The right to a second opinion on Artificial Intelligence diagnosis—Remedying the inadequacy of a risk-based regulation

Ploug, Thomas; Holm, Søren

Published in:
Bioethics

DOI (link to publication from Publisher):
10.1111/bioe.13124

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Ploug, T., & Holm, S. (2023). The right to a second opinion on Artificial Intelligence diagnosis—Remedying the inadequacy of a risk-based regulation. *Bioethics*, *37*(3), 303-311. https://doi.org/10.1111/bioe.13124

Received: 20 April 2022 | Revised: 13 October 2022 | Accepted: 26 October 2022

DOI: 10.1111/bioe.13124

ORIGINAL ARTICLE

bioethics WILEY

The right to a second opinion on Artificial Intelligence diagnosis—Remedying the inadequacy of a risk-based regulation

Thomas Ploug[1] | Søren Holm[2,3]

[1]Centre of Applied Ethics and Philosophy of Science, Department of Communication and Psychology, Aalborg University, Copenhagen, Denmark

[2]Centre for Social Ethics and Policy, School of Law, University of Manchester, Manchester, UK

[3]Centre for Medical Ethics, University of Oslo, Oslo, Norway

Correspondence
Thomas Ploug, Centre of Applied Ethics and Philosophy of Science, Department of Communication and Psychology, Aalborg University, A C Meyers Vænge 15, 2450 Kbh. SV,Denmark.
Email: ploug@hum.aau.dk

Abstract

In this paper, we argue that patients who are subjects of Artificial Intelligence (AI)-supported diagnosis and treatment planning should have a right to a second opinion, but also that this right should not necessarily be construed as a right to a physician opinion. The right to a second opinion could potentially be satisfied by another independent AI system. Our considerations on the right to second opinion are embedded in the wider debate on different approaches to the regulation of AI, and we conclude the article by providing a number of reasons for preferring a rights-based approach over a risk-based approach.

KEYWORDS

AI regulation, AI rights, AI risks, Artificial Intelligence, second opinion

1 | INTRODUCTION

The use of modern Artificial Intelligence (AI)—Machine Learning and Deep Learning models—for diagnostics and treatment planning holds considerable promise. Such models hold out the prospect of improving patient outcomes and reducing health care costs.[1] Government agencies and scientific institutions alike have issued reports detailing the potential of AI for medical diagnostics and treatment planning,[2] and several AI diagnostic algorithms have already been granted regulatory approval by the FDA.[3] However, the use of AI diagnostic systems has also raised ethical concerns. First, they are not perfectly accurate and may lead to over/underdiagnosis and treatment. A recent comprehensive meta-analysis of AI diagnostic systems in medical imaging and histopathology found, in some instances, 'the diagnostic performance of deep learning models to be equivalent to that of health-care professionals'.[4] That is, some AI systems are currently equivalent but not superior to doctors in this specific context and will therefore still misdiagnose. Second, they may be biased and produce morally unjustified differential treatment of patients. A prediction algorithm widely used in health care was shown to exhibit a racial bias that, if remedied, would increase the percentage of black patients being enroled in a high-risk care management programme in primary care from 17.7% to 46.5%.[5]

[1]Matheny, M. E., Whicher, D., & Thadaney Israni, S. (2020). Artificial intelligence in health care: A report from the National Academy of Medicine. *JAMA*, *323*(6), 509–510; Scott, I. A. (2019). Hope, hype and harms of Big Data. *Internal Medicine Journal*, *49*(1), 126–129.

[2]National Academy of Medicine. (2019). *Artificial intelligence in health care: The hope, the hype, the promise, the peril*.

[3]Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44–56.

[4]Liu, X., Faes, L., Kale, A. U., Wagner, S. K., Fu, D. J., Bruynseels, A., Mahendiran, T., Moraes, G., Shamdas, M., Kern, C., Ledsam, J. R., Schmid, M. K., Balaskas, K., Topol, E. J., Bachmann, L. M., Keane, P. A., & Denniston, A. K. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. *The Lancet Digital Health*, *1*(6), e271–e297.

[5]Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453; Ploug, T., & Holm, S. (2020a). The four dimensions of contestable AI diagnostics—A patient-centric approach to explainable AI. *Artificial Intelligence in Medicine*, *107*, 101901.

The problems and perils of AI systems have attracted significant attention from a host of public and private actors, and several guidelines for the responsible use of AI have been issued. There is general consensus that AI use and development require regulation, but no clear consensus on exactly how this should be done. A systematic review found 84 documents containing ethical principles and guidelines for the use of AI.[6] Most recently, the EU Commission has issued a proposal for the Regulation of AI.[7] The proposal adopts a risk-based approach. AI systems must be classified according to three categories of risk: (1) unacceptable risk, (2) high risk and (3) low or minimal risk. The prohibition against systems having an unacceptable risk covers, among others, AI systems that may be used for manipulation through subliminal techniques (not further defined in the proposal), systems that may cause physical or psychological harm to vulnerable groups and systems used for social scoring by public authorities. The high-risk systems are those that constitute a high risk to the health and safety or fundamental rights of people. They are further divided into systems used in eight different areas, including systems used for determining access to private or public services and benefits. AI systems for diagnostic and treatment planning purposes are likely to fall in the high-risk category. They must be developed on the basis of high-quality data, be sufficiently transparent, be designed so as to allow for human oversight and be consistent in their performance. Developers must also implement a risk management system, that is, continuously map, evaluate and handle risks. However, is a risk-based approach an adequate solution for the use of AI in the health care sector?

