Ethical Challenges of Artificial Intelligence in Medicine and the Triple Semantic Dimensions of Algorithmic Opacity with Its Repercussions to Patient Consent and Medical Liability

Rafaella Nogaroli and José Luiz de Moura Faleiros Júnior

Abstract Artificial intelligence algorithms have the potential to diagnose some types of skin cancer or to identify specific heart-rhythm abnormalities as well as (or even better) than board-certified dermatologists and cardiologists. However, one of the biggest fears in the healthcare sector in the Era of AI in Medicine is the so-called *black box medicine*, given the obscurity in the way information is processed by algorithms. More broadly, it is observed that there are three different semantic dimensions of algorithmic opacity relevant to Medicine: (1) *epistemic opacity* for the insufficient physicians understanding of the rules an AI system is applying to make predictions and decisions; (2) *opacity for the lack of medical disclosure* about the AI systems to support clinical decisions and patient's unawareness that automated decision-making are being carried out with their personal data; (3) *explanatory opacity* for the unsatisfactory explanation to patients about the technology used to support professional decision-making. Therefore, the aim of this study is to analyze each type of opacity, considering hypothetical scenarios and its repercussions in terms of medical malpractice and patient's informed consent. From this, it will be defined ethical challenges of using AI in the healthcare sector and the importance of medical education.

R. Nogaroli (✉)
Federal University of Paraná, Curitiba, PR, Brazil
e-mail: rafaellanogaroli@ufpr.br

J. L. d. M. Faleiros Júnior
University of São Paulo, São Paulo, Brazil
e-mail: jfaleiros@usp.br

H. Sousa Antunes et al. (eds.), *Multidisciplinary Perspectives on Artificial Intelligence and the Law*, Law, Governance and Technology Series 58,
https://doi.org/10.1007/978-3-031-41264-6_12

1 Introduction: Advantages of Artificial Intelligence (AI) in Medicine

The Digital Age of Medicine created the concept of *smart health*, following the transformation phenomenon from traditional Medicine towards *P4-Medicine* (preventive, predictive, personalized and participatory) (Hood 2013; Holzinger et al. 2015).[1] In this new scenario, health care is no longer essentially limited to the treatment of pathologies (a task that has never been abandoned, of course) and is now focused on the adoption of measures aimed at preventing diseases (preventive medicine) (Balicer and Cohen-Stavi 2020) or making it possible to anticipate the diagnosis (predictive medicine). Regarding personal treatment, the patient is seen in a more individualized way (and less generic, therefore), based on his genetic and health data (personalized medicine). Finally, the doctor-patient relationship ceases to be something punctual and starts to develop in a continuous manner, with the patient's active participation (participatory medicine) (Flores et al. 2013).[2] With digital tools, patients can take a more active, participatory role in their health care and wellness decision-making. In this way, the diabetic patient can constantly monitor his blood glucose, enabling, in real time, algorithms to analyze the personal data provided, supporting the physician in faster, more efficient and personalized therapeutic decisions, regarding drug administration or dietary (Nogaroli and Kfouri-Neto 2021a).

The transformation of medical care in this more proactive/participative, preventive, precise model and focused on the individuality of each patient became possible from the combination of large volumes of health data and Artificial Intelligence algorithms. Human life, in the wake of the third millennium, will be conditioned to algorithms for solving problems and making more accurate decisions. Eric Topol, in his books about the present and future of Medicine (Topol 2013, 2016, 2019a, b), points to several scientific studies that attest the enormous AI's ability to diagnose some types of skin cancer, or identify specific heart rhythm abnormalities, as well or perhaps even better than dermatologists and cardiologists (Huang et al. 2020).

During the COVID-19 pandemic, AI also demonstrated its great potential in medical imaging around the globe. Due to the rapid increase in number of new and suspected COVID-19 cases, as an alternative to relieve pressure on radiologists and prevent further spread of the disease, AI-based algorithms were developed across

[1] "The convergence of patient-activated social networks, big data and their analytics, and systems medicine has led to a P4 medicine that is predictive, preventive, personalized, and participatory. Medicine will focus on each individual. It will become proactive in nature. It will increasingly focus on wellness rather than disease".

[2] By adding the "participatory" component, P4 Medicine "maximizes the effectiveness of systems medicine by expanding its application out from hospitals and clinics into homes, workplaces and eventually schools. With the addition of self-monitoring (activity, weight and calorie intake) and self-assessments in the participatory component, new quantities and forms of data will be aggregated and mined to generate new insight into health and disease. These insights will drive the development of new technologies, analytic tools and forms of care".

the globe that supported these professionals in quickly identifying the pathogen disease by analyzing computed tomography images of symptomatic patients of COVID-19 (Harmon et al. 2020).[3] Besides, in the last years, predictive algorithms have been used to target treatment more effectively toward high-risk patient groups for the prevention of major chronic disease complications. This approach is being applied in other relevant domains to allow for the identification of populations at risk and early identification of impending complications of multiple acute and chronic illnesses (Nogaroli and Nalin 2021; Nogaroli and Silva 2021).

Nowadays, IBM is one of the major companies that creates more technological solutions for the healthcare sector and developed the so-called *Watson for Oncology*, a solution powered by information from relevant guidelines, best practices, and medical journals and textbooks. Watson evaluates the information from a patient's medical record, along with medical evidence (scientific papers and clinical studies), thus showing possible treatment options for cancer patients, classified by confidence level. In the end, it will be up to the doctor to analyze the conclusions reached by the AI and decide which is the best treatment option for that specific patient (IBM Healthcare and Life Sciences 2021).[4]

The brief demonstration of these examples of Artificial Intelligence being incorporated into medical practice is to illustrate some of the various benefits that this technology can provide to the healthcare sector. These potential benefits, however, are accompanied by relevant ethical and legal concerns to be faced. AI brings many benefits to the healthcare sector, but its risks cannot be ignored, which are even many of them intrinsic to the technology itself. In September 2021, it was published an UN report that analyses how AI tools affect people's right to privacy and other human rights. The UN High Commissioner for Human Rights called for a moratorium on the AI systems, considering that the technology in several sectors has caused serious human rights risks and, therefore, it would be needed a pause in creating new AI tools until authorities can demonstrate that there are no significant

[3] The algorithms were programed with thousands of tomography images labeled training images in two general classes: (1) COVID-19 and (2) Not COVID-19. Images marked as "Not COVID-19" represented cases of patients with healthy lungs. Preliminary studies indicate chest CT has a high sensitivity for detection of COVID-19 lung pathology and several groups have demonstrated the potential for AI-based diagnosis, reporting as high as 95% detection accuracies. Also included were examples of patients with other lung diseases, such as lung cancer, tuberculosis, bronchiectasis, and pneumonia of non-viral etiology.

