Received: 13 November 2020 | Revised: 31 March 2021 | Accepted: 20 May 2021

DOI: 10.1111/bioe.12918

SPECIAL ISSUE: PROMISES AND CHALLENGES OF MEDICAL AI

bioethics WILEY

Explicability of artificial intelligence in radiology: Is a fifth bioethical principle conceptually necessary?

Frank Ursin | Cristian Timmermann | Florian Steger

Institute of the History, Philosophy and Ethics of Medicine, Ulm University, Ulm, Germany

Correspondence
Frank Ursin, Institute of the History, Philosophy and Ethics of Medicine, Ulm University, Parkstraße 11, 89073 Ulm, Germany.
Email: frank.ursin@uni-ulm.de

Abstract

Recent years have witnessed intensive efforts to specify which requirements ethical artificial intelligence (AI) must meet. General guidelines for ethical AI consider a varying number of principles important. A frequent novel element in these guidelines, that we have bundled together under the term explicability, aims to reduce the black-box character of machine learning algorithms. The centrality of this element invites reflection on the conceptual relation between explicability and the four bioethical principles. This is important because the application of general ethical frameworks to clinical decision-making entails conceptual questions: Is explicability a free-standing principle? Is it already covered by the well-established four bioethical principles? Or is it an independent value that needs to be recognized as such in medical practice? We discuss these questions in a conceptual-ethical analysis, which builds upon the findings of an empirical document analysis. On the example of the medical specialty of radiology, we analyze the position of radiological associations on the ethical use of medical AI. We address three questions: Are there references to explicability or a similar concept? What are the reasons for such inclusion? Which ethical concepts are referred to?

KEYWORDS

black box, explainability, machine learning, medical ethics, principlism, transparency

1 | INTRODUCTION

Efforts to specify the ethical issues of artificial intelligence (AI) predominantly rely on a principled approach. A recent and comprehensive review of guidelines for ethical AI found 11 overarching ethical principles, each summarizing further principles, 63 in total.[1] The most frequent issue was communicating to patients how results were achieved, identified in 73 out of 84 guidelines (87%). To support this demand, a series of terms are used: transparency, understandability, comprehensibility, intelligibility, demonstrability, explainability, and interpretability. These terms have been bundled under the umbrella concept of explicability.[2]

The high frequency with which these terms are stated in general guidelines invites us to explore how they have been translated into practice, taking the specialty of radiology as an example. We will first provide an empirical overview on white papers issued by

[1]Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399.

[2]Robbins, S. (2019). A misdirected principle with a catch: Explicability for AI. *Minds and Machines*, *29*, 495–514.

radiological associations by addressing three questions: (a) Are there references to explicability or a similar idea? (b) What are the reasons for such inclusion? (c) Which ethical principles are referred to? In the Discussion, we conduct a conceptual-ethical analysis of the relation between explicability and the four ethical principles by addressing the following question: Is it conceptually necessary to treat explicability as a principle for assessing medical AI? This is important because it has recently been stated that a five-principle framework that includes explicability is suitable for assessing ethical AI in general.[3] Interestingly, this approach has been directly applied to radiology, where explicability may accompany the four traditional principles of biomedical ethics.[4]

In the conceptual-ethical analysis, we proceeded step-wise addressing three questions: First, what are the advantages of recognizing explicability as a fifth principle? Second, is explicability already covered by any of the four traditional bioethical principles autonomy, beneficence, non-maleficence, and justice?[5] Lastly, does explicability have a value in itself? Before analyzing how the advent of medical AI in radiology relates to the four bioethical principles we need to specify the key idea that is on trial. A crucial element of medical examinations is to be able to communicate findings to the patient at a sufficient level to facilitate informed consent about future procedures. Informed consent to a medical diagnosis or treatment contains five aspects: (a) information disclosure to safeguard autonomous decisions, (b) the patient's capacity to understand the information, (c) voluntariness of the decision, (d) the competence to make decisions, and (e) the decision itself.[6] At first sight, two agents are involved in this communication and two different types of explanations are required. Physicians demand explicability as domain experts who need to assume the responsibilities of avoiding harm and informing patients. Patients may demand explicability as autonomous agents that want to decide over a therapeutic process or simply inquire about their condition and its assessment. The use of complex tools to assist diagnosis demands a degree of openness on how doctors reach a certain conclusion. Therefore, technology developers come into play as a third agent when physicians require them to explain how an AI system arrived at a diagnosis.

To encourage this openness reference is made to the concepts we have grouped under "explicability." Although these concepts all aim at improving communication, they have different ethical implications. For instance, transparency usually appeals to not withholding information. At a minimum level, transparency involves mostly negative duties. In contrast, explainability and demonstrability not only ask health professionals to refrain from hiding information but also demand that information is made understandable to patients. In other words, these two concepts involve the positive duty to deliver the information and request a substantial effort to make sure patients understand how this information came about and what it implies. In the following we refer to this latter, more demanding task in our use of the principle of explicability.

In the literature, there is a split opinion on whether to include such an additional principle. On the one hand, explicability is considered crucial for ethical AI on technological grounds.[7] The moral responsibility of clinicians to provide reasons or a rationale for decisions in individual cases has traditionally been emphasized in this regard.[8] Fostering trust in the results by being able to understand how they were achieved is considered important to increase the acceptance of diagnosis.[9] On the other hand, commentators state that the ability to produce accurate results for decisions in medicine is more important than the ability to explain how results are produced.[10] The technological background is that there is a tradeoff between accuracy and explicability: the more explicable an AI system is, the less accurately it performs.[11] Here, the need to offer medical accuracy is seen as the prevailing value.

We contribute to the discussion by analyzing how the inclusion of explicability is justified in the medical domain by taking radiology as an example. Within the wide field of AI, machine learning (ML) is most often used in radiology due to its capabilities to autonomously label imaging data after training.[12] ML can use supervised learning to generate an output within predefined classifications using an algorithm whose parameters are computed during a training phase.[13] The most often used type of ML in radiology is deep learning (DL) that uses multiple hidden layers between the input and output

[3]Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*.https://doi.org/10.1162/99608f92.8cd550d1; Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People - An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*, 689–707; European Commission. (2020). *White Paper on Artificial Intelligence - A European approach to excellence and trust*. European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0065&qid=1605015492763&from=EN (accessed March 1, 2021); European Commission. (2019). *Ethics guidelines for trustworthy AI*. European Commission. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60651 (accessed October 15, 2020).

