Artificial intelligence: Who is responsible for the diagnosis?

Emanuele Neri¹ · Francesca Coppola² · Vittorio Miele³ · Corrado Bibbolino⁴ · Roberto Grassi⁵

Received: 13 December 2019 / Accepted: 16 January 2020 / Published online: 31 January 2020
© Italian Society of Medical Radiology 2020

Abstract

The aim of the paper is to find an answer to the question “Who or what is responsible for the benefits and harms of using artificial intelligence in radiology?” When human beings make decisions, the action itself is normally connected with a direct responsibility by the agent who generated the action. You have an effect on others, and therefore, you are responsible for what you do and what you decide to do. But if you do not do this yourself, but an artificial intelligence system, it becomes difficult and important to be able to ascribe responsibility when something goes wrong. The manuscript addresses the following statements: (1) using AI, the radiologist is responsible for the diagnosis; (2) radiologists must be trained on the use of AI since they are responsible for the actions of machines; (3) radiologists involved in R&D have the responsibility to guide the respect of rules for a trustworthy AI; (4) radiologist responsibility is at risk of validating the unknown (black box); (5) radiologist decision may be biased by the AI automation; (6) risk of a paradox: increasing AI tools to compensate the lack of radiologists; (7) need of informed consent and quality measures. Future legislation must outline the contours of the professional’s responsibility, with respect to the provision of the service performed autonomously by AI, balancing the professional’s ability to influence and therefore correct the machine, limiting the sphere of autonomy that instead technological evolution would like to recognize to robots.

Keywords Artificial Intelligence · Robotics · Ethics · Radiology

Introduction

We are living in the automation society that automates more tasks and automates to a larger extent than before [1].

The boost to automation originates from the introduction of artificial intelligence in many human tasks. A typical example is represented by the so-called home automation provided by systems like Alexa, Cortana, Google Assistant, Apple Home, etc. [2]. Such systems allow to automate and

repeat human tasks as switching on and off the house light, scheduling a music playlist, switching on and off the heating in our house, etc.: all processes that can be programmed and that intelligent systems can learn to perform independently after performing them during a learning phase, which could be defined the “ground truth.” At the base of these automatic actions, there are machine/deep learning algorithms that enter into our daily life and, from our actions, learn to suggest common and new behaviors in our living.

Automation is a benefit until these automated actions are carried out in harmony with what we want, but when the automated actions go further, suggesting or carrying out unwanted actions or causing damage to things and people, questions of responsibility arise which have been unexplored up until now.

The aim of this editorial is to find an answer to the question “who or what is responsible for the benefits and harms of using this technology?”

✉ Emanuele Neri
emanuele.neri@med.unipi.it

¹ Diagnostic Radiology 3, Department of Translational Research, University of Pisa, Pisa, Italy

² Radiology Unit, Department of Diagnostic and Preventive Medicine, Sant’ Orsola Malpighi University Hospital, Bologna, Italy

³ Department of Emergency Radiology, University Hospital Careggi, Florence, Italy

⁴ SNR Foundation, Rome, Italy

⁵ Department of Radiology, University of Campania “Luigi Vanvitelli”, Naples, Italy

Who or what is the agent of responsibility?

When human beings make decisions, the action itself is normally connected with a direct responsibility by the agent who generated the action. You have an effect on others, and therefore, you are responsible for what you do and what you decide to do.

But if you do not do this yourself, but an artificial intelligence system, it becomes difficult and important to be able to ascribe responsibility when something goes wrong.

A typical example is the case of a self-driving car or an airplane using AI [3]; it should be asked: If the automation system of the car or the airplane autopilot causes an accident, who is responsible?

The Greek philosopher and polymath Aristotle gives an answer to this problem [4]. Since Aristotle, there are at least two traditional conditions for attributing responsibility for an action, the so-called control condition and the epistemic condition. So, in the control condition, you are responsible if you do it (or have done it), if you are the agent of the action, if you have caused the action, if you have a sufficient degree of control over the action, but to attribute a full responsibility, an epistemic condition, as “you know what you are doing” or “you are aware what you are doing (or knew what you were doing),” is required.

Aristotle argued in the *Nicomachean Ethics* that the action must have its origin in the agent and that one must not be ignorant of what one is doing [5].

AI technologies do not meet traditional criteria for full moral action (and hence preconditions for responsibility) such as freedom and consciousness, and therefore, they also cannot be (held) responsible [6].

With regard to the two Aristotelian conditions, it is thus assumed that it does not make sense to demand that the AI agent act voluntarily or without ignorance, since an AI agent lacks the preconditions for this: An AI cannot really act “freely” (as in “having free will”) or “know” (as in “being aware of”) what it is doing.