[4] "Watson for Oncology was developed in concert with Memorial Sloan Kettering Center (MSK). To date, it has invested nearly 15 million pages of medical content, including more than 200 medical textbooks and 300 medical journals. By combining MSK's world-renowned cancer expertise with the analytical speed of IBM Watson, the tool has the potential to transform how doctors provide individualized cancer treatment plans and to help improve patient outcomes. In 2015, nearly 44,000 oncology research papers were published in medical journals around the world, or more than 120 new papers each day, outpacing the ability of humans to keep up with the proliferation of medical knowledge. Watson's machine learning capability means it is continuously learning about oncology over time, and doctors have access to peer-reviewed studies, clinical guidelines and expert perspectives, enabling them to make more specific and nuanced treatment decisions more quickly, based on the latest data."

issues with accuracy or discriminatory impacts and that the AI systems comply with robust privacy and data protection standards (United Nations 2021).

In recent decades, with the exponential creation of new predictive algorithms in medical practice, it is also possible to observe a moment of a global crisis in the credibility of this technology in Medicine. There is a scenario of potential expressive AI risks in supporting the medical professional decision, considering several factors, including deficiency in the process of creating and validating algorithms, relevant degree of fallibility, unpredictability and algorithmic opacity (Topol 2019a, b).

Consequently, the present study proposes to investigate the potential risks of implementing AI in clinical practice, as well as the definition of ethical principles to be followed during the development of the technology and, after being introduced in the market, throughout its useful life cycle. From this, this paper will seek to draw some conclusions about the future of Artificial Intelligence algorithms in Medicine and the importance of medical education in digital health and new technologies.

2 Triple Semantic Dimensions of Algorithmic Opacity and Its Repercussions to Patient Consent and Medical Liability

One of the biggest fears in the health sector in the Era of artificial intelligence is the so-called 'black box medicine', given the obscurity in the way information is processed by the algorithms. More broadly, it is observed that there are three different semantic dimensions of algorithmic opacity relevant to Medicine: (1) *epistemic opacity*; (2) opacity for the *lack of medical disclosure*; and (3) *explanatory opacity*. Therefore, it is important to analyze each type of opacity, considering hypothetical scenarios and its repercussions in terms of medical malpractice and patient's informed consent.

(1) *Epistemic opacity*: there is a relevant complexity for physicians' understanding about how personal data are processed by algorithms, which can discover patterns within such a large number of variables that it becomes extremely difficult—or even impossible—for a human mind to understand it. In fact, this is a problem present in most Artificial Intelligence systems and it is called by Frank Pasquale by 'black box problem', in his book 'The Black Box Society' (Pasquale 2015, pp. 6–7). Thus, epistemic opacity occurs when there is not sufficient understanding of the rules that an AI system is applying to make classifications, predictions and decisions. As an example, this opacity can originate physician's lack of comprehension about the machine learning process to arrive at a certain diagnosis or prediction about his patient's clinical condition. The lack of transparency is also associated with the problem of reliability of algorithms predictions, and it raises understandable fears regarding the implementation of the technology in medical practice.

There are two symbolic cases that exemplify the black box problem and the unpredictable behaviors arising from AI self-learning and the unreliability of the results generated by the algorithms. During an experiment conducted in 2002 by scientists at the Magna Science Center, in England, an unforeseen event occurred: two intelligent robots were placed in an arena to simulate a scenario of 'predators' and 'prey', in order to see if the robots would be able to benefit from the experience acquired from machine learning to develop new hunting and self-defense techniques. However, Gaak, one of the robots, that was unintentionally left unattended for 15 min managed to escape and it adopted an unpredictable behavior, founding a way out through the arena wall and reached the parking lot, where it ended up being hit by a car (Čerka et al. 2015).

It is also relevant to mention the incident reported by Sameer Singh, an assistant professor in the Department of Computer Science at the University of California (UCI), in the United States, in which a student created an algorithm to categorize pictures of huskies and wolves. Initially, it seemed that the algorithm was able to classify the two animals almost perfectly. However, after numerous and subsequent cross-analysis, Singh found out that the algorithm was identifying wolves based only on the snow in the background of the images and not on the animal's own characteristics (UCI Beall Applied Innovation 2017).

Undoubtedly, damages could rise to immeasurable levels if we consider the risks presented in the two cases above in the context of AI algorithms in Medicine. Now take for example a poorly programmed and tested algorithm, or one with expressive degree of fallibility, in the cognitive technology that was used in some countries to diagnose patients infected with the new coronavirus. Because of this, Nicholson Price and Roger Allan Ford explain that one of the biggest fears of the healthcare sector at this stage of artificial intelligence stems precisely from the unpredictable situations arising from black box medicine, given the obscurity in the way information is processed by the algorithms (Ford and Price 2016). Therefore, when algorithmic systems are implemented in clinical practice, it is essential that physicians know their limitations and what is effectively taken into account for predictions. Understanding the limits of algorithms will help physicians to better judge their decisions and proposals, thus avoiding simplistic and reductionist views, in addition to preventing patients from becoming 'hostages' of automated decisions made in the black box of algorithms.

In addition, it is necessary to emphasize that AI in diagnostic analysis is not perfect. No matter how efficient an 'intelligent' system is for medical diagnosis and clinical predictions, it will continue to present a significant margin of inaccuracy, which can lead to adverse results. For example, Watson for Oncology is not 100% accurate. There is a significant inaccuracy margin of around 10%, according to a clinical research conducted by a team of 15 doctors at Manipal Hospitals in India over 3 years of 1000 patients diagnosed with cancer. In cases where there was disagreement between the AI and the doctors, the medical professionals changed in 63% of the cases their own diagnoses to follow the one given by Watson. There is a central point for this reflection: the AI system altered the final decision of oncologists in several cases. On the other hand, the same survey revealed that in

37% of cases the physicians did not change its own diagnosis, in disagreement with the result obtained by Watson (Bicudo 2021).

In this scenario, imagine a patient diagnosed with cancer and his doctor believe, at first, that he has a certain type of cancer. However, after putting the patient's clinical data into predictive software, such as Watson for Oncology, this one gives another result, saying the patient has a different type of cancer. Then, a question arises from it: if the physician follows or disregards the result of the AI, and damage to the patient occurs, after inappropriate diagnosis and treatment, should the professional be held responsible? In other words, would be possible to consider a case of medical malpractice in the event of the supervenience of a harmful result for the patient that, in theory, could be avoided, if the diagnosis proposed by the AI had been followed? This complex issue has already been discussed in recent papers (Nogaroli and Silva 2020; Nogaroli and Nalin 2021).

In order to answer properly this question, some basic concepts need to be initially indicated about medical liability for misdiagnosis. For the purposes of analyzing liability in AI services, the primary element of a medical malpractice claim is the breach of a legal duty to adhere to a professional standard of care, which is 'a set of guidelines specifying the appropriate or required treatment methods for a given condition based on medical research and professional practice' (Jorstad 2020). Moreover, in most jurisdictions, the law does not hold doctors legally responsible for all diagnostic errors. A misdiagnosis or delayed diagnosis itself is not evidence of medical negligence (Kfouri-Neto 2021). Skillful professionals can make diagnostic errors even when using reasonable care. When the doctor carries out a good examination of his patient, with all the healthcare data, medical exams and means available, and still makes a diagnostic error, the professional will not be held responsible. An obligation of infallibility or absolute accuracy cannot be imposed on the physician.