[4]Akinci D'Antonoli, T., Weikert, T. J., Sauter, A. W., Sommer, G., & Stieltjes, B. (2019). *Ethical considerations for artificial intelligence implementation in radiology*. https://epos.myesr.org/poster/esr/ecr2019/C-2553 (accessed March, 1 2021); Akinci D'Antonoli, T. (2020). Ethical considerations for artificial intelligence: An overview of the current radiology landscape. *Diagnostic and Interventional Radiology (Ankara, Turkey), 26*, 504–511.

[5]Beauchamp, T. L., & Childress, J. F. (2019). *Principles of biomedical ethics*. Oxford University Press.

[6]Faden, R. R., Beauchamp, T. L., & King, N. M. P. (1986). *A history and theory of informed consent*. Oxford University Press.

[7]Floridi & Cowls, op. cit. note 3.

[8]Swartout, W. R. (1983). XPLAIN: A system for creating and explaining expert consulting programs. *Artificial Intelligence, 21*(3), 285–325.

[9]Holzinger, A., Carrington, A., & Müller, H. (2020). Measuring the quality of explanations: The system causability scale (SCS): Comparing human and machine explanations. *Künstliche Intelligenz, 34*, 193–198.

[10]London, A. J. (2019). Artificial intelligence and black-box medical decisions: Accuracy versus explainability. *The Hastings Center Report, 49*, 15–21.

[11]Reyes, M, Meier, R., Pereira, S., Silva, C. A., Dahlweid, F. M., Tengg-Kobligk, H. V., Summers, R. M., & Wiest, R. (2020). On the interpretability of artificial intelligence in radiology: Challenges and opportunities. *Radiology. Artificial Intelligence, 2*, e190043.

[12]Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. (2018). Artificial intelligence in radiology. *Nature Reviews Cancer, 18*, 500–510.

[13]SFR-IA Group, CERF & on behalf of the French Radiology Community (2018). Artificial intelligence and medical imaging 2018: French Radiology Community white paper. *Diagnostic and Interventional Imaging, 99*, 727–742, p. 728.

layers.[14] Afterwards, the accuracy of the algorithm is tested on new data and thus the algorithm's capability to generalize. Applications in radiology include medical imaging (computer-aided detection of abnormalities and its characterization, organ segmentation, etc.), workflow optimization, triaging, decision support, and improving image quality.[15]

By examining white papers on ethical AI authored by radiological associations, we are able to elaborate the following argument. First, radiological associations consider explicability important due to the new technological challenges of medical AI in imaging (e.g., its black-box character, risk of bias in the training data, overfitting and generalization problems). Second, radiological associations justify the claim for explicability with a variety of reasons: avoiding harm to patients, need to justify judgements, and building trust in AI. Third, avoiding harm is demanded by the principle of non-maleficence and informed consent is required from the principle of respecting the patient's autonomy. Fourth, the four bioethical principles are sufficient within the medical domain. We conclude that the concept of explicability merely offers an additional safety protocol for the technological specificities of medical AI.

2 | MATERIALS AND METHODS

In a document analysis, we investigated how the inclusion of explicability and similar ideas are justified in white papers and statements on medical AI authored by radiological associations. Radiology was chosen on three grounds. First, radiology is among the medical specialties where AI systems are most advanced.[16] There are already over 80 Food and Drug Administration (FDA) cleared AI algorithms in radiology that are commercially available and used in clinical practice.[17] Second, most of the medical AI research output comes from this field.[18] Third, nearly three quarters of medical AI research using DL deals with diagnostic imaging.[19] Within the field it is considered urgent to develop new publication venues and reporting guidelines that are specific for AI and ML research since the topic represented 25% of published research in one eminent journal in 2018.[20] Radiologists perceive themselves as being on the forefront of the digital era in medicine and may "now guide the introduction of AI in healthcare."[21]

White papers authored by radiological associations were searched in PubMed with the formula (artificial intelligence) AND (radiology) AND ((white paper) OR statement) on October 8, 2020 (n = 97; Figure 1). After removing duplicates and checking for eligibility, seven white papers or statements were included in the document analysis. The document analysis aimed at answering three questions: (a) Are there references to explicability or a similar idea? (b) What are the reasons for including explicability or a similar idea? (c) Which ethical principles are referred to?

To answer question (a), a close reading of the full text of the included articles was conducted. To answer question (b), reasons for including explicability were extracted by a process similar to a systematic review of reasons.[22] Recurrent reasons were pooled only minimally, because the small dataset did not show much variance. The results are charted in Table 1. To answer question (c), we identified which ethical guidelines or superordinate policies the authors of white papers refer to. Superordinate policies are guidelines for ethical AI, reporting guidelines for research, or legal regulations like the European Union's (EU) General Data Protection Regulation (GDPR).[23]

As a second step, to analyze the results in detail and answer the conceptual-ethical question of whether the inclusion of explicability is conceptually necessary, we surveyed how explicability is conceived in selected guidelines for ethical AI. The inclusion criteria were that the guidelines include explicability or a similar principle and also apply the four bioethical principles. We built on the findings of three recent reviews of general guidelines for ethical AI. These reviews were found by manual searches and they allow us to gain an overview of the involved principles: Jobin et al. (2019) conducted a scoping review of 84 guidelines for ethical AI.[24] Their content analysis revealed 11 overarching principles of which transparency was the most common. Floridi and Cowls (2019) synthesized six guidelines authored by high-profile initiatives with 49 principles into a five-principle approach for ethical AI.[25] Hagendorff (2020) analyzed 22 guidelines with 18 principles to highlight overlaps and omissions.[26]

[14]Geis, J. R., Brady, A., Wu, C. C., Spencer, J., Kohli, M., Ranschaert, E., Jaremko, J. L., Langer, S. G., Borondy Kitts, A., Birch, J., Shields, W. F., van den Hoven van Genderen, R., Kotter, E., Wawira Gichoya, J., Cook, T. S., Morgan, M. B., Tang, A., & Safdar, N. M. (2019). *Ethics of AI in radiology. European and North American Multisociety Statement*. https://www.acr.org/-/media/ACR/Files/Informatics/Ethics-of-AI-in-Radiology-European-and-North-American-Multisociety-Statement--6-13-2019.pdf (accessed March 1, 2021), p. 43.