If this assumption holds, then our only option is to make humans responsible for what the AI technology does.

Key point 1 Using AI the radiologist is responsible for the diagnosis.

A help from the laws

On February 16th, 2017, the European Parliament approved a resolution with recommendations to the Commission on Civil Law Rules on Robotics [7, 8]. It is a resolution that somehow anticipates the current situation

that involves the professional responsibility of the radiologist. In fact, the resolution refers to the role of robotics in surgery, but the similarities of the relationship between robots and the surgeon with the relationship between artificial intelligence and radiologist are so strong that they are superimposable.

The resolution was based on a study requested by the European Parliament’s Committee on Legal Affairs and commissioned, supervised and published by the Policy Department for “Citizens’ Rights and Constitutional Affairs.” The study highlights the importance of a resolution for the immediate creation of a legislative instrument governing robotics and artificial intelligence, to anticipate any scientific developments foreseeable over the medium term and which could be adapted to track progress. We can therefore state that the resolution correctly predicted the current situation.

The study furthermore explored the liability for damages caused by an autonomous robot and stated that the expression “robots’ liability” should be banned, since it implies that the robot might itself incur civil liability for any damage caused. Instead the concept of “vicarious liability for the robot(s)” was proposed.

In the paragraph dedicated to medical robots, the resolution of the European parliament underlines: the importance of appropriate education, training and preparation for health professionals, such as doctors and care assistants; the need to define the minimum professional requirements to use a robot; the respect of the principle of the supervised autonomy of robots; the need for training for users to allow them to familiarize themselves with the technological requirements in this field; the risk of self-diagnosis of patients using mobile robot; and, consequently, the need for doctors to be trained in dealing with self-diagnosed cases. The use of such technologies should not diminish or harm the doctor–patient relationship.

Key point 2 Radiologists must be trained on the use of AI since they are responsible for the actions of machines.

The need for a trustworthy AI

In order to use artificial intelligence safely, as a support to the activity of the radiologist, it is necessary that it be trustworthy and validated in clinical practice [9].

In 2018, the European Commission established the High-Level Experts Group on Artificial Intelligence with the general objective to support the implementation of the European Strategy on Artificial Intelligence, including the elaboration of recommendations on future-related policy development and on ethical, legal and societal issues related to AI, including socio-economic challenges.

Based on fundamental rights and ethical principles, the guidelines list seven key requirements that AI systems should meet in order to be trustworthy: (1) the human agency and oversight, ensuring that an AI system does not undermine human autonomy; (2) the technical robustness and safety; (3) privacy and data governance; (4) transparency; (5) diversity, non-discrimination and fairness; (6) societal and environmental well-being; and (7) accountability [10].

The group recommended that the development, deployment and use of AI systems should adhere to the ethical principles of *respect for human autonomy, prevention of harm, fairness/equity and explicability*.

In respect to the key requirement, a trustworthy AI in radiology should not influence the autonomous decision-making process of the radiologist in image interpretation and patient's workflow (image protocol, exam prioritization, etc.); should not alter the imaging data, as in case of pre-processing from raw data to DICOM images, as well as in case of post-processing where the data could be presented in a non-interpretable format (i.e., radiomics features) [11]; and should be technically robust and safe, i.e., the resilience of software against security attacks and the reproducibility and reliability of software results in case of AI-assisted diagnosis. Important key factors of a trustworthy AI for radiology are also the data governance, the privacy and the transparency, for which an AI system should guarantee at the same time quality, integrity and confidentiality of the data processed and be explicable about the processes/algorithms used to process patient's data.

The theme of explicability is very important as an AI system is considered a kind of “black box,” of which perhaps mathematical and logical processes are understood, but not clearly the transformations of the data contained in it after various passages within the neural network (i.e., the convolutional networks).

Radiologists are familiar with digital imaging and informatics, and those with an imaging informatics special interest are frequently involved in AI algorithms development and clinical validation. Therefore, they have the potential to look into the “black box” and guide the research and development process, ensuring the respect of rules.

Key point 3 Radiologists involved in R&D have the responsibility to guide the respect of rules for a trustworthy AI.

Professional risks for the radiologist

Use of the “Black box” in clinical practice

In radiology, the clinical use of computer-aided diagnosis (CAD) is well known; multiple studies have shown advantages, limitations and risks of image interpretation

in a CAD paradigm, as first reader, second reader, or in concurrent reading [12–15]. In all cases, the final responsibility is in the hands of the radiologist, and a debate is still open about including CAD results into the radiology report and informing patients that the diagnosis is supported by automated software.