We argue in this analysis that a risk-based approach is inadequate for the protection of patients being subjected to AI-supported diagnostics and treatment planning. We suggest that a risk-based approach must be supplemented with individual patient rights and, in particular, a right to a second opinion. This right should not necessarily be construed as a right to a physician opinion. We provide a number of reasons for thinking that it could be satisfied by another independent AI system. A right to a second opinion entails a duty to provide and facilitate a second opinion. We consider who the duty-bearers should be and argue that the duty should mainly be discharged by the health care system. An exhaustive analysis of this issue is, however, beyond the scope of this paper. In making the case for the right to a second opinion, we also sketch the case for a patient right to be offered AI diagnostics and treatment planning in situations where such diagnoses cannot be given a physician-like explanation. For reasons of simplicity, we shall refer in the following only to AI diagnostic systems, but we take most—if not all—of our arguments to apply equally to AI systems for treatment planning.

2 | SECOND OPINION

There are many different contexts in which physicians seek input from other physicians or health care professionals in relation to the diagnosis and treatment decision-making for a particular patient. The 'second opinion' is, however, a very specific type of process. Seeking a second opinion can be requested by either the treating physician or the patient, but it involves getting an independent assessment of a diagnosis or treatment options or both, that is, an assessment that is not influenced by the views of the treating physician and their team. Second opinions can be sought by patients for many reasons, but they are typically only sought when there is something significant at stake for the patient and some reason for the patient to be uncertain about the diagnosis made or treatment recommendation given by the 'first' physician. Studies indicate that this uncertainty on the part of the patient is what makes obtaining a second opinion important for the patient.[8]

In current practice, a second opinion can only be sought for nonurgent and elective treatments, because it is only in that context that there is sufficient time for another physician to be approached to provide a second opinion, and for that physician to study the patient's notes, and potentially examine the patient or acquire new diagnostic tests, before providing the independent second opinion.[9] However, as discussed below in the section 'The automation bias and deskilling arguments', one of the justifications for instituting a right to a second AI opinion in certain contexts is that these are contexts where automation bias and/or deskilling are likely to occur. This will include a range of acute treatment contexts since automation bias is more likely to occur when the decision-making is made under time pressure. Facilitating a second AI opinion in these acute contexts will therefore require a different approach than is currently used for obtaining a second opinion.

Second opinions raise particular resource allocation or payment issues and issues concerning the implementation of second opinion systems in the health care context. Who should pay for the costs of the second opinion? Should it be the patient or the third-party payer in the particular health care system? Does the health care system have an obligation to facilitate second-opinion requests? In many

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

[7]European Commission. (2021). EUR-Lex—52021PC0206—EN. *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*.

[8]Greenfield, G., Pliskin, J. S., Feder-Bubis, P., Wientroub, S., & Davidovitch, N. (2012). Patient–physician relationships in second opinion encounters—The physicians' perspective. *Social Science & Medicine*, *75*(7), 1202–1212; Payne, V. L., Singh, H., Meyer, A. N. D., Levy, L., Harrison, D., & Graber, M. L. (2014). Patient-initiated second opinions: Systematic review of characteristics and impact on diagnosis, treatment, and satisfaction. *Mayo Clinic Proceedings*, *89*(5), 687–696; Philip, J., Gold, M., Schwarz, M., & Komesaroff, P. (2011). An exploration of the dynamics and influences upon second medical opinion consultations in cancer care. *Asia-Pacific Journal of Clinical Oncology*, *7*(1), 41–46; Shmueli, L., Davidovitch, N., Pliskin, J. S., Hekselman, I., Balicer, R. D., & Greenfield, G. (2019). Reasons, perceived outcomes and characteristics of second-opinion seekers: Are there differences in private vs. public settings? *BMC Health Services Research*, *19*(1), 238.