[15]SFR-IA Group et al., op. cit. note 13; Alexander, A., Jiang, A., Ferreira, C., & Zurkiya, D. (2020). An intelligent future for medical imaging: A market outlook on artificial intelligence for medical imaging. *Journal of the American College of Radiology, 17*, 165–170; European Society of Radiology (ESR) (2019). What the radiologist should know about artificial intelligence - an ESR white paper. *Insights into Imaging, 10*, 44.

[16]Crawford, K., Roel, D., Dryer, T., Fried, G., Green, B., Kazunias, E., Kak, A., Mathur, V., McElroy, E., Nill Sánchez, A., Raji, D., Lisi Rankin, J., Richardson, R., Schultz, J., Myers West, S., & Whittaker, M. (2019). *AI Now 2019 Report*. AI Now Institute. https://ainowinstitute.org/AI_Now_2019_Report.pdf (accessed March 1, 2021), p. 54.

[17]American College of Radiology. (2020). *FDA cleared AI algorithms*. https://www.acrdsi.org/DSI-Services/FDA-Cleared-AI-Algorithms (accessed March 1, 2021).

[18]European Society of Radiology (ESR), op. cit. note 15.

[19]Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology, 2*, 230–243.

[20]Bluemke, D. A., Moy, L., Bredella, M. A., Ertl-Wagner, B. B., Fowler, K. J., Goh, V. J., Halpern, E. F., Hess, C. P., Schiebler, M. L. & Weiss, C. R. (2020). Assessing radiology research on artificial intelligence: A brief guide for authors, reviewers, and readers-from the Radiology Editorial Board. *Radiology, 294*, 487–489.

[21]Pesapane, F., Codari, M., & Sardanelli, F. (2018). Artificial intelligence in medical imaging: Threat or opportunity? Radiologists again at the forefront of innovation in medicine. *European Radiology Experimental, 2*, 35.

[22]Strech, D., & Sofaer, N. (2012). How to write a systematic review of reasons. *Journal of Medical Ethics, 38*, 121–126.

[23]European Parliament & European Council. (2016). *General Data Protection Regulation*. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN (accessed March 1, 2021).

[24]Jobin et al., op. cit. note 1.

[25]Floridi & Cowls, op. cit. note 3.

[26]Hagendorff, T. (2020). Publisher correction to: The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines, 30*,457–461.

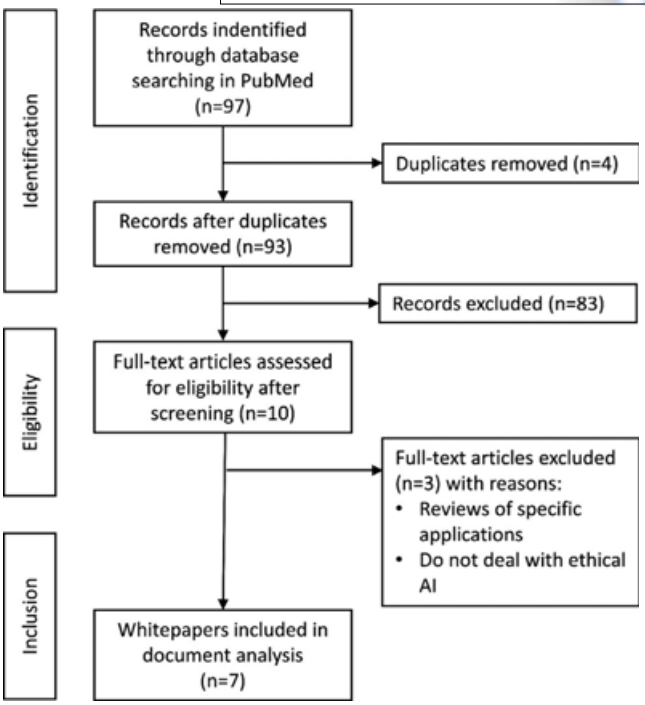

FIGURE 1 Flowchart of search strategy following the PRISMA (2009) guideline (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)

3 | RESULTS

The seven white papers we analyzed were issued between 2018 and 2020. Explicability is a recurrent idea therein, but the meaning and scope differ significantly. Some white papers address transparency as the disclosure of competing interests in developing medical AI, others as obtaining informed consent from patients regarding their data for training purposes, or as the technological particularities of "black-box algorithms," i.e., their opacity. In what follows we will treat each of the seven documents separately beginning with national, to international and multisociety statements. This is a narrative synthesis of our empirical findings.

1. An AI working group within the Canadian Association of Radiologists has elaborated recommendations for the introduction and implementation of AI in imaging.[27] The working group had the mandate to discuss and deliberate on practice, policy, and patient care issues. In terms of education, the working group recommends training radiologists in the understanding of medical AI. Because algorithms are prone to bias, the authors warn of significant ethical issues. These ethical issues are not discussed in detail except data privacy and misdiagnosing that can harm patients.
2. The same AI working group issued another white paper specifically on the ethical and legal issues related to AI in radiology.[28] They provide a framework for ethical and legal issues like patient data (privacy, confidentiality, ownership, and sharing), algorithms (levels of autonomy, liability, and jurisprudence), and practice (best practices and current legal framework). The transparent and safe design of algorithms in order to minimize their black-box character is discussed in the context of James H. Moor's approach of just consequentialism. According to Moor, policy should be guided by balancing the principle of justice with the expected positive consequences (reducing harm, increasing happiness).[29] In its consequentialist emphasis for the common good this approach goes beyond basic moral virtues like Google's former motto "Don't be evil."
3. An AI group of the French Radiology Society and the French College of Radiology Teachers have issued a white paper about AI in radiology on behalf of the French radiology community.[30] The authors explicitly promote that algorithms should comply with ethical principles and have therefore translated the 23 Asilomar principles to the needs of radiologists. Important for the topic of explicability are two principles: "Failure transparency: If an AI system causes harm, it should be possible to ascertain why" (Principle no. 7) and "Judicial transparency: Any involvement by an autonomous system in judicial/medical decision-making should provide a satisfactory explanation auditable by a competent human authority" (Principle no. 8).[31] A central argument concerns the justification of results provided by AI. The authors state that diagnosis must be justified in healthcare. If the result of an algorithm cannot be explained by the physician to the patient, the algorithm must not be used, not even as a second opinion. Due to the black-box character of neural networks, intelligibility and demonstrability raise ethical challenges.[32] Transparency and comprehensibility may help to overcome these challenges. The authors suggest that AI systems "be founded on the principle of justification, based on possibilities instead of probabilities, in order to maintain some level of demonstrability of the results."[33] By demanding transparent justification of judgements, the authors refer to the principle of explicability. Furthermore, they demand that a "radiologist must understand the technical basis of a tool,"[34] thereby pointing to education and structured training of radiologists. While explicability is often discussed in this statement, its centrality is down-played because the authors state that radiologists remain responsible for any decision. It is argued that neither machines nor

