integrating AI in medicine. Careful consideration must be given to possible situations and issues that may arise due to the rapid growth of technology. A call to action is needed for further discussion about the applications of AI technology and policy proposals by legal professionals, physicians, ethics experts, and computer engineers to elucidate the limitations and potential problems of AI in medicine.

---

71. American Medical Association, *AMA Code of Medical Ethics*, AMA (2001), https://www.ama-assn.org/sites/default/files/media-browser/principles-of-medical-ethics.pdf (emphasis added).