[9]Arthur, A. S., Mocco, J., Linfante, I., Fiorella, D., Hussain, M. S., Jovin, T. G., Nogueira, R., Schirmer, C., Barr, J. D., Meyers, P. M., De Leacy, R., & Albuquerque, F. C. (2018). Stroke patients can't ask for a second opinion: A multi-specialty response to The Joint Commission's recent suspension of individual stroke surgeon training and volume standards. *Journal of NeuroInterventional Surgery*, *10*(12), 1127–1129; Gagin, R., HaGani, N., Zigelboim, E., & Shinan-Altman, S. (2019). Patient-initiated second opinions during acute hospital care. *Patient Experience Journal*, *6*(3), 66–73.

health care systems, there will not be free access to second opinions, but third-party payers will have criteria for the situations in which they will pay for a second opinion. In some jurisdictions, third-party payers must offer to pay for second opinions in relation to particular procedures.[10] It is beyond the scope of this paper to provide a full analysis and discussion of when a third-party payer should fund an AI second opinion, but criteria will have to be put in place relating to factors such as the importance of a decision for the treatment of the patient and the likelihood that a particular AI system provides erroneous advice. Below, we argue that there is a positive right to a second opinion and that they should be facilitated by the health care system. A full account of what this duty on behalf of the health care system entails is also beyond the scope of this paper. In jurisdictions where physician-provided second opinions are already available for free or at minimal cost, our arguments straightforwardly entail that this should be extended to appropriate AI-provided second opinions.

3 | THE RIGHT TO A SECOND OPINION AS A RIGHT TO PROTECT ONESELF

The right to a second opinion, we are discussing, is a positive right to have an *independent* opinion on the soundness and adequacy of one's diagnosis and treatment plan. That is, it is not simply a right not to be prevented from seeking a second opinion. The contours of the right will become clearer as the analysis progresses. Why should patients be granted such a right in relation to AI diagnostics?

3.1 | The inevitability of AI harm

One set of reasons for the right to a second opinion comes from peoples' right to protect themselves against harm. AI-driven diagnostics may cause harm if it is inaccurate and biased. A risk-based regulation may take steps to minimise these harms, but they are unlikely to be fully eliminated for a combination of technical and ethical reasons.

Perfect accuracy of AI diagnostics is unachievable. It requires that the diagnostic models are taking all relevant parameters into account, it requires immense amounts of accurate data for training purposes and it requires accurate and complete input data for a particular patient. With the limitations on the availability of 'noise-free' data sets and under the pressure of a competitive market, the *development* of AI diagnostic systems cannot in practice be expected to achieve perfect accuracy. Under the constraints of data collection in real-life, clinical situations, all actual *use* of AI diagnostic systems cannot be expected to produce completely accurate diagnostics. The inevitability in practice of inaccurate diagnostics is reflected in empirical studies. As mentioned in the introduction, recent studies suggest that although AI systems may outperform physicians in a few cases, the general performance of AI diagnostics in fields such as medical imaging and histopathology is on a par with physicians—and hence not perfectly accurate. Other studies suggest that the combination of physician and AI diagnostics does not achieve perfect accuracy either.[11]

However, even if it was possible at some point in the future to achieve perfect accuracy and completely unbiased diagnostics of AI systems as standalone systems or of the combination of physicians and AI systems, and even if it was possible to somehow ascertain that this point had been reached, it would arguably be unethical to regulate AI now based on the prediction that it will achieve perfect accuracy. This would almost certainly lead to underregulation, that is, a failure to prevent preventable harm. It seems that we pass an important ethical threshold for implementation already when AI-driven diagnostics is at least as good as average physician diagnostic decision-making in the particular domain. At this point, there may be potential benefits in terms of reduced health care costs, and beyond this point, there may be improvements in patient outcomes. An ethically justifiable, risk-based approach to the regulation of current AI use therefore cannot, for good reason, guarantee that patients subjected to diagnostics involving AI will not come to suffer inaccurate diagnostics and consequently come to suffer the associated harms. An adequate regulation of AI should grant patients a right to the means to protect themselves *effectively* against this harm, as far as it is possible.

3.2 | The inadequacy of competing rights

There are several rights that might seem to do the job here. The right to provide or refuse informed consent is one such right. By providing or withholding consent to AI diagnostics, patients can act on what they take to be acceptable risks in relation to AI diagnostics. Similarly with the right to be diagnosed entirely by physicians. It has been suggested that patients should have the right to withdraw from AI diagnostics (not from the involvement of AI equipment altogether) and insist that the act of providing a diagnosis and/or a treatment plan is conducted entirely by physicians.[12] Whereas the right to provide or deny informed consent is rooted in ethical considerations, the right to be diagnosed entirely by physicians is best justified as a protection of the citizen as a political actor. It is a right that allows individuals to act on their societal concerns about the role of AI in future society in so far as these concerns satisfy conditions of

[10]Pieper, D., Heß, S., & Mathes, T. (2018). Bestandsaufnahme zu Zweitmeinungsverfahren in der Gesetzlichen Krankenversicherung (GKV). *Das Gesundheitswesen*, *80*(10), 859–863.