[27]Tang, A., Tam, R., Cadrin-Chênevert, A., Guest, W., Chong, J., Barfett, J., Chepelev, L., Cairns, R., Mitchell, J. R., Cicero, M. D., Gaudreau Poudrette, M., Jaremko, J. L., Reinhold, C., Gallix, B., Gray, B., Geis, R., & Canadian Association of Radiologists (CAR) Artificial Intelligence Working Group (2018). Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology. *Canadian Association of Radiologists Journal = Journal l'Association canadienne des radiologistes*, *69*, 120–135.

[28]Jaremko, J. L., Azar, M., Bromwich, R., Lum, A., Alicia Cheong, L. H., Gibert, M., Laviolette, F., Gray, B., Reinhold, C., Cicero, M., Chong, J., Shaw, J. Rybicki, F. J., Hurrell, C., Lee, E., Tang, A., & Canadian Association of Radiologists (CAR) Artificial Intelligence Working Group (2019). Canadian Association of Radiologists white paper on ethical and legal issues related to artificial intelligence in radiology. *Canadian Association of Radiologists Journal = Journal l'Association canadienne des radiologistes*, *70*, 107–118.

[29]Moor, J. H. (1999). Just consequentialism. *Ethics and Information Technology*, *1*, 61–65.

[30]SFR-IA Group et al., op. cit. note 13.

[31]Ibid: 740.

[32]Ibid: 737.

[33]Ibid.

[34]Ibid: 735.

TABLE 1 Justifications of including explicability in white papers on medical artificial intelligence (AI) authored by radiological associations

No.	Issuing association	Are there references to explicability or a similar idea?	What are the reasons for including explicability or a similar idea?	Which ethical principles are referred to?
1	Canadian Association of Radiologists (2018)	Understandability (implicit)	Radiologist must understand value, weaknesses, and potential errors of AI systems performing image analysis; opacity of AI systems	Not specified
2	Canadian Association of Radiologists (2019)	Transparency	Fear of suspicion and public backlash; disclosing financial interests for for-profit AI systems; opacity of AI systems	Not specified
3	Artificial Intelligence Group of the French Radiology Society and French College of Radiology Teachers on behalf of the French Radiology Community (2018)	Transparency; comprehensibility; intelligibility; demonstrability	Failure and judicial transparency; improving quality and safety; justification of judgements; opacity of AI systems	23 Asilomar principles (2017)
4	Royal Australian and New Zealand College of Radiologists (2019)	Transparency; explainability; interpretability	Radiologist must be capable of interpreting the basis on which a result was reached; understand and explain how a result can impact patient care; opacity of AI systems	Not specified
5	European Society of Radiology (2019)	Understandability (implicit)	Radiologist must understand how and why decisions were made; opacity of AI systems	Asimov's three laws of robotics (1942)
6	Joint European and North American Multisociety Statement (2019)	Transparency; interpretability; explainability	Patient and provider trust in AI as to how decisions are made; EU's General Data Protection Regulation demands consent to automated decision making; opacity of AI systems	Not specified
7	International Society of Radiographers, Radiological Technologists, and the European Federation of Radiographer Societies (2020)	Understandability (implicit)	Avoid errors; important for decision making; opacity of AI systems	Not specified

algorithms are autonomous moral or legal entities. Therefore, the physician is responsible for diagnostic procedures even if decision-making is automated. AI systems are considered merely as assistants that answer specific questions with no high level of autonomy. This can be observed in one of the "ten principles of AI in radiology" the authors created for the French radiology community: "AI tools should be used as a complement to the imaging study process to improve quality and safety in radiology."[35] We conclude that "improving quality and safety" as well as the Asilomar principles 7 and 8 resemble the principles of beneficence and non-maleficence.

4. The Royal Australian and New Zealand College of Radiologists authored a statement called "Ethical principles for artificial intelligence in medicine" intended to guide more stakeholders than only radiologists.[36] Beside "Principle Four: Transparency and Explainability" there are eight further principles: safety, privacy and protection of data, avoidance of bias, respecting human values, decision making on diagnosis and treatment, teamwork, responsibility for decisions, and governance. Transparency and explainability mean that a physician must be able to understand and explain a result produced by an algorithm.[37]
5. The European Society of Radiology refers to Isaac Asimov's three laws of robotics, transfering them to medical AI imaging software.[38] Accordingly, the first law indicates that "AI tools should achieve the best possible diagnosis to improve patient's healthcare."[39] Secondly, "AI must be properly trained,"[40] and a radiologist must ensure clinically meaningful outputs. Third, AI software may rapidly become outdated and obsolescent when technology evolves, and therefore be replaced by the new state of the art. Strong emphasis is put upon the black-box character of AI when discussing the question "who is responsible for the diagnosis," especially if it is incorrect.[41]
6. In the European and North American Multisociety Statement,[42] the four bioethical principles are not explicitly mentioned, but the ethical use of AI in radiology "should promote wellbeing, minimize harm, and ensure that the benefits and harms are distributed among the possible stakeholders in a just manner."[43] This statement resembles three of the four bioethical principles, i.e., beneficence, non-maleficence, and justice. Patient and provider trust must be secured by transparency, interpretability, and explainability, clarifying how decisions are made by AI systems. ML algorithms do have biases due to training data and entail biases in using them. Automation bias leads physicians to favor the suggestion made by a machine.[44] Explicability is considered to mitigate both aforementioned biases. It is referred to the EU's GDPR with its emphasis on consent to automated decision making.[45]
7. The joint statement of the International Society of Radiographers, Radiological Technologists, and the European Federation of Radiographer Societies was included because it supplements the statements of the radiological associations through another perspective on the clinical application of AI.[46] Explicability or a similar concept are not mentioned explicitly, but it is stated that the radiographer must "understand how algorithms arrive at decisions and probability errors within these decisions to enable effective communication of findings to patients."[47]