However, when the misdiagnosis is gross, revealing inadmissible ignorance or negligence, it leads to the medical liability. The inexcusable diagnostic error can come from several causes: (a) superficial examination of the patient; (b) inexcusable ignorance of the physician with elementary information from medical science; (c) not resorting to the auxiliary diagnostic means made available to the professional; (d) disregards evident symptoms that required additional exams for a better determination of the clinical condition. Thus, the key is determining whether the physician acted competently, which involves an evaluation of what the professional did and did not do in arriving at a certain diagnosis.

When analyzing the problem of medical liability for diagnostic errors in the context of AI, according to lessons from Nicholson Price, the doctor can be held responsible if he is not diligent in using the technology (Price et al. 2019). In the same sense, Fruzsina Molnár-Gábor argues that if doctors recognize, based on their medical expertise, that the result provided by the AI is incorrect in that specific case, they should not consider it as a basis for their clinical decision. On the other hand, the physician's lack of diligence in thoughtlessly discarding the result obtained by the AI system may constitute a criterion for liability (Molnár-Gábor 2020).

Thus, it is possible to conclude that, in order to verify whether a doctor has acted with negligence in a specific case, the standards of professional conduct required at the time of medical practice must be analyzed. In summary, the physician using the technology will be in a difficult position to justify: (1) why he followed the diagnosis or course of action suggested by the AI or (2) why—and based on what factors—he deviated from the algorithmic recommendation. The medical professional is free to choose his means of diagnosis and therapy proposals, but he is also responsible for his choices (Nogaroli and Nalin 2021).

Beyond that, when algorithmic systems are implemented in clinical practice, it is essential that physicians know their limitations and what is taken into account for algorithm predictions. Understanding the limits of the technology will help physicians to better judge their decisions and proposals, thus avoiding simplistic and reductionist views based in the black box of the algorithms. Lack of in-depth knowledge of the benefits and risks of healthcare technologies can translate into worse outcomes for patients due to a lack of medical understanding about which tools add value to their practice or how to properly integrate AI into the clinical workflow.

As an example, some hospitals in US implemented the so-called *AI Dying Algorithms*, which use patient's health data and analyze around 5000 clinical risk factors to predict the chances of survival among hospitalized individuals, screening patients with palliative needs or even determining the time until death of patients with terminal or incurable diseases. There are potential benefits of these algorithms as a tool to support medical decision in the indication of palliative care, in order to avoid undue extension of life and provide terminal patients with the option of living the end of life with better quality, through the indication of palliative care.

However, it is possible to observe expressive complications with this type of AI algorithm such as the one called *Jvion CORE*, created by the company Jvion for medical decision in the indication of palliative care. It has already been implemented in several oncology clinics in the United States (Jvion CORE 2021). However, there is a serious problem of using the *AI Dying Algorithms* in clinical practice. *Jvion CORE* presents an approximately 40% accuracy in its predictions about patients flagged as high risk to die in the following month. In order words, there is expressive percentage of 60% of algorithmic fallibility (Robbins 2021).

Therefore, Eric Topol states that algorithms can help patients and their physicians make decisions about the course of medical treatment, both in palliative situations and in those where cure is the goal (Topol 2019a, b). However, the author states that there is not 'a particularly good use for AI unless and until it is shown that the algorithm being used is extremely accurate' (Robbins 2021). Besides, there is also a mismatch between the task of these models: predicting a patient's odds of death and how they're actually being used to try to identify who will benefit most from an advance care planning conversation (medical recommendation for palliative care). Consequently, there are considerable doubts about the role that artificial intelligence can play in the context of palliative care (Nogaroli and Kfouri-Neto 2021b).

Last but not least, there is another relevant effect of *epistemic opacity* that deserves special consideration. Physicians have a legal duty to provide a certain

standard of skill and care to their patients but have no obligation under law to guarantee the cure or other concrete results. Though, there is a risk that the physician does not understand the limitations of the AI system, using it as an end in itself—not as a tool—and, more than that, pass on the guarantee of total success to his patient precisely because of the technology used. Then, it arises the discussion about the possibility of considering a medical obligation of result, based on the promise of infallibility of the AI tool used in the clinical practice. As an illustration, it is worth mentioning that was discussed in the US about physicians who used the Da Vinci robotic platforms in surgeries and ensured positive results for patients, providing information only about the benefits of the technological tool (Nogaroli and Kfouri-Neto 2019, 2020).

The same logic seems to be applicable for the hypothesis of the physician using Artificial Intelligence tools, such as IBM's Watson, creating in the patient the expectation that he will have an extremely accurate diagnosis of cancer and the best treatment proposal due to the use of AI, which, acting better than human beings, would be able to bring about a favorable result, practically guaranteeing the cure. In this scenario, there would be a violation of the ethical principle of 'human control of technology', since the professional does not understand *AI-as-a-tool* to support clinical decision-making, bringing the technology as a guaranteed success in medical practice. This result in the breach of the patient's legitimate expectation and the possible qualification of the legal obligation nature for the doctor as an obligation of result.

(2) *opacity for the lack of medical disclosure*: in the second semantic dimension of algorithmic opacity particularly relevant to Medicine, it is observed that there is considerable risk that AI algorithms are used to support medical decision without the patients' knowledge, and patient's unawareness that automated decision-making and profiling activities about them are being carried out with their personal data. In this scenario, first of all, it is important to consider that medical disclosure is the structured process of transparent communication between patients and physicians involved during medical care. However, a lot of criticisms have arisen because patients are often not informed or asked to consent to the use of Artificial Intelligence algorithms in their health care (Robbins and Brodwin 2021). In fact, some physicians use a paternalistic discourse that they don't need to inform patients about all the resources used in the clinical decision process. Following this logic, the medical professional, in theory, could indicate palliative care for a patient, informing some aspects of their clinical condition and making medical recommendations without the need to disclose the specific information about the use of an *AI Dying Algorithm* (Cohen 2020).

Though, informing the patient and providing their consent represents one of the mechanisms for the realization of the fundamental right to the free development of the human personality, having an instrumental nature as it is a way of realizing the right to autonomy. Currently, the modern doctrine around the globe about medical liability defends patients' consent as an instrument that allows, in addition to the interests and medical-therapeutic objectives, to increase respect for the person in its holistic dimension. Patients need to be provided with the essential

information to properly understand his health condition or possible treatments available, so that he can exercise the faculty of consenting to the proposed treatment or intervention, choosing another of the existing alternatives, although less indicated by the attending professional, or even refusing to be treated. This doctrinal notion is a trend of thought that has taken shape in various jurisdictions around the globe in the last decades (Pereira 2004).