[11]Salim, M., Wåhlin, E., Dembrower, K., Azavedo, E., Foukakis, T., Liu, Y., Smith, K., Eklund, M., & Strand, F. (2020). External evaluation of 3 commercial artificial intelligence algorithms for independent assessment of screening mammograms. *JAMA Oncology*, *6*(10), 1581–1588; Wu, N., Phang, J., Park, J., Shen, Y., Huang, Z., Zorin, M., Jastrzębski, S., Févry, T., Katsnelson, J., Kim, E., Wolfson, S., Parikh, U., Gaddam, S., Lin, L. L. Y., Ho, K., Weinstein, K. D., Reig, B., Gao, Y., Toth, H., … Geras, K. J. (2020). Deep neural networks improve radiologists' performance in breast cancer screening. *IEEE Transactions on Medical Imaging*, *39*(4), 1184–1194; Zhou, Q., Zuley, M., Guo, Y., Yang, L., Nair, B., Vargo, A., Ghannam, S., Arefan, D., & Wu, S. (2021). A machine and human reader study on AI diagnosis model safety under attacks of adversarial images. *Nature Communications*, *12*(1), 1–11

[12]Ploug, T., & Holm, S. (2020b). The right to refuse diagnostics and treatment planning by artificial intelligence. *Medicine, Health Care and Philosophy*, *23*(1), 107–114.

rationality and public reason. In exercising this political right, patients may also protect themselves against suffering the harms of inaccurate and biased decision-making.

Both these rights may, however, be considered *ineffective* for three reasons. First, because they do not empower patients to take action against the cause of the harm—only against the harm itself. They empower patients to act so as to avoid the consequences of inaccurate and biased decision-making, but they do not empower patients to correct the actual inaccuracy or bias. Second, they are rights that preclude the patients from enjoying the benefits of AI system use. Exercising these rights not only rule out the potential suffering of the relevant harms but also the possibility of enjoying improved health care outcomes. Third, if these benefits become significant—as they likely will as AI system performance progresses in the future—the protective powers of these rights may *de facto* be traded off. The gains may push people to choose AI-driven diagnostics. Several studies on how the provision of informed consent may become routinised—that is, provided as an unreflective, habitual act—suggest that this phenomenon is partly driven by getting access to the rewards that can only be obtained by consenting.[13]

The right to a second opinion does the job without incurring these problems. It is a right that empowers patients to take steps to ensure more accurate and unbiased diagnostics without being deprived of the opportunity of enjoying the potential benefits of AI use. Also, since it does not rule out enjoying the benefits of AI use, it may not to the same extent tempt patients to give up on this right as the benefits of AI involvement get bigger. The latter point is, however, ultimately a question that must be settled through empirical studies.

3.3 | The right to a second opinion as a positive right

The right to a second opinion does not straightforwardly follow from patients' right to protect themselves against harm. The right to a second opinion is a positive right in the sense that it confers on the health care services a moral obligation to provide an option—the setting up of a procedure and system for second opinions. This positive right is not directly entailed by the right to protect oneself against harm.

The problem is not unique. It also applies to the right to provide or refuse informed consent. Neither the right to protect oneself against harm nor the right to autonomy entails a right to be informed about all relevant aspects of a procedure and offered a choice. However, it seems that informed consent is both in principle and in practice an *acceptable method* for ascertaining that these rights are not violated.

[13]Ploug, T., & Holm, S. (2013). Informed consent and routinisation. *Journal of Medical Ethics*, *39*(4), 214–218; Ploug, T., & Holm, S. (2015). Routinisation of informed consent in online health care systems. *International Journal of Medical Informatics*, *84*(4), 229–236.

Similarly, there is a need here for independent reasons for assuming that the right to a second opinion will both in principle and in practice be an acceptable method for protecting individuals' rights. We have already argued that the right to a second opinion is particularly effective in protecting individuals against harmful AI diagnostics. In a later section, we aim to show that there are no compelling reasons against this right.

4 | THE RIGHT TO A SECOND OPINION IN RELATION TO AI USE IN THE CLINICAL SETTING

The general structure of our argument in favour of a right to a second opinion is rather simple:

Premise 1: Individuals have a right to protect themselves effectively against harm.

Premise 2: AI will inevitably cause harm to individuals through inaccurate and biased decision-making.

Premise 3: Competing rights such as informed consent and the right to withdraw altogether from AI involvement in diagnosis and treatment planning are not as effective as a right to a second opinion in protecting individuals against harms caused by inaccurate and biased decision-making.

Premise 4: A right to a second opinion is an effective way for individuals to protect themselves against some of the harm caused by inaccurate and biased decision-making.

Premise 5: There are no compelling reasons not to introduce the right to a second opinion.

Conclusion: Individuals (should) have a right to a second opinion in relation to AI use.

In premise 2, the argument is made relative to AI use. Physicians may also cause harm through inaccurate and biased decision-making, so why limit it to AI use? There are several intertwined reasons for thinking that AI systems present a special case.

4.1 | The extended explainability problem

In the literature, there has been significant analysis of the importance of explainability in relation to the implementation of AI in health care. These analyses have been focused on the black-box issue. There is, however, a much more important normative issue at stake, as we will show below.