4 | DISCUSSION

4.1 | Explicability in white papers of radiological associations

The reasons why explicability was deemed important for medical AI differed significantly except for one element. All white papers justified explicability due to the technological peculiarities of AI/ML systems. These systems are black boxes that are characterized by a tradeoff between accuracy and explicability. This tradeoff results in a dilemma, i.e., one has to decide what the prevailing value is: a high degree of accuracy entails opacity while increasing explicability comes at the cost of accuracy. The radiological associations tend to value explicability, because they see AI/ML systems merely as complementary tools for the radiologist who remains responsible for medical decisions. The basic attitude common to all associations is that the responsibility towards patients demands that radiologists are able to explain how decisions are reached and to double-check whether AI processes indeed benefit the patient. We conclude that this resembles the bioethical principle of beneficence.

We found that the four bioethical principles are only implicitly referred to.[48] There was no reference to one of the recent general

[35]Ibid: 738.

[36]The Royal Australian and New Zealand College of Radiologists. (2019). *Ethical principles for artificial intelligence in medicine*. https://www.ranzcr.com/documents/4952-ethical-principles-for-ai-in-medicine/file (accessed March 1, 2021).

[37]Ibid: 5.

[38]European Society of Radiology (ESR), op. cit. note 15.

[39]Ibid.

[40]Ibid.

[41]Ibid.

[42]The societies are the American College of Radiology, the European Society of Radiology, the Radiological Society of North America, the Society for Imaging Informatics in Medicine, the European Society of Medical Imaging Informatics, the Canadian Association of Radiologists, and the American Association of Physicists in Medicine.

[43]Geis, J. R., Brady, A. P., Wu, C. C., Spencer, J., Ranschaert, E., Jaremko, J. L., Langer, S. G., Borindy Kitts, A., Shields, W. F., van den Hoven van Genderen, R., Kotter, E., Wawira Gichoya, J., Cook, T. S., Morgan, M. B., Tang, A., Safdar, N. M., & Kohli, M. (2019). Ethics of artificial intelligence in radiology: Summary of the Joint European and North American Multisociety Statement. *Radiology, 293*, 436–440, p. 437; Geis et al. (2019), op. cit. note 14, p. 10.

[44]Geis et al. (2019), op. cit. note 14, p. 35.

[45]Ibid: 30.

[46]The European Federation of Radiographer Societies. Artificial Intelligence and the Radiographer/Radiological Technologist Profession. (2020). A joint statement of the International Society of Radiographers and Radiological Technologists and the European Federation of Radiographer Societies. *Radiography (London), 26*, 93–95.

[47]Ibid: 94.

[48]European and North American Multisociety Statement, op. cit. note 14.

ethical frameworks for ethical AI that use a five-principle approach.[49] There was only reference to two ethical frameworks, i.e., the 23 Asilomar principles and Asimov's three laws for robotics.[50] These frameworks have been translated to the needs of radiologists. Important for explicability are the Asilomar principles 7 (Failure transparency: If an AI system causes harm, it should be possible to ascertain why.) and 8 (Judicial transparency: Any involvement by an autonomous system in judicial decision-making should provide a satisfactory explanation auditable by a competent human authority.). The American writer Isaac Asimov (1920–1992) formulated the three laws of robotics in the short story "Runaround":

> First Law: A robot may not injure a human being or, through inaction, allow a human being to come to harm. Second Law: A robot must obey the orders given it by human beings except where such orders would conflict with the First Law. Third Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.[51]

The Asilomar principles have in common with Asimov's laws the aim of harm reduction as does the bioethical principle of non-maleficence.

The repeated references to the harm reducing character of explicability should not lead us to think that this concept has only an instrumental character. A commitment to harm reduction has a long history in radiology due to the extensive use of radioactive materials and radiation in the first decades of the last century.[52] Yet we need to recognize that radiologists value explicability also for intrinsic reasons, such as promoting the greater good of patients, the satisfaction of scientific curiosity or learning the details about alternative processes to identify a pathogen. Therefore, not all reasons in favor of explicability are grounded on medical ethics, but also on epistemic interest within the radiological community.

We found that explicability is considered crucial to support clinical decisions. What this means in clinical practice becomes clear from a recent survey among radiologists: 56% said that they currently use some sort of AI in at least one of the following five domains: (a) triaging images to first review critical patients; (b) optimizing workflow for overall productivity; (c) partly automating image analysis; (d) providing clinicians with decision support; (e) enhancing imaging quality.[53]

In the same survey, two barriers were identified that impede radiologists from applying medical AI more broadly: first, radiologists express skepticism about its current diagnostic capabilities, and second, the lack of regulatory approval. However, the FDA in the US and the Medical Device Regulation in the EU already allow approval to be given to medical AI as a "Software as Medical Device".[54] The American College of Radiology provides a list of over 80 already cleared AI algorithms in radiology.[55] However, there are still paramount challenges concerning continual learning ML tools like catastrophic forgetting, i.e., new data interferes with what the model has already learned and decreases its performance.[56] Therefore, the FDA has so far only granted approvals for locked systems.

4.2 | Explicability in reporting guidelines

Transparency was mentioned in conjunction with two reporting guidelines.[57] This is instructive for the research domain as there have been repeated calls for AI systems to be robust and accurate. On the one side there is the Standards for Reporting Diagnostic Accuracy (STARD) statement, which aims at improving completeness and transparency of accuracy studies.[58] On the other side there is the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement, which aims at studies that report predictive models.[59] Transparency is mentioned in both checklists, but they do not aim primarily on AI systems.

The recent Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence (SPIRIT-AI) and Consolidated Standards of Reporting Trials–Artificial Intelligence (CONSORT-AI) extensions are important because they will shape how research in medical AI will be reported in the future.[60] These two reporting checklists are intended for applications like triage, diagnosis,

[49]Floridi & Cowls, op. cit. note 3.

[50]Future of Life Institute. (2017). *Asilomar AI principles*. https://futureoflife.org/ai-principles/ (accessed March 1, 2021); Moran, M. (2008). Three laws of robotics and surgery. *Journal of Endourology*, *22*, 1557–1560.