Thus, the type of algorithmic opacity due to non-disclosure does not concern the intrinsic characteristics of AI systems but has its origin in the risks to the patient's informative self-determination, that is, it derives from the way in which the medical decision regarding the diagnosis, prognosis and treatment proposals supported by AI can be carried out by the physician without the patients being aware of it, neither during the medical intervention nor after the harmful event. It is important to consider that there may be medical liability due for the deprivation suffered by the patient in his self-determination, because he was deprived of the opportunity to ponder the risks and advantages of an AI algorithm prediction about his clinical condition (Nogaroli and Dantas 2020). In conclusion, physician must inform the patient about the fact that the diagnosis, prognosis, treatment proposal or even his indication for palliative care are supported by several factors and resources, including an Artificial Intelligence algorithm (Nogaroli and Dantas 2021). This includes the ideal of shared decision-making in medicine.

(3) *explanatory opacity*: in addition to the physician's duty to disclose the information that he uses an AI algorithm to support his clinical decision, he also needs to explain about the technology used, according to the degree of understanding of each patient. If patients don't received this properly explanation in Medical AI, it may occur the so-called *explanatory opacity*. There is a divergence in the doctrine about the amount of information that must be given to the patient, in order to the doctor comply with his duty to inform. However, we have already defended in a recent paper that, with the evolution of new technologies in the healthcare sector, physicians need to understand that the right to adequate information (which corresponds to a duty to inform) also includes consent to the use of new technologies, based on the patient's knowledge of their functioning, objectives, advantages, costs, risks and alternatives (Nogaroli and Dantas 2020). Thus, there is a demand for a new interpretation of the principle of patient self-determination in the context of new technologies: we moved away from the simple right to receive medical information, and we are going towards a greater informational range, since there is a right to explanation and justification (Astromskė et al. 2020).

Therefore, if we go back to Watson for Oncology's factual hypothesis, even if medical negligence is not configured, if the professional only informs—*but does not adequately explain* to the patient about the use of the technology to support the medical decision, he may be held responsible for the deprivation suffered by the patient in his self-determination, since the opportunity to consider the advantages and risks of treatment proposed or medical diagnosis supported by AI algorithms was taken away from the patient.

Explainability can be understood as "a characteristic of an AI-driven system allowing a person to reconstruct why a certain AI came up with the presented

predictions" (Amann et al. 2020). Nevertheless, it is essential to point out that explainability is not a purely technological issue, instead it invokes a host of medical, legal, ethical, and societal questions that require thorough exploration. Taking AI-based clinical decision support systems as a case in point, there is an ethical and legal obligation for the doctor to inform and explain for his patient something like: 'look Mr. John, at first I see that your clinical condition indicates that you have a specific type of cancer, but we tried a certain chemotherapy treatment without much success. Therefore, we could put your personal data into Watson for Oncology and the AI would make a cross-reference with its huge database, in order to show us an eventual diverse diagnosis, or bring others recent treatments proposals based on confidence levels. But look Mr. John, Watson has a certain fallibility degree, and it has other risks...'.

This is the appropriate model of the process of obtaining the patient's consent in AI, explaining and dialoguing with him to clarify the nuances of the diagnosis and prognosis process supported by the technology. To sum up, for the physician not to be held responsible for violating the duty to inform, it is essential to pay special attention to the process of obtaining informed consent, converting it into an *informed choice process*, following the idea of a true process of dialogue between doctor and patient. From the beginning of the decision to use the AI-based algorithm, there is a need for an explanation and justification for those affected by the technology.

3 Ethical Dimensions of Using Artificial Intelligence (AI) in the Healthcare Sector: Setting the Parameters for Data-Informed Duties in Tort Law

The development and implementation of AI tools in Medicine are opening the doors to new ethical and legal challenges. These challenges include how to evaluate algorithm performance and to determine where AI can be safely and efficiently applied to clinical practice. There are three examples of ethical issues relate to: "(1) Biases in training data; (2) The potential replacing of human health care providers with AI tools; (3) Responding to an AI intervention that has failed. If we develop an AI tool that influences a clinical decision, and a poor decision was made, how do we (as humans) respond?" (Marcetich 2020). As mentioned before, designing machine learning tools used to support clinical decision-making can be thought of as an experiment whose risks need to be carefully evaluated before implementation in clinical practice.

In June 2021, the World Health Organization (WHO) published its guidance on *Ethics and Governance of Artificial Intelligence for Health* (World Health Organization 2021). The report reflects the WHO's intention to anchor their guidance within a human rights framework and it makes direct references to the Universal Declaration of Human Rights by exploring the question of autonomy, protecting populations from harm, and ensuring inclusiveness and equity. It states

that 'ethical considerations and human rights must be placed at the center of the design, development, and deployment of AI technologies for health'. The document offers 6 primary principles for the use of AI in Medicine: (1) protect autonomy; (2) promote human well-being, human safety and the public interest; (3) ensure transparency, explainability and intelligibility; (4) foster responsibility and accountability; (5) ensure inclusiveness and equity; (6) promote AI that is responsive and sustainable.

Another strong point of the report is its detailed analysis of the risks and limitations of AI. Two major problems are raised: (1) the potential of discrimination; and (2) bias when datasets used to train AI fail to reflect the real world, and there is a lack of transparency in the data source used to program these algorithms, without the explanation of how they cross-reference the data and effectively reach to a certain result. In fact, nowadays AI is booming in Medicine but it's also facing a credibility crisis because the algorithms are 'often trained on small, single-origin data samples with limited diversity; some even reused the same data for training and testing, a cardinal sin that can lead to misleadingly impressive performance' (Ross 2021a, b).

The failure to test AI models on data from different sources—a process known as external validation—is common in studies published in leading medical journals. According to a research team from the University of Cambridge in England, an ever-growing list of papers rely on 'limited or low-quality data, fail to specify their training approach and statistical methods, and don't test whether they will work for people of different races, genders, ages, and geographies' (Ross 2021a, b). This results in an algorithm that appears highly accurate in a specific study, but does not work to the same level of accuracy when exposed to real-world variables, across different types of patients in several locations.

In a recent interview, Eric Topol presented worries about how AI might worsen some inequities and discrimination, since 'algorithms are not biased, but the data we put into those algorithms, because they are chosen by humans, often are' (Time Magazine 2019). There is the potential risk for discrimination of the AI algorithms in Medicine, since they can be programmed based on data from scientific studies and electronic health records of certain populations where some races predominate. Thus, there is a risk that decisions are contaminated by significant biases (Ledford 2019; Obermeyer et al. 2019). As an example, it is argued that black women with breast cancer are more likely to be diagnosed late by the FDA approved algorithms in the market, precisely because they were programmed with data from a population where probably did not have black women, or it had very few (Brodwin 2021). This is something very serious and important to reflect on, since programming the algorithms with healthcare data from different populations and geographic locations is essential, considering the expressive variations in the way the diseases manifest in different races.