Many modern AI systems—Machine Learning and Deep Learning models—are black-boxes. The complexity of these models means that diagnostic and treatment suggestions cannot be easily explained. All the parameters feeding into a classification cannot be explicated. The reasoning cannot be fully replicated stepwise. A physician's diagnostic decisions are arrived at in a different way. They have been argued to follow—as a cognitive matter of fact—the pattern of an inference

to the best possible explanation.[14] Inference to the best possible explanation is a sort of inference that moves from premises detailing the signs, symptoms and indicators by way of a consideration of the adequacy of competing explanatory hypotheses to a conclusion establishing the hypothesis that best explains the symptoms. This process is different from that of arriving at a diagnosis through a model based on the weighing of numerous parameters of a certain health condition in that all premises for an explanation can be made explicit and the reasoning can be fully replicated stepwise. The physician can therefore engage in a dialogical process of reason-giving when asked to explain a diagnosis. This is the case even if physicians make judgements by drawing reflexively on professional experience. Saying that 'I have seen many patients with symptoms like yours and they all had X' is also giving a reason for the diagnosis.

There is, however, a more fundamental, morally relevant difference between the explainability of physician and AI diagnostics. The question of what explainability is cannot be separated from the act of providing an explanation in a particular situation with constraints of different sorts. Thus, in providing an explanation to the patient, a physician must satisfy the informational requirements following from the patient's right to make informed choices. The physician must ensure that the patient receives adequate information, that this information is (sufficiently) understood by the patient, and that the patient is not under undue influence. In satisfying these requirements, the physician must judge the patient's cognitive powers and informational interests, with special attention to their particular situation and their particular vulnerabilities. The guiding aim of this situational judgement and the provision of information is that patients should be empowered to make choices that protect them against harm and allows them to act on their wider preferences and interests.

The point to be made here is that the explanation of a diagnosis or treatment plan to a patient is an exercise that not only involves explaining why the diagnosis was the best possible explanation of the set of signs, symptoms and indicators. It is also an exercise that is governed by the patients' rights and that draws upon highly situational, evaluative judgements. Explaining diagnostic choices and treatment decisions is a normative—and more specifically, a moral—act. The current endeavours in explainable AI to a large extent focus on how to detect and visualise key features and parameters in ML and Deep Learning classifications.[15] However, it is anticipated that AI diagnostic systems will outperform physicians in particular medical fields long before they are able to provide explanations with special sensitivity towards human vulnerabilities and patient rights. As some scholars have argued, it is simply easier to make AI systems perform technical tasks than to make them share our moral outlook.[16] This is the case whether or not an AI system is technically an unexplainable black-box.

4.2 | The right to be offered inexplainable AI diagnostics

But does it really matter that physician-like explanations cannot be provided by an AI system on its own? After all, it is a widespread view that AI systems should only be used as clinical decision support systems (CDSS) or as second opinions on physician diagnoses and treatment plans.[17] In this view, physicians must be able to spell out why a diagnosis is the best possible explanation of a set of symptoms without reference to the diagnosis of the AI CDSS. That is, they cannot simply claim that a diagnosis is the best explanation because it was determined by an AI CDSS with a performance that is statistically superior to that of physicians. So defined, the independence of AI physician diagnostics entails full physician explainability. Thus, if physicians were to decide a diagnosis independently of AI CDSS in this way, they could also, in the encounter with patients, provide explanations taking all relevant normative constraints into account. Their independence would be compromised only if the AI CDSS somehow led them to decide on a diagnosis without being able to account for why it was the best possible explanation of the set of symptoms.

However, insisting that physician diagnostics be in this way completely independent of AI systems use may be morally unjustified. AI diagnostic systems may, in the not-too-distant future, perform significantly better than physicians. Would it, in such circumstances, be justifiable not to offer patients AI-generated diagnoses simply because they cannot be reconstructed by physicians? There are strong arguments against this. First, the primary obligation of health care institutions is to provide the best possible health care, and this value is deeply embedded in the public´s expectations to health care institutions.[18] The public will therefore expect AI systems to be used if they perform better than physicians and improve care. Protecting the public´s expectations should be done for reasons of both autonomy and trust. Maintaining a state of the world—a state of health care—that fits peoples' expectations makes the world reliable, and reliability not only allows individuals to plan and pursue their own longer-term goals and values accordingly, but arguably also drives trust.[19] So, in offering high-performance AI-driven diagnostics, the health care institutions would not only protect patients' health but also their autonomy and trust.

[14]Dragulinescu, S. (2016). Inference to the best explanation and mechanisms in medicine. *Theoretical Medicine and Bioethics*, *37*(3), 211–232.

[15]Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health*, *3*(11), e745–e750.

[16]Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

[17]Kempt, H., & Nagel, S. K. (2022). Responsibility, second opinions and peer-disagreement: Ethical and epistemological challenges of using AI in clinical diagnostic contexts. *Journal of Medical Ethics*, *48*(4), 222–229.

[18]Coulter, A. (2005). What do patients and the public want from primary care? *BMJ*, *331*(7526), 1199–1201; Naidu, A. (2009). Factors affecting patient satisfaction and healthcare quality. *International Journal of Health Care Quality Assurance*, *22*(4), 366–381.

[19]Francis, L. P. (1992). Consumer expectations and access to health care. *University of Pennsylvania Law Review*, *140*, 1881–1917.