[51]Asimov, I. (1950). *I, Robot*. Garden City, N.Y.: Doubleday, p. 40.

[52]Cho, K. W. (2016). Ethical foundations of the radiological protection system. *Annals of the ICRP*, *45*, 297–308.

[53]Alexander et al., op. cit. note 15.

[54]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 into Imaging*, *9*, 745–753; The European Parliament and the Council of the European Union. (2017). *Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC*. http://data.europa.eu/eli/reg/2017/745/2020-04-24 (accessed March 1, 2021); U.S. Food and Drug Administration. (2019). *Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD): Discussion paper and request for feedback*. https://www.fda.gov/media/122535/download (accessed March 1, 2021).

[55]American College of Radiology, op. cit. note 17.

[56]Lee, C. S., & Lee, A. Y. (2020). Clinical applications of continual learning machine learning. *The Lancet Digital Health*, *2*, e279–e281.

[57]Tang et al., op. cit. note 27, p. 125.

[58]Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L., Lijmer, J. G., Moher, D., Rennie, D., de Vet, H. C. W., Kressel, H. Y., Rifai, N., Golub, R. M., Altman, D. G., Hooft, L., Korevaar, D. A., & Cohen, J. F. (2015). STARD 2015. An updated list of essential items for reporting diagnostic accuracy studies. *Radiology*, *277*, 826–832.

[59]Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Annals of Internal Medicine*, *162*(1), 55.

[60]Liu, X., Cruz Rivera, S., Moher, D., Calvert, M. J., Denniston, A. K., & SPIRIT-AI and CONSORT-AI Working Group (2020). Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. *Nature Medicine*, *26*, 1364–1374; Cruz Rivera, S., Liu, X., Chan, A. W., Denniston, A. K., Calvert, M. J., SPIRIT-AI and CONSORT-AI Working Group, SPIRIT-AI and CONSORT-AI Steering Group, & SPIRIT-AI and CONSORT-AI Consensus Group (2020). Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nature Medicine*, *26*, 1351–1363.

prognostication, decision support, and treatment recommendations. Both checklists resemble the intended applications and the attitudes towards explicability that are found in the radiological white papers.

The SPIRIT-AI extension emphasizes explaining the input and output data of AI interventions.[61] For example, it should be specified what level of expertise is required for users (SPIRIT-AI item 11a iv) and how the output of AI interventions will contribute to the decision making in clinical practice (SPIRIT-AI item 11a vi). There is recurrent emphasis on performance errors like decreases in accuracy or erroneous predictions (SPIRIT-AI item 22).

The CONSORT-AI extension demands similar aspects for the reporting of clinical trials that evaluate interventions with an AI component.[62] Like SPIRIT-AI item 22, CONSORT-AI item 19 demands the analysis of errors to prevent harms. The emphasis on performance errors in both guidelines mirrors the wish for safety of medical AI "in recognition that these systems, unlike other health interventions, can unpredictably yield errors that are not easily detectable or explainable by human judgement."[63]

4.3 | Explicability in guidelines for ethical AI

Out of the 84 guidelines reviewed by Jobin et al. (2019) there are 73 that address explicability or similar concepts as a requirement for ethical AI.[64] In their thematic analysis, the codes under which explicability was pooled were most commonly transparency, followed by explainability, explicability, understandability, interpretability, communication, disclosure, and showing. It has been found that there are significant differences in the meaning and justification of these terms. Relevant for medical AI are the aspects of communication and disclosure to increase explainability (the fact that AI is used, evidence-base for AI use, limitations, auditability), minimization of harm, benefits for legal reasons, and fostering trust. Besides transparency, in more than half of the reviewed guidelines there were references to justice and fairness, non-maleficence, responsibility, and privacy. Taken together with our results derived from the radiological white papers, we conclude that there is a big overlap in both the terminology and the justifications for including explicability or a similar concept in radiological white papers and general guidelines for ethical AI.

We noted that explicability has not systematically been introduced as an additional principle to the four bioethical principles, as Floridi and Cowls (2019) proposed.[65] These authors synthesized six guidelines authored by high-profile initiatives into a five-principle approach for ethical AI.[66] They explicitly built upon Beauchamp's and Childress' classic work and added explicability.[67] Floridi and Cowls (2019) argue that explicability complements the four bioethical principles because it meets the technological needs of ethical AI. In their understanding, explicability incorporates both intelligibility ("How does an AI system work?") and accountability ("Who is responsible for the way it works?").[68] Overall, we found five ethical frameworks in the literature that are based on the four bioethical principles and added explicability as a principle.[69]

The introduction of this approach can also be observed from within the radiological community. Akinci D'Antonoli and colleagues endorse the five-principle approach with explicit reference to the four bioethical principles with the addition of explicability.[70] The sources for this five-principle approach are first and foremost Beauchamp and Childress, complemented by the AI4People framework, the European Commission's guidelines for trustworthy AI, the 23 Asilomar principles, and the 10 principles of the Montreal Declaration for a Responsible Development of Artificial Intelligence.[71] The majority of these sources tend to emphasize explicability in one way or another. Most influential is the AI4People framework. It has been developed by a multi-stakeholder group of the European Commission, the European Parliament, civil society organizations, industry and the media under the lead of Luciano Floridi.[72]

The question arises as to why explicability is considered important for medical AI both in general guidelines and in radiological white papers. Two reasons are conceivable, namely (a) the explanation that explicability has been adopted from superordinate policies and (b) the justification with arguments highlighting the new technical peculiarities of "black-box algorithms." Regarding (b), radiological associations deem it important to emphasize explicability on technological grounds, as algorithms are becoming exponentially more complex in contrast to the relatively linear technological improvements of machines traditionally used in the profession. Regarding (a), there is the FDA in the US as a regulatory body that requires explicability for "Software as Medical Device."[73] In the EU there are the article 22 of the GDPR ("right to explanation") and the Medical Device Regulation of 2017.[74] The principle of transparency, the limited interpretability of ML and DL, and explainability of AI outcomes are explicitly

[61]Cruz Rivera et al., ibid: 1358–1359.

[62]Liu et al., op. cit. note 60.

[63]Cruz Rivera et al., op. cit. note 60, p. 1360.