Furthermore, in a recent study, it was found that between 2012 and 2020 only 73 of 161 AI products approved by Food and Drug Administration (FDA) in the US have publicly disclosed the amount of data used to validate the product, with only 7 of them reporting the racial makeup for their study populations. Moreover, among 10 AI products approved for breast imaging, only 1 publicly disclosed the racial

demographics of the dataset used to detect suspicious lesions and assess cancer risk (Ross 2021a, b). In another study conduct by Stanford University between January 2015 and December 2020, it was observed that almost all of the FDA approved AI devices (126 of 130) underwent only retrospective studies at their submission. None of the 54 high-risk devices were evaluated by prospective studies and only 17 device studies reported that demographic subgroup performance was considered in their evaluations. It was concluded in this second study that more than the importance of evaluating the performance of AI devices in multiple clinical sites and across representative populations, it is also essential encouraging prospective studies. The reason for this conclusion is that 'prospective studies with comparison to standard of care reduces the risk of harmful overfitting and more accurately captures true clinical outcomes. Post-market surveillance of AI devices is also needed for understanding and measurement of unintended outcomes and biases that are not detected in prospective, multi-center trials' (Wu et al. 2021).

Discussion about the need for specific regulation regarding algorithms is a recurrent doctrinal theme in a lot of areas, including the healthcare sector (Benjamens et al. 2020). Its impacts challenge the understanding of the State's own role in controlling technological development. If, on the one hand, it is expected that innovation will bring improvements to the overall quality of life, on the other hand, there is no denying that facing the issue from a regulatory point of view is a challenge (Tomasevicius and Ferraro 2020). Structuring a comprehensive approach to assess the current state of technological development does not seem like a plausible path for some more detailed demands and discussions about law-making affairs in this complex scenario, whereas tort law doctrine has been seeking to establish a systematic model for the delimitation of risk assessment contours in the development of applications centered on Artificial Intelligence systems.

Frank Pasquale suggests the parameterization of *data-informed duties* for the creation of standard models that may support accountability assessments. In the author's words, 'such standards are particularly important given the potential for inaccurate and inappropriate data to contaminate machine learning' (Pasquale 2019). In this respect, it appears that data-driven heuristic process, if contaminated early in the processing stages, might generate biased results. In other words, data curation of inputs must prevail and be observed throughout the entire algorithmic processes—which must also be auditable—otherwise the final substrates obtained after processing such data (the so-called 'outputs') might not be reliable.

Essentially, the parameterization of standard models no longer depends on regulatory efforts for the vast array of algorithmic structures, which vary in several aspects, and offers greater freedom for the development of self-regulated metrics for each type of activity. In this context, it would be possible to work with comparative bases that would offer more precise and well-mapped conditions to determine the performance in compliance with the equivalent risk duly measured for the type of algorithmic activity in question.

Stuart Russell and Peter Norvig's had already dealt with the troublesome 'quantification of uncertainties' in the context of AI algorithm's predictions: 'Agents may need to handle uncertainty, whether due to partial observability, uncertainty

nondeterminism, or a combination of the two' (Russell and Norvig 2016). In summary, the conjectures from which data-informed duties are conceived are in line with a very important guideline, proposed by Frank Pasquale as the 'fourth law of robotics' (*explainability*) (Pasquale 2017). His idea reinforces the need to overcome the *black box problem* (Pasquale 2015). As mentioned before, this is a problem usually identified by the use of machine learning techniques that provide uncontrolled and unsupervised improvement of these applications, to the point of becoming so complex that even their own creators do not understand them (Asaro 2011).[5]

Civil liability deals with uncertainty and the unpredictability. Traditionally, such derive from the application of integral risk theory as a basis for redressing torts specifically based on guardianship dangers and the precautionary principle (Calo 2015). The same logic, if transferred to the context of AI algorithms, would provide some peculiar consequences. On the subject, Yaniv Benhamou and Justine Ferland have already pointed out five observations about the data-informed duties (Benhamou and Ferland 2021).

1. A first observation of the authors is that, with regard to the requirements imposed on algorithmic actors (owner, operator, retailer and designer) (Balkin 2015), it is necessary to comply with duties of care, which concern: (a) the choice of a particular technology, in light of the tasks that need to be performed and the operator's own skills and abilities; (b) the planned organizational framework, in particular with regard to adequate follow-up; and (c) maintenance, including safety checking routines. Failure to comply with such obligations could trigger strict liability, regardless of whether the operator is also responsible for creating or elevating the risks of a certain technology (Benhamou and Ferland 2021). Considering this, it seems to be also important for physicians or hospitals—in a position of algorithm operator—to comply with these duties.
2. Benhamou and Ferland also point out that manufacturers, including those who act incidentally as algorithmic supervisors,[6] must observe the following standards of conduct (Benhamou and Ferland 2021): (b.1) design, describe and market products in a way that allows them to fulfill data-informed duties, making risks more predictable (foreseeability) (Karnow 2016)[7]; and (b.2) properly

[5] Commenting on the practical difficulties of the difficulty of identifying the developer creator, see.

[6] In Brazil, the concept is found in article 5, item VII, of the LGPD: "Art. 5th. (...) VII - operator: natural or legal person, under public or private law, who processes personal data on behalf of the controller." [Originally: "Art. 5°. (...) VII - operador: pessoa natural ou jurídica, de direito público ou privado, que realiza o tratamento de dados pessoais em nome do controlador."]

[7] The author points out that "predictability and foreseeability are, in practice, vague and peculiar notions, and people with different experiences and beliefs about how the world works will treat different things as "predictable." In any event humans are poor at predicting odds, and generally are not accurate estimating the likelihood of future events. Perhaps we may get better at predicting the behavior of autonomous robots as we interact with them; actions that appear at first random may begin to cluster in their frequencies, revealing theretofore unanticipated patterns that will help future prediction."

monitor the product after it has been put into circulation, in light of the characteristics of emerging digital technologies, in particular due to their openness and dependence on the general digital environment, including obsolescence, the emergence of malware or even its vulnerability to possible external attacks.

3. The so-called *supervision*, in the context of monitoring specific duties that hierarchically superior may even be due to the administrative police power of the State (Scherer 2016), in what Pasquale calls 'oversight' in his newest book (Pasquale 2020). That could be achieved by carrying out audits and studies of the specific algorithm, even after its market release. Thus, as a result of the implementation of supervised monitoring systems, the identification of anomalies and the prior parameterization of the systems would be expected to 'warn' about the occurrence of unexpected behaviors, as well as the observation of specific evolution trends from machine learning to predict such behaviors. Once such monitoring is implemented, the obligation to inform potential victims appears as a duty attached to objective good faith (Wischmeyer 2020).
4. If feasible, the authors argue that producers should be compelled to include mandatory backdoors in their algorithms (Liao et al. 2020). Other designations for this are the expressions 'emergency brakes by default (or by design)', 'shut down features', or features that allow operators or users to 'turn off the AI' by manual commands, or make it 'unintelligent' by simply pressing a panic button. Failure to guarantee such tools and control options could be considered a design defect to justify a breach of the general precautions that are to be expected of them, opening up the possibility of imposing civil liability due to the fact that the algorithm is to be considered faulty. In fact, depending on the circumstances, manufacturers or operators could also be forced to 'turn off' the AI as part of their algorithmic monitoring and auditing tasks.
5. Similar to existing after-sales duties, which are composed of warnings and instructions for recalling defective products, producers/manufacturers might also assume support and correction duties—corollaries of auditability and transparency principles (Pasquale 2019)—in line with other recent developments on the potential obligation of software developers to update unsafe algorithms, for as long as the technology is on the market (i.e., beyond any contractual stipulations on warranty period) (Wolters 2019).