However, a clear distinction between CAD and AI must be traced, since CAD is designed to perform specific tasks on the basis of a training set and AI power and promise is that useful features can exist that are not currently known or are beyond the limit of human detection. In practice, AI brings new features from images that radiologists, as human beings, cannot detect or quantify. A typical example is radiomics, where a texture analysis can generate hundreds of features that a human being cannot generate and interpret [16].

In clinical practice, radiologists will be asked to monitor AI system outputs and validate AI interpretations; so *they risk to carry the ultimate responsibility of validating what they cannot understand*.

Key point 4 Radiologist responsibility is at risk of validating the unknown (black box).

Automation bias

The recent North America and European multi-society paper about “Ethics of artificial intelligence in radiology” reports the risk of the automation bias [17].

Automation bias is the tendency for humans to favor machine-generated decisions, ignoring contrary data or conflicting human decisions. Automation bias leads to errors of omission and commission, where omission errors occur when a human fails to notice, or disregards, the failure of the AI tool. High decision flow rates, where decisions are swiftly made on radiology examinations, and the radiologist is reading examinations rapidly, predispose to omission errors. This is compounded by AI decisions made based on features that are too subtle for humans to detect.

Commission errors occur when the radiologist erroneously accepts or implements a machine's decision in spite of other evidence to the contrary.

The typical feature of the act performed by the radiologist, as a professional, is the autonomy of decisions on the provision of service and the technical tools to be used, and the personality of the service [18]. It is clear that these peculiarities must be harmonized with the automation of AI.

Key point 5 Radiologist decision may be biased by the AI automation.

Shortage of radiologists and inappropriate training

Much discussion has been raised in the media about the introduction of artificial intelligence in radiological practice, suggesting that radiologists could become useless or even disappear as a specialty [19–21]. This could lead to a lack of motivation for young doctors to pursue a career in radiology, with a real imminent risk of not having enough radiologist specialists.

An additional risk for young doctors in training is the reduction in training opportunities. In fact, if the use of artificial intelligence systems in clinical practice can accelerate interpretation times, it can also reduce the number of examinations/images that a radiologist in training will be able to interpret. In other words, the greater the automatic tasks performed by artificial intelligence, the less those performed by the radiologist.

The paradox of this situation would be the need to implement ever greater AI tasks to compensate for the progressive lack of radiologists.

Key point 6 Risk of a paradox: increasing AI tools to compensate the lack of radiologists.

Radiologist–patient relationship and data integrity

A contingent problem with the introduction of AI and of no less importance is transparency toward patients. They must be informed that the diagnosis was obtained with the help of the AI.

All this poses problems to the patient's informed consent and the adoption of quality measures of the patient's care providers (public or private institutions) on AI systems.

Quality measures should relate to software robustness and data security, as well as constant updating of software and hardware, avoiding equipment obsolescence. Particular attention should also be paid to image processing, guaranteeing its integrity during the analysis process with neural networks, thus avoiding a modification of the raw data.

Key point 7 Need for informed consent and quality measures.

Conclusions

Artificial intelligence is entering the radiological discipline very quickly and will soon be in clinical use. The European Society of Radiology stated that the most likely and immediate impact of AI will be on the management of radiology workflows, improving and automating acquisition protocols, appropriateness (with clinical decision support systems), structured reporting, up to the ability to interpret the big

data of image biobanks connected to tissue biobanks, with the development of radiogenomics [22].

But the fundamental problem of an ethical and not harmful AI still remains.

Are there solutions? In a recent article published by Thomas PS et al., the application of a defined Seldon algorithm is described. The authors propose a framework for designing machine learning algorithms and show how it can be used to construct algorithms that provide their users with the ability to easily place limits on the probability that the algorithm will produce any specified undesirable behavior [23].

Perhaps the solution is to create an ethical AI, subject to a constant action control, as indeed happens for the human conscience: an AI subjected to a vicarious civil liability, written in the software and for which the producers must guarantee the users, so that they can use AI reasonably and with a human-controlled automation.

It is clear that the future legislation must outline the contours of the professional's responsibility, with respect to the provision of the service performed autonomously by AI, balancing the professional's ability to influence and therefore correct the machine, limiting the sphere of autonomy that instead technological evolution would like to recognize to robots.

Compliance with Ethical Standards

Conflicts of interest The authors declare that they have no conflict of interest.

Research involving human participants and/or animals No human participants and/or animals.

Informed consent Not applicable.