Second, we also take the requirement of independent explainability to be unjustifiable because we believe that one of the primary reasons for insisting on full physician-like explainability—protecting the patients against harm—can be given *sufficient weight* by introducing the right to a second opinion from another AI system (see below). Third, and relatedly, there are alternative and *sufficiently adequate* 'explainability-like' notions that AI systems can readily satisfy.[20] Recent writings on the right to contest AI decision-making suggest that this right cannot be exercised properly unless patients are provided with four types of information about an AI system involved in their diagnostics. Patients must be afforded information about the (1) the AI system's use of data, (2) the system's potential biases, (3) the system performance and (4) the division of labour between the system and health care professionals. This information does not amount to a physician-like explanation, but it is an explanation that empowers patients not only to act on their right to contest AI diagnostics but also to act on the right to a second opinion. The combination of these rights may thus be sufficient to protect patients against the harm that they may incur from AI diagnostics that cannot be given a physician-like explanation.

An important observation must be made here. The right substantiated in this section complements a legal right in the European Union's General Data Protection Regulation (GDPR). Article 22 of the GDPR guarantees individuals a right 'not to be subject to a decision based solely on automated processing'.[21] In the words of GDPR, what we have been arguing here is that—at least in health care—individuals should have a complementary right to be offered a decision based solely on automated processing if it is likely to produce a better outcome than through human participation. In other words, patients should both have a right not to be subject to decision-making based solely on automated processing and a right to choose to be subject to decision-making based solely on automated processing.

4.3 | The automation bias and deskilling arguments

If AI systems in the not-too-distant future perform significantly better than physicians, then it seems likely that AI diagnostics will inevitably come to strongly influence physician diagnostics in an everyday clinical setting. There is ample evidence of the phenomenon of automation bias in relation to already existing CDSS, that is, physicians overrelying on or substituting the advice of the CDSS for their own professional judgement.[22] The evidence suggests that automation bias is driven by a range of factors including time pressure, physician self-confidence and complacency, trust in and experience with CDSS, task experience and complexity.[23] Automation bias is by definition the surrendering of independent physician diagnostics. It entails that *full* physician explainability cannot be achieved. Extrapolating this evidence to a situation in which AI diagnostic systems are performing significantly better than physicians, automation bias will likely become a more widespread phenomenon.

The use of CDSS may also, over time, lead to the deskilling of physicians, that is, the loss of skills necessary for physicians to perform their job adequately.[24] Deskilling has been defined as the experience of reduced discretion, autonomy, decision-making quality and knowledge.[25] A study found that the introduction of electronic medical records (EMRs) into clinical practice had deskilling outcomes for primary care physicians. In particular, they experienced a loss of clinical knowledge, more stereotyping of patients and also reduced confidence in making clinical decisions as a consequence of using EMR.[26] There is a potential vicious circle here with deskilling reinforcing automation bias and vice versa.

The upshot is this. Even if one could produce sufficiently strong moral reasons for maintaining a system with physicians having to make their own independent diagnostics and thereby retain the ability to provide explanations via reason-giving, the phenomena of automation bias and deskilling make it unlikely that this independence can be *fully* upheld in practice if AI systems are introduced into clinical care without particular attention to these issues. Clearly, there is room for further studies and debate here. Three things should be noted, however. First, that while the right to be offered AI diagnostics presupposes that AI systems outperform physicians—being more accurate and/or less biased—the argument from automation bias and deskilling does not rest on this presupposition. We have argued here that significantly better performance by AI systems may increase the extent of automation bias and deskilling, but the evidence suggests that automation bias and deskilling are already existing phenomena,

[20]Ploug & Holm, op. cit. note 5.

[21]European Commission. (2016). EUR-Lex—32016R0679—EN. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).*

[22]Alberdi, E., Povyakalo, A., Strigini, L., & Ayton, P. (2004). Effects of incorrect computer-aided detection (CAD) output on human decision-making in mammography. *Academic Radiology*, *11*(8), 909–918; Bogun, F., Anh, D., Kalahasty, G., Wissner, E., Bou Serhal, C., Bazzi, R., Douglas Weaver, W., & Schuger, C. (2004). Misdiagnosis of atrial fibrillation and its clinical consequences. *The American Journal of Medicine*, *117*(9), 636–642; Bond, R. R., Novotny, T., Andrsova, I., Koc, L., Sisakova, M., Finlay, D., Guldenring, D., McLaughlin, J., Peace, A., McGilligan, V., Leslie, S. J., Wang, H., & Malik, M. (2018). Automation bias in medicine: The influence of automated diagnoses on interpreter accuracy and uncertainty when reading electrocardiograms. *Journal of Electrocardiology*, *51*(6), S6–S11; Golchin, K., & Roudsari, A. (2011). Study of the effects of clinical decision support system's incorrect advice and clinical case difficulty onusers' decision making accuracy. In Borycki, E., Bartle-Clar, J. A., Househ, M. S., Kuziemski, C. E., & Schraa, E. G. (Eds.). *International perspectives in health informatics* (pp. 13–16). IOS Press. Tsai, T. L., Fridsma, D. B., & Gatti, G. (2003). Computer decision support as a source of interpretation error: The case of electrocardiograms. *Journal of the American Medical Informatics Association*, *10*(5), 478–483.