Frank Pasquale investigates the potential liability in the context of the use of inaccurate or inappropriate data (faulty data) in training sets for machine learning: 'firms using faulty data can be required to compensate those harmed by that data use—and should be subject to punitive damages when such faulty data collection, analysis, and use is repeated or willful (Pasquale 2019).' The punitive function of civil liability raises controversial aspects to be considered in the context of this brief study. This is because, particularly in the common law experience, punitive and dissuasive benefits have a wider application and are accepted, both by the doctrine and by the Courts. Although the topic is controversial and even though punitive damages are only one of the various options to consider a deterrent effect of potential

liability, it is inevitable to observe the relevance of the discussion to the complex technological context in which Artificial Intelligence algorithms are inserted.

Preserving the complementarity of tort law and regulation of data collection, analysis, and use is very appropriate to help it avoid preventable accidents and expands opportunities for those harmed by new technologies to demand accountability (Faleiros Júnior 2021). Nowadays, tort law is moving towards to promote not only liability but also accountability, which has a prospective function and is more robust and based on multiple functions, especially the precautionary one. This scenario presented by Pasquale reinforce, in one hand, the important concern with the desirable compliance, considered from governance structures and data curation aimed at the continuously verification of the quality of the collection used into the AI algorithms. On the other hand, this context turns out to a triple reflection: (1) if it is possible to assume that AI diagnosing will be covered under health providers' current malpractice insurance policies, or if the introduction of AI diagnosing into clinical practice will likely prompt insurance providers to decline coverage for such activities; (2) the potential civil liability of the physician as an algorithmic operator who repeatedly observes its ineffectiveness or becomes aware that the AI uses biased data collections (faulty data); (3) the importance of medical education in AI, digital health and new technologies to prevent adverse events.

4 Concluding Notes: The Future of Artificial Intelligence (AI) in Medicine and the Importance of Medical Education in Digital Health and New Technologies

It was observed in the present study that the valuable development of P4-Medicine from the use of predictive algorithms cannot be unaccompanied by the need for reflection about the risks and a special medical diligence in using the technology as a tool to support decision making. Moreover, it was concluded that there are three different semantic dimensions of algorithmic opacity relevant to Medicine: (1) *epistemic opacity* for the insufficient physicians understanding of the rules an AI system is applying to make predictions and decisions; (2) opacity for the *lack of medical disclosure* about the use of AI systems and patient's unawareness that automated decision-making and profiling activities about them are being carried out with their personal data; and (3) *explanatory opacity* for the unsatisfactory explanation to patients about the technology used to support professional decision-making. Therefore, the aim of this paper was to analyze each type of opacity, considering hypothetical scenarios and its repercussions in terms of medical malpractice and patient's informed consent.

Regarding *epistemic opacity*, questions were presented about to what extent a doctor might rely on AI and the legal consequences if the physician adhered to the recommendation or overruled the machine, leading to the significant consideration about the determination of the standard of medical diligence must be an issue always

open to debate in each specific medical malpractice case. This is because, in each situation, the degree of accuracy of an algorithm and its goal are different. It could also be concluded in the present paper that there is a possibility of qualifying the physician's obligation as an obligation of result when there is a violation of the ethical principle of 'human control of technology', that is, in the face of non-understanding of AI as a tool to support clinical decision-making (AI-as-a-tool), with the consequent breach of the patient's legitimate expectation of technology as a guarantee of success. It was also observed that *opacity for the lack of medical disclosure* and *explanatory opacity* demand reflections on the impact of the ethical principles of explanation and justification, in order to understand a new model of patient consent in AI and the violation of the medical duty of qualified information.

In this context, the above-mentioned issues—specially, the consequences of the triple semantic dimensions of algorithmic opacity—represent an enormous challenge to educators in the health sciences. AI can help medical professionals by amalgamating large amounts of healthcare data and supporting their decision-making process about diagnosis and recommend treatments. Nevertheless, physicians need the ability to interpret the results and properly communicate a recommendation to the patient. Physicians need to learn how to better use and interpret AI algorithms, including in this learning process the comprehension of in which situations an algorithm should be effectively used in their practice, and, above all, how much confidence should be placed in an algorithmic recommendation, in each concrete case.

Thus, new skills and expertise are required as we move to an age of Artificial Intelligence in the healthcare environment. Physicians' lack of in-depth knowledge about the benefits and risks of healthcare technologies can translate into worse outcomes for patients due to little or none understanding of which AI tools add value to their activities or how to integrate AI in a way suitable for the clinical workflow. This task calls for a new model of educating the new generation of experts with deep interdisciplinary training in Medicine, ethics, and technologies. Therefore, AI needs to be seamlessly integrated across different aspects of the medical education curriculum.

The American Medical Association (AMA) noted that from 2000 to 2015 there were 15 national reports calling for medical education reform (Beck 2015). In US, there are several initiatives for incorporating new technologies—such as AI tools—in medical education: (1) *Duke Institute for Health Innovation*: medical students work together with data experts to develop care-enhanced technologies made for physicians; (2) *University of Florida*: radiology residents work with a technology-based company to develop computer-aided detection for mammography; (3) *Carle Illinois College of Medicine*: offers a course by clinical scientists and engineers to learn about new technologies; (4) *Sharon Lund Medical Intelligence and Innovation Institute*: organizes a summer course on all new technologies in health care, open to medical students; (5) *Stanford University Center for Artificial Intelligence in Medicine*: involves graduate and postgraduate students in solving heath care problems with the use of machine learning (Paranjape et al. 2019); (6) *Rocky Vista University College of Osteopathic Medicine*: offers courses to train medical

students in AI, remote monitoring, ethics, informatics, telemedicine, analytics, and entrepreneurship (Aungst and Patel 2020).