[64]Jobin et al., op. cit. note 1.

[65]Floridi & Cowls, op. cit. note 3.

[66]Ibid.

[67]Beauchamp & Childress, op. cit. note 5.

[68]Floridi & Cowls, op. cit. note 3, p. 8.

[69]Ibid; Floridi et al., op. cit. note 3; Akinci D'Antonoli, op. cit. note 4; European Commission (2019), op. cit. note 3; European Commission (2020), op. cit. note 3.

[70]Akinci D'Antonoli et al., op. cit. note 4; Akinci D'Antonoli, op. cit. note 4.

[71]Beauchamp & Childress, op. cit. note 5; Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke P., & Vayena, E. (2018). *AI4People's Ethical Framework for a Good AI Society: Opportunities, risks, principles, and recommendations*. https://www.eismd.eu/wp-content/uploads/2019/11/AI4People%E2%80%99s-Ethical-Framework-for-a-Good-AI-Society_compressed.pdf (accessed March 1, 2021); Floridi et al., op. cit. note 3; European Commission (2019), op. cit. note 3; Future of Life Institute, op. cit. note 50; University of Montréal. (2018). *Montreal Declaration for a Responsible Development of Artifical Intelligence*. https://www.montrealdeclaration-responsibleai.com/the-declaration (accessed March 1, 2021).

[72]Floridi & Cowls, op. cit. note 3; Floridi et al. (2018), ibid.

[73]U.S. Food and Drug Administration, op. cit. note 54.

[74]The European Parliament and the Council of the European Union, op. cit. note 54; Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "Right to Explanation". *AI Magazine*, *38*, 50–57; European Parliament & European Council, op. cit. note 23.

emphasized in the European Commission's white paper on artificial intelligence.[75] There, it is indicated that the EU was closely involved in developing the Organization for Economic Co-operation and Development's (OECD's) ethical principles for AI that were adopted by the G20 in June 2019.[76] In the OECD's recommendation on AI, transparency and explainability are a joint principle (no. 1.3). We conclude that the appearance of explicability can be explained by the framing of superordinate policies and later on the concept was justified by the need to address the technological peculiarities of AI/ML systems.

4.4 | Explicability and the four principles of biomedical ethics

At least since the 1970s, a principlism approach to medical ethics gained ground to become today's dominant approach. Despite their wide acceptance, there have been efforts to enrich the four bioethical principles. It has been argued that there are special conditions in each respective specialty that require a different set of principles. For example, in public health seven principles are considered important, while in global health nursing practice 10 principles are proposed.[77] In nuclear medicine, 16 principles for ethical AI are proposed, among which we find the four bioethical principles and explicability.[78] In intensive care there are five principles, including explicability.[79]

Is explicability a free-standing principle of biomedical ethics? In light of the different needs of medical specialties, at first sight, one may be tempted to answer yes. The technological particularities of AI may require the adaptation of specific principles. As discussed above, the black-box character of AI is a special challenge for informed consent. Moreover, AI may make it more difficult for physicians to maintain oversight of the different diagnosis processes and fulfill their medical responsibility of non-harm.

There are good reasons to avoid expanding the number of principles. Beauchamp and Childress base their principles on common morality that can be conceived as universal norms. Specifying additional principles may come at the risk of losing this already debatable universal character. Augmenting the number of principles may have the effect of watering down the importance of each principle and makes it more complex to follow their interactions. As long as the properties of explicability are covered by at least one of the four principles of biomedical ethics, explicability may not have to be recognized as a free-standing principle.

The white papers we examined tend to justify explicability in terms of harm reduction. The principle of explicability can thereby be subsumed under the principles of non-maleficence and beneficence, giving explicability mostly an instrumental character. By asking doctors to explain the processes involved in reaching their conclusions, patients can count on a certain degree of human oversight in AI assisted decision-making.[80] Furthermore, by referring to their professional duties, doctors have an argument to insist that technology producers develop explicable AI-systems to adequately fulfill their responsibility of avoiding harm.[81] As an element facilitating informed consent, explicability also has an intrinsic value. Patients may value intrinsically that procedures were followed correctly independently of the outcome. In this position, explicability relates to the principles of justice and respect for autonomy. Health professionals would be failing to respect patients as autonomous agents if they do not recognize them as agents capable of receiving and processing the information that affects them. A communicative process that truly recognizes others as autonomous agents requires a dialogue seeking mutual understanding.[82] In addition, independently of whether their condition can be treated or not, patients may want to understand their individual situation and how it was assessed.

In relation to the principle of justice, patients may claim that they are being discriminated against when they are not given similar opportunities to clear their doubts compared to others.[83] Patients may also appeal to a right to justification when they are concerned that decisions that negatively affect another fundamental right, such as access to healthcare, are being made on unjust or erroneous grounds.[84] Under a social arrangement where people have agreed on a set of rights, everyone has a legitimate claim to a justification of why they are being denied a protected good that they can reasonably expect to access.

Does explicability have an independent value that needs to be recognized as such in medical practice? If we recognize that patients have a right to justification as a demand of justice, the principle of explicability would be reduced to a subcomponent of the broader principle of justice. In spite of this conclusion, an explicit reference to the concept of explicability has an added value. It highlights the fact that communicating to patients involves a greater effort than merely disclosing information. It requires a reflection on how findings were reached and a substantial effort to effectively communicate the involved epistemic processes and their implications in view of the patient's knowledge base. These

[75]European Commission (2020), op. cit. note 3.

[76]Ibid; OECD. (2019). *Recommendation of the Council on Artificial Intelligence*. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (accessed March 1, 2021).

[77]Schröder-Bäck, P., Duncan, P., Sherlaw, W., Brall, C., & Czabanowska, K. (2014). Teaching seven principles for public health ethics: Towards a curriculum for a short course on ethics in public health programmes. *BMC Medical Ethics*, *15*, 73; McDermott-Levy, R., Leffers, J., & Mayaka, J. (2018). Ethical principles and guidelines of global health nursing practice. *Nursing Outlook*, *66*, 473–481.

[78]Currie, G., Hawk, K. E., & Rohren, E. M. (2020). Ethical principles for the application of artificial intelligence (AI) in nuclear medicine. *European Journal of Nuclear Medicine and Molecular Imaging*, *47*, 748–752.