[23]Goddard, K., Roudsari, A., & Wyatt, J. C. (2011). Automation bias—A hidden issue for clinical decision support system use. In Borycki, E., Bartle-Clar, J. A., Househ, M. S., Kuziemski, C. E., & Schraa, E. G. (Eds.). *International Perspectives in Health Informatics* (pp. 17–22). IOS Press; Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, *19*(1), 121–127. Goddard, K., Roudsari, A., & Wyatt, J. C. (2014). Automation bias: Empirical results assessing influencing factors. *International Journal of Medical Informatics*, *83*(5), 368–375.

[24]Cabitza, F., Rasoini, R., & Gensini, G. F. (2017). Unintended consequences of machine learning in medicine. *JAMA*, *318*(6), 517–518; Vellido, A. (2019). Societal issues concerning the application of artificial intelligence in medicine. *Kidney Diseases*, *5*(1), 11–17.

[25]Hoff, T. (2011). Deskilling and adaptation among primary care physicians using two work innovations. *Health Care Management Review*, *36*(4), 338–348.

[26]Ibid.

and this is all our argument requires. Second, it should also be noted that the occurrence of automation bias is used here as grounds for a right to a second opinion specifically in relation to AI use. It is an argument in favour of limiting the right to a second opinion to AI use—not for the right to a second opinion as such. Third, and finally, note that both the argument from the right to be offered physician inexplainable diagnostics and the argument from automation bias and deskilling are each on their own sufficient ground for the right to a second opinion in relation to AI diagnostics.

If significant deskilling occurs, this will also affect clinical encounters where the decision-making of the physician is not supported by an AI system. Such cases may strengthen the argument for a general right to a second opinion to physician diagnosis and treatment planning.

4.4 | The argument from systematic errors

AI systems make systematic errors. That is, given the same input data, a 'locked' AI model will provide the same output. If the output is inaccurate, then it will be inaccurate whenever the model is provided with the same input data. Applied to AI diagnostics, if the AI system provides an inaccurate diagnosis or treatment plan, then it will do so whenever it is fed the same input data. The systematic character of AI errors means that an AI model cannot meaningfully act as second opinion for itself. An 'unlocked' AI model will change and adapt over time, but the time span between a primary diagnosis and a second opinion is usually fairly short, and it is therefore unlikely that a system will have changed significantly in that period. It would, in some sense, be worrying if an AI system provided one diagnosis and then a different diagnosis based on the same data within a short time span.

While physicians may make systematic diagnostic errors, it seems a fair assumption that these errors are at least sometimes unsystematic in exactly the sense that provided with the same signs, symptoms and indicators, they will not necessarily make the same diagnostic error twice. It is, in other words, possible to avoid making the same mistake twice. However, the ability to correct previous mistakes means that physicians can act as second opinions for themselves. It makes sense for a patient to contest the rightness of a diagnosis and ask the treating physician to reconsider these. Findings in behavioural psychology suggest that it may be difficult to change beliefs. Thus, the cognitive bias known as confirmation bias denotes the tendency of human beings to look for information that confirms existing beliefs.[27] Whereas such cognitive biases may certainly make it difficult to change one's preconceptions, they do not make it impossible.

This difference between physician and AI diagnostics has implications for the right to a second opinion. If patients may be subjected to AI diagnostics without physician-like explainability, then they do not have the option of contesting the decision and asking for their diagnosis to be reconsidered as they would have in the case of physician diagnostics. We believe that this counts in favour of granting patients a right to a second opinion specifically in relation to AI diagnostics.

5 | DOES A SECOND OPINION HAVE TO BE AI INDEPENDENT?

As defined throughout the previous sections, the right to a second opinion has been left unspecific as to whom should provide a second opinion. It may have been tacitly assumed that it would have to be a physician, but does it really have to?

On the one hand, if AI diagnostic systems outperform physicians in the future, and if AI will pervade health care and automation bias becomes a widespread phenomenon, then the right to a second opinion may well be a right to be offered an AI-generated second opinion. Moreover, the right to a second opinion is a right that comes at a cost. It will be an added burden to health care professionals in their everyday clinical work. An AI-generated second opinion could alleviate this burden. Qua the problem of systematic errors, an AI second opinion would have to come from a wholly independent AI system. It would have to be based on a model or algorithm being substantially independent of the model or algorithm having provided the diagnosis. Developing such a system will certainly also carry costs, but we foresee a market situation with many competing AI systems for diagnostics across a range of medical fields. In this situation, AI-generated second opinions may come at a reasonable price.