In conclusion, an overriding issue for the future of AI in Medicine rests with how well medical education can be assured. As AI and its application become mainstream in the healthcare sector, medical students, residents, fellows, and practicing physicians need to have better knowledge of AI. The integration of digital health into formal education offers a novel means to engage in interprofessional education opportunities. Determining how to build out digital health education and to integrate into the formal curriculum will be a topic of debate in the coming years. To ensure that AI-based clinical decision lives up to its promises, there is a need to sensitize developers, healthcare professionals, and legislators to the challenges and limitations of opaque algorithms in the healthcare sector and to foster medical education moving forward to the Age of Artificial Intelligence in Medicine.[8]

References

Amann J, Blasimme A, Vayena E, Frey D, Madai VI (2020) Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 20:310

Asaro P (2011) A body to kick, but still no soul to damn: legal perspectives on robotics. In: Lin P, Abney K, Bekey G (eds) Robot ethics: the ethical and social implications of robotics. The MIT Press, Cambridge, pp 169–186

Astromskė K, Peičius E, Astromskis P (2020) Ethical and legal challenges of informed consent applying artificial intelligence in medical diagnostic consultations. AI Soc 36:509–520

Aungst TD, Patel R (2020) Integrating digital health into the curriculum—considerations on the current landscape and future developments. J Med Educ Curric Dev 7:238212051990127

Balicer RD, Cohen-Stavi C (2020) Advancing healthcare through data-driven medicine and artificial intelligence. In: Nordlinger B, Villani C, Rus D (eds) Healthcare and artificial intelligence. Springer, Cham, pp 9–15

Balkin JM (2015) The path of robotics law. Calif Law Rev Circuit 6:45–60

Beck M (2015) Innovation is sweeping through U.S. medical schools. Wall Street J. https://www.wsj.com/articles/innovation-is-sweeping-through-u-s-medical-schools-1424145650. Accessed 17 Oct 2021

Benhamou Y, Ferland J (2021) Artificial intelligence & damages: assessing liability and calculating the damages. In: D'Agostino P, Piovesan C, Gaon A (eds) Leading legal disruption: artificial Intelligence and a toolkit for lawyers and the law. Thomson Reuters Canada, Toronto, pp 1–20

Benjamens S, Dhunnoo P, Meskó B (2020) The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Digit Med 3:118

[8] *See* generally, on AI and Healthcare, in this book A T Freitas—Data-driven approaches in healthcare: challenges and emerging trends; M S Fernandes and J R Goldim—Artificial Intelligence and Decision Making in Health: Risks and Opportunities; and M N Duffourc and D S Giovanniello—The Autonomous AI Physician: Medical Ethics and Legal Liability. *See* also, on the black box effect, in this book E Magrani and P G F Silva—The Ethical and Legal Challenges of Recommender Systems Driven by Artificial Intelligence; and M N Duffourc and D S Giovanniello—The Autonomous AI Physician: Medical Ethics and Legal Liability.

Bicudo L (2021) Inteligência artificial descobre 1.000 casos de câncer com precisão de 90%. https://www.startse.com/noticia/nova-economia/tecnologia-inovacao/inteligencia-artificial-descobre-1-000-casos-de-cancer-com-precisao-de-90. Accessed 8 Oct 2021

Brodwin E (2021) Google debuts an AI-powered tool to analyze skin conditions. https://www.statnews.com/2021/05/18/google-dermatology-assist-skin-app/. Accessed 10 Oct 2021

Calo R (2015) Robotics and the lessons of cyberlaw. Calif Law Rev 103:513–563

Čerka P, Grigienė J, Sirbikytė G (2015) Liability for damages caused by artificial intelligence. Comput Law Secur Rev 31:376–389

Cohen G (2020) Informed consent and medical artificial intelligence: what to tell the patient? Harv Public Law Work Pap 108:1425–1469

Faleiros Júnior JLM (2021) Discriminação por algoritmos de Inteligência Artificial: a responsabilidade civil, os vieses e o exemplo das tecnologias baseadas em luminância. In: Barbosa MM, Braga-Netto F, Silva MC, Faleiros Júnior JLM (eds) Direito digital e inteligência artificial. Foco, Indaiatuba, pp 969–1000

Flores M, Glusman G, Brogaard K, Price ND, Hood L (2013) P4 medicine: how systems medicine will transform the healthcare sector and society. Pers Med 10:565–576

Ford RA, Price N (2016) Privacy and accountability in black-box medicine. Mich Telecommun Technol Law Rev 23:1–43

Harmon SA, Sanford TH, Xu S, Turkbey EB, Roth H, Xu Z, Yang D, Myronenko A, Anderson V, Amalou A, Blain M, Kassin M, Long D, Varble N, Walker SM, Bagci U, Ierardi AM, Stellato E, Plensich GG, Franceschelli G, Girlando C, Irmici G, Labella D, Hammoud D, Malayeri A, Jones E, Summers RM, Choyke PL, Xu D, Flores M, Tamura K, Obinata H, Mori H, Patella F, Cariati M, Carrafiello G, An P, Wood BJ, Turkbey B (2020) Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. Nat Commun 11:4080

Holzinger A, Röcker C, Ziefle M (2015) From smart health to smart hospitals. In: Holzinger A, Röcker C, Ziefle M (eds) Smart health: open problems and future challenges. Springer, Cham, pp 1–20

Hood L (2013) Systems biology and P4 medicine: past, present, and future. Rambam Maimonides Med J 4:e0012

Huang S, Yang J, Fong S, Zhao Q (2020) Artificial intelligence in cancer diagnosis and prognosis: opportunities and challenges. Cancer Lett 471:61–71

IBM Healthcare and Life Sciences (2021) The future of health is cognitive. Harnessing data and insight to deliver better health, value and individual engagement. https://www.ibm.com/downloads/cas/LQZ0O1WM. Accessed 17 Sep 2021

Jorstad KT (2020) Intersection of artificial intelligence and medicine: tort liability in the technological age. J Med Artif Intell 3:1–17

Jvion CORE (2021) Care optimization and recommendation enhancement. https://jvion.com/approach/the-jvion-core/. Accessed 20 Oct 2021

Karnow CEA (2016) The application of traditional tort theory to embodied machine intelligence. In: Calo AR, Froomkin M, Kerr I (eds) Robot law. Edward Elgar, Cheltenham, pp 51–77

Kfouri-Neto M (2021) Responsabilidade civil do médico. Thomson Reuters Brazil, São Paulo, pp 97–100

Ledford H (2019) Millions of black people affected by racial bias in health-care algorithms. Nature 574:608–609

Liao C, Zhong H, Squicciarini A, Zhu S, Miller D (2020) Backdoor embedding in convolutional neural network models via invisible perturbation. In: Proceedings of the 10th ACM conference on data and application security and privacy. ACM, New York, pp 97–108

Marcetich M (2020) Data pulse: a brief tour of artificial intelligence in healthcare. New Degree Press, Washington, DC

Molnár-Gábor F (2020) Artificial intelligence in healthcare: doctors, patients and liabilities. In: Wischmeyer T, Rademacher T (eds) Regulating artificial intelligence. Springer, Cham, pp 337–360

Nogaroli R, Dantas E (2020) Consentimento informado do paciente frente às novas tecnologias da saúde (telemedicina, cirurgia robótica e inteligência artificial). Lex Med Rev Port Direito Saúde 17:25–63