[79]Beil, M., Proft, I., van Heerden, D., Sviri, S., & van Heerden, P. V. (2019). Ethical considerations about artificial intelligence for prognostication in intensive care. *Intensive Care Medicine Experimental*, *7*, 70.

[80]European Commission (2020), op. cit. note 3, pp. 12–13.

[81]Cho, op. cit. note 52.

[82]Freire, P. (1973). *¿Extensión o comunicación? La concientización en el medio rural*. Tierra Nueva, Siglo XXI.

[83]Beauchamp & Childress, op. cit. note 5, pp. 250–251.

[84]Forst, R. (2016). The justification of basic rights. *Netherlands Journal of Legal Philosophy*, *45*, 7–28.

goals can be achieved without treating explicability as a principle. By being aware of the challenges of explicability doctors may gain clarity on whether they are indeed addressing the four principles when informing patients.

5 | CONCLUSIONS

Is a fifth bioethical principle conceptually necessary against the backdrop of medical AI? As we have shown, the properties of explicability are already covered by the four bioethical principles and therefore there is no need for explicability as a fifth principle for biomedical ethics. However, considering explicability honors the epistemic value of AI/ML systems instrumentally for harm prevention and obtaining informed consent.

More specifically, we have examined the reasons that justify the addition of explicability as follows: prevention of harm in case of performance errors, doctors must understand how a result is obtained and communicate it to patients, transparent decision making, the need for informed consent in healthcare, superordinate policies like the EU's General Data Protection Regulation require it, and the technological peculiarities of AI/ML ("black-box algorithms"). Almost all of these reasons are justifications that are already covered by the principles of respect for autonomy, beneficence, and non-maleficence. The adoption from superordinate policies and the new technological peculiarities of AI/ML are explanations for the promotion of explicability in the specialty of radiology.

The conceptual analysis that builds upon the document analysis of white papers shows that explicability mostly is a vehicle for the principle of non-maleficence because there is a wish to reduce harms inflicted by performance errors of medical AI. This is directed at both patients and radiologists because there is the fear of increased liability in case medical AI's high level of automation leads to errors. This is in line with most of the radiological associations that consider the radiologist responsible for diagnosis and prognosis. For cases in which opaque AI systems are used, the radiologist must understand and be capable of explaining how a result was reached. Moreover, liability issues demand from radiologists a substantial amount of trust towards technology producers. This is why radiological associations consider training and educating radiologists in understanding algorithmic decision making important. Furthermore, the opacity of AI/ML systems as a new technological challenge requires elaboration of the explicability of medical AI in imaging. Lastly, superordinate legal policies like the EU's General Data Protection Regulation require the explicability of medical AI. This implies that opaque AI/ML systems must not be applied in clinical practice unless they provide a certain degree of explicability. What this certain degree should be must be the topic of further research.

The original question was whether there is a need to add a fifth principle, i.e., explicability, to the four bioethical principles of autonomy, beneficence, non-maleficence, and justice. Since the instrumental justification of explicability (avoiding harm and facilitating informed decision-making) is already included in non-maleficence and informed consent follows from the principle of respect for autonomy, the four established bioethical principles are sufficient to guide good medical practice. The intrinsic value of explicability is addressed by the concepts of both justice and autonomy. Independently of treatment availability, autonomy requires to know about the situation one is in and why others have assessed it in such a way. Justice requires that others are treated as subjects one can engage in a communicative process as peers of equal standing.[85] Unwillingness to adequately explain issues that are important to another person, particularly in the context of patient-physician relationship, would fail to meet this demand of justice. The wish to integrate explicability into the specialty of radiology is understandable, but medical ethics already has the key principles to handle the new technological specificities of medical AI.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Frank Ursin https://orcid.org/0000-0002-9378-3811
Cristian Timmermann https://orcid.org/0000-0001-7935-2823
Florian Steger https://orcid.org/0000-0001-8108-1591

AUTHOR BIOGRAPHIES

Frank Ursin has been a research associate at the Institute of the History, Philosophy and Ethics of Medicine at Ulm University since 2016. Before that he was a researcher at the Institute for History and Ethics of Medicine at Martin-Luther-University Halle-Wittenberg (2015–2016). He earned a PhD scholarship from the Gerda Henkel Foundation for a PhD project in Ancient History (2012–2015 in Halle/S.), which was completed in 2016. He studied Ancient History, Journalism and Philosophy at Leipzig University (Magister Artium 2011). His main research interests are ancient and early modern medical history as well as ethical questions concerning the digitalization of medicine.

Cristian Timmermann has been a research associate at the Institute of the History, Philosophy and Ethics of Medicine at Ulm University since November 2020. Before coming to Ulm, he was a post-doctoral research fellow at the Jacques Loeb Centre for History and Philosophy of the Life Sciences at the Ben-Gurion University of the Negev in Israel (2013–2014), at the Institute for Philosophical Research at the Universidad Nacional Autónoma de México (2014–2016), and at the Interdisciplinary Center for Studies in Bioethics at the Universidad de Chile (2017–2020). His areas of specialization are medical ethics, research ethics, theories of justice, and applied philosophy in agriculture.

[85] Fraser, N. (1998). Social justice in the age of identity politics: Redistribution, recognition, and participation. In G. B. Peterson (Ed.), *Tanner lectures on human values* (vol. 19, pp. 1–67). University of Utah Press.

FLORIAN STEGER has been Full Professor and Director of the Institute of the History, Philosophy and Ethics of Medicine at Ulm University since 2016. Before that, he was in the same function at the Institute for History and Ethics of Medicine at the Martin-Luther-University Halle-Wittenberg since 2011. He is chairman of the Research Ethics Committee at Ulm University and the Commission "Responsibility in the Conduct of Science" (Good Scientific Practice). He was appointed Leibniz-Professor at Leipzig University in 2014 and was a member of the Junge Akademie in 2009–2014. He earned his habilitation at the University Erlangen-Nuremberg in 2008 and his PhD at the Ruhr-University Bochum in 2002. His research interests are medical ethics, the history of medicine throughout all eras, and the relation between arts and science.

How to cite this article: Ursin, F., Timmermann, C., & Steger, F. (2022). Explicability of artificial intelligence in radiology: Is a fifth bioethical principle conceptually necessary? *Bioethics*, 36, 143–153. https://doi.org/10.1111/bioe.12918