References

1. Coeckelbergh M (2019) Artificial intelligence, responsibility attribution, and a relational justification of explainability. *Sci Eng Ethics*. <https://doi.org/10.1007/s11948-019-00146-8>
2. Hoy MB (2018) Alexa, Siri, Cortana, and More: an introduction to voice assistants. *Med Ref Serv Q* 37(1):81–88. <https://doi.org/10.1080/02763869.2018.1404391>
3. Kallioinen N, Pershina M, Zeiser J, Nosrat Nezami F, Pipa G, Stephan A, König P (2019) Moral judgements on the actions of self-driving cars and human drivers in dilemma situations from different perspectives. *Front Psychol* 1(10):2415. <https://doi.org/10.3389/fpsyg.2019.02415>
4. Prus R (2015) Aristotle's theory of deviance and contemporary symbolic interactionist scholarship: learning from the past, extending the present, and engaging the future. *Am Sociol* 46(1):122–167. <https://doi.org/10.1007/s12108-014-9250-9>
5. Price AW (1985) Aristotle's ethics. *J Med Ethics* 11(3):150–152. <https://doi.org/10.1136/jme.11.3.150>

6. Giubilini A, Savulescu J (2018) The artificial moral advisor The “Ideal Observer” meets artificial intelligence. *Philos Technol* 31(2):169–188. <https://doi.org/10.1007/s13347-017-0285-z>
7. [https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU\(2016\)571379_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf)
8. Winfield AFT, Jirotko M (2018) Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philos Trans A Math Phys Eng Sci.* 376(2133):20180085. <https://doi.org/10.1098/rsta.2018.0085>
9. Ho CWL, Soon D, Caals K, Kapur J (2019) Governance of automated image analysis and artificial intelligence analytics in healthcare. *Clin Radiol* 74(5):329–337. <https://doi.org/10.1016/j.crad.2019.02.005>
10. <https://ec.europa.eu/futurium/en/ai-allianceconsultation/guidelines/1#Human%20agency>
11. Pesapane F, Volonté C, Codari M, Sardanelli F (2018) Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. *Insights Imaging* 9(5):745–753. <https://doi.org/10.1007/s13244-018-0645-y>
12. Doi K (2007) Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput Med Imaging Graph* 31(4–5):198–211
13. Regge D, Halligan S (2013) CAD: how it works, how to use it, performance. *Eur J Radiol* 82(8):1171–1176
14. Al Mohammad B, Brennan PC, Mello-Thoms C (2017) A review of lung cancer screening and the role of computer-aided detection. *Clin Radiol* 72(6):433–442
15. Katzen J, Dodelzon K (2018) A review of computer aided detection in mammography. *Clin Imaging* 52:305–309
16. Kohli M, Prevedello LM, Filice RW, Geis JR (2017) Implementing machine learning in radiology practice and research. *AJR Am J Roentgenol* 208(4):754–760
17. Geis JR, Brady AP, Wu CC et al (2019) Ethics of artificial intelligence in radiology: summary of the joint European and North American multisociety statement. *Can Assoc Radiol J* 70(4):329–334. <https://doi.org/10.1016/j.carj.2019.08.010>
18. Salloch S (2017) Who’s afraid of EBM? Medical professionalism from the perspective of evidence-based medicine. *Med Health Care Philos* 20(1):61–66. <https://doi.org/10.1007/s11019-016-9726-1>
19. Neri E, de Souza N, Brady A, Bayarri AA, Becker CD, Coppola F, Visser J (2019) What the radiologist should know about artificial intelligence—an ESR white paper. *Insights Imaging.* <https://doi.org/10.1186/s13244-019-0738-2>
20. Morgenstern M (2017) Automation and anxiety. *The Economist* <https://www.economist.com/special-report/2016/06/25/automation-and-anxiety>. Accessed 7 Aug 2017
21. Mukherjee S (2017) A.I Versus M.D. *New Yorker.* <http://www.newyorker.com/magazine/2017/04/03/ai-versus-md>. Accessed 7 Aug 2017
22. Neri E, de Souza N, Brady A, Bayarri AA, Becker CD, Coppola F, Visser J (2019) What the radiologist should know about artificial intelligence - an ESR white paper. *Insights Imaging.* <https://doi.org/10.1186/s13244-019-0738-2>
23. Thomas PS, CastrodaSilva B, Andrew G, Barto AG, Giguere S, Brun Y, Brunskill E (2019) Preventing undesirable behavior of intelligent machines. *Science* 366(6468):999–1004. <https://doi.org/10.1126/science.aag3311>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.