Nogaroli R, Dantas E (2021) The rise of robotics and artificial intelligence in healthcare: news challenges for the doctrine of informed consent. Med Law 40:15–61

Nogaroli R, Kfouri-Neto M (2019) Responsabilidade civil pelo inadimplemento do dever de informação na cirurgia robótica e telecirurgia: uma abordagem de direito comparado (Estados Unidos, União Europeia e Brasil). In: Dadalto L, Bezerra J, Rosenvald N (eds) Responsabilidade Civil e Medicina. Foco, Indaiatuba, pp 173–203

Nogaroli R, Kfouri-Neto M (2020) Estudo comparatístico da responsabilidade civil do Médico, hospital e fabricante na cirurgia assistida por robô. Essay. In: Nogaroli R, Kfouri-Neto M (eds) Debates contemporâneos em direito médico e da saúde. Thomson Reuters Brazil, São Paulo, pp 33–67

Nogaroli R, Kfouri-Neto M (2021a) Algoritmos de inteligência artificial na predição do quadro clínico de pacientes e a responsabilidade civil médica por omissão de cuidados paliativos. In: Dadalto L (ed) Cuidados paliativos: aspectos jurídicos. Foco, Indaiatuba, pp 163–190

Nogaroli R, Kfouri-Neto M (2021b) Inteligência artificial nas decisões clínicas e a responsabilidade civil médica por eventos adversos no contexto dos hospitais virtuais. In: Barbosa MM, Braga-Netto F, Silva MC, Faleiros Júnior JLM (eds) Direito digital e inteligência artificial: diálogos entre Brasil e Europa. Foco, Indaiatuba, pp 1079–1107

Nogaroli R, Nalin P (2021) Diagnóstico para covid-19 com inteligência artificial: novos desafios sobre princípios contratuais e responsabilidade médica em tempos de pandemia. Rev Eletrônica Direito Cent Univ Newton Paiva 43:256–279

Nogaroli R, Silva RG (2020) Inteligência artificial na análise diagnóstica: benefícios, riscos e responsabilidade do médico. In: Nogaroli R, Kfouri-Neto M (eds) Debates contemporâneos em direito médico e da saúde. Thomson Reuters Brazil, São Paulo, pp 69–91

Nogaroli R, Silva RG (2021) Inteligência artificial na análise diagnóstica da COVID-19: possíveis repercussões sobre a responsabilidade civil do médico. In: Rosenvald N, do Rêgo MFCE, Densa R (eds) Coronavírus e responsabilidade civil: impactos contratuais e extracontratuais. Foco, Indaiatuba, pp 293–300

Obermeyer Z, Powers B, Vogeli C, Mullainathan S (2019) Dissecting racial bias in an algorithm used to manage the health of populations. Science 366:447–453

Paranjape K, Schinkel M, Nannan Panday R, Car J, Nanayakkara P (2019) Introducing artificial intelligence training in medical education. JMIR Med Educ 5:e16048

Pasquale F (2015) The black box society: the secret algorithms that control money and information. Harvard University Press, Cambridge

Pasquale F (2017) Toward a fourth law of robotics: preserving attribution, responsibility, and explainability in an algorithmic society. Univ Md Leg Stud Res Pap 78:1–12

Pasquale F (2019) Data-informed duties in AI development. Columbia Law Rev 119:1917–1940

Pasquale F (2020) New laws of robotics: defending human expertise in the age of AI. The Belknap Press of Harvard University Press, Cambridge

Pereira AGD (2004) O consentimento informado na relação médico-paciente. Coimbra Editora, Coimbra

Price WN, Gerke S, Cohen IG (2019) Potential liability for physicians using artificial intelligence. JAMA 322:1765–1766

Robbins R (2021) An experiment in end-of-life care: tapping AI's cold calculus to nudge the most human of conversations. https://www.statnews.com/2020/07/01/end-of-life-artificial-intelligence/. Accessed 20 Oct 2021

Robbins R, Brodwin E (2021) An invisible hand: patients aren't being told about the AI systems advising their care. https://www.statnews.com/2020/07/15/artificial-intelligence-patient-consent-hospitals/. Accessed 10 Oct 2021

Ross C (2021a) As the FDA clears a flood of AI tools, missing data raise troubling questions on safety and fairness. https://www.statnews.com/2021/02/03/fda-clearances-artificial-intelligence-data/. Accessed 14 Oct 2021

Ross C (2021b) Machine learning is booming in medicine. It's also facing a credibility crisis. https://www.statnews.com/2021/06/02/machine-learning-ai-methodology-research-flaws/?utm_source=STATNewsletters&utm_campaign=37c8993853-MR_COPY_02&utm_medium=email&utm_term=0_8cab1d7961-37c8993853-153220734#. Accessed 10 Sept 2021

Russell S, Norvig P (2016) Artificial intelligence: a modern approach. Pearson, Boston

Scherer M (2016) Regulating artificial intelligence systems: risks, challenges, competencies, and strategies. Harv J Law Technol 29:353–400

Time Magazine (2019) Cardiologist eric topol on how AI can bring humanity back to medicine. March 25, 2019. https://time.com/collection/life-reinvented/5551296/cardiologist-eric-topol-artificial-intelligence-interview/. Accessed 10 Oct 2021

Tomasevicius FE, Ferraro AV (2020) Le nuove sfide dell'umanità e del diritto nell'era dell'intelligenza artificiale. Rev Direitos Cult 15:401–413

Topol EJ (2013) The creative destruction of medicine: how the digital revolution will create better health care. Basic Books, New York

Topol EJ (2016) The patient will see you now: the future of medicine is in your hands. Basic Books, New York

Topol EJ (2019a) Deep medicine: how artificial intelligence can make healthcare human again. Basic Books, New York

Topol EJ (2019b) High-performance medicine: the convergence of human and artificial intelligence. Nat Med 25:44–56

UCI Beall Applied Innovation (2017) Husky or wolf? Using a black box learning model to avoid adoption errors. http://innovation.uci.edu/2017/08/husky-or-wolf-using-a-black-box-learning-model-to-avoid-adoption-errors/. Accessed 8 Oct 2021

United Nations (2021) Urgent action needed over artificial intelligence risks to human rights. https://news.un.org/en/story/2021/09/1099972. Accessed 20 Oct 2021

Wischmeyer T (2020) Artificial intelligence and transparency: opening the black box. In: Wischmeyer T, Rademacher T (eds) Regulating artificial intelligence. Springer, Cham, pp 75–101

Wolters PTJ (2019) The obligation to update insecure software in the light of Consumentenbond/Samsung. Comput Law Secur Rev 35:295–305

World Health Organization (2021) WHO guidance – ethics and governance of artificial intelligence for health. https://www.who.int/publications/i/item/9789240029200. Accessed 20 Sep 2021

Wu E, Wu K, Daneshjou R, Ouyang D, Ho DE, Zou J (2021) How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. Nat Med 27:582–584