On the other hand, if AI diagnostics is not fully explainable by a physician, then in the interest of securing maximal transparency, the right to a second opinion could be construed as a right to an independent physician opinion. Getting an independent physician opinion may become increasingly difficult in a future with increasing levels of automation bias and deskilling. A physician opinion would also make demands on physician resources and therefore add costs to health care at the clinical level. It may thus potentially involve hard choices of prioritisation in the everyday clinical work. For reasons of trust, a physician opinion may, however, be the preferred choice of patients.

Where do these considerations leave us in terms of deciding whether the right to a second opinion should be a right to an AI or physician opinion. Offering the choice between an AI and physician opinion would be a way of balancing the conflicting concerns. Patients would be in principle be empowered to balance their preferences in performance, explainability and trust. As AI diagnostics is developed and implemented extensively in clinical care along with a right to a second opinion, evidence of the performance and explainability, the level of automation bias and deskilling, the extent to which patients exercise their right to a second opinion and the costs of maintaining a certain regulatory framework will gradually become available. Such evidence may tip the scales in relation to the set of options that patients should be offered. It may very well be the case that patients should be offered only an AI-generated second opinion.

[27]Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

6 | CONCLUSION

6.1 | A risk- or rights-based approach to regulation

In the introduction, we suggested that a risk-based approach to AI regulation is inadequate. A risk-based approach operates at the level of the population. It protects individuals and their interests only indirectly. On the backdrop of our considerations throughout the article, we are now in a position to clarify exactly why such an approach is inadequate. Essentially, a risk-based approach provides insufficient protection of the individual because some of the interests in need of protection can only be protected at the individual level, where a patient interacts with a health care system. There are four main arguments detailing this.

First and foremost, a risk-based approach is inadequate because it does not empower individuals to protect themselves against the harm that they may suffer. A risk-based approach can either halt AI development and use or impose restrictions based on an evaluation of the risk. Halting AI development would, for reasons already rehearsed, be unjustified. Imposing restrictions to minimise harm will not nullify the risk. The risk of serious harm in the wake of using AI diagnostic systems will most certainly persist. A risk-based approach can only deal with this risk at the population level. It cannot address the specific risks that a specific patient may be subjected to in a specific clinical situation with a specific implementation of an AI diagnostic system. The risk in a specific situation can, however, be addressed. It can be addressed by the patients if they have the appropriate rights to do so. A right to a second opinion would be one such right.

Second, an existing risk-based regulation is inadequate because it cannot address the uncertain and unpredictable dynamic effects of introducing AI systems in health care and elsewhere. A risk-based approach to AI regulation ultimately works by assessing the risks of AI systems. It cannot—for good reason—evaluate and regulate the wider effects of introducing AI systems in different societal contexts, including health care, given the uncertainty surrounding such effects. That is, it cannot fully take into account the risks of automation bias and deskilling and the effects on public trust. There are two problems here. First, the full extent of dynamic effects such as automation bias and deskilling cannot be predicted accurately prior to the implementation of a specific system. Second, many AI systems are going to be implemented in the clinical setting, and the types of automation bias and deskilling that these interacting systems are going to create are likely to produce even more unpredictability. The dynamic unpredictability ensuing from both of these problems cannot be resolved reasonably by risk assessments prior to the introduction of a specific AI system. A rights-based approach can address such dynamic effects. Rights can be designed in ways that addresses the uncertainties of possible dynamic effects. Rights can be designed in ways that provide individuals with ongoing protection in the face of varying dynamic effects. As we have shown here, the right to a second opinion empowers individuals to act to protect themselves against the harms of automation bias and to maintain trust and to give priority to certain kinds of explainability over others.

Third, a risk-based approach to AI regulation can hardly be adjusted in ways where the consequences of such adjustments are predictable at the individual level. It is a top-down approach with unpredictability at the level of individual patient protection. If, for instance, a certain set of criteria for risk classification turns out to have devastating consequences for the costs of developing and using AI diagnostic systems, then adjustments in the risk criteria may have unpredictable consequences at the individual level. That is, it may be hard to predict what the consequences of lowering the threshold of performance or bias will be for individuals. By contrast, a rights-based approach is a bottom-up process where the adjustments of regulation start at the individual patient level. The consequences for patient protection are the starting point for adjustments of the regulation. Thus, as we have shown, the right to a second opinion may be adjusted with the regard to the question of whether a physician and/or an AI system should be offered as the second opinion in view of the costs of these different solutions for the health care system. As is evident, the consequences for individual patient protection would be evident in such an adjustment of the regulation.

Fourth and finally, a risk-based approach is fundamentally a paternalistic approach to the protection of individuals. It introduces a centralised system for risk assessment that, so to speak, bypasses individuals. That is, the acceptability and unacceptability of risks are determined away from the individual. It is determined without individuals having a right to define and act on what they take to be acceptable and unacceptable risks. The paternalism of a risk-based approach is inevitable but becomes a problem because there are mechanisms by which we can enable much more individual decision-making. The right to a second opinion remedies the general paternalism of a risk-based approach by empowering individuals to act on their own perceptions of risks.

For these reasons, we are convinced that a risk-based approach to the regulation of AI is inadequate in the sense that it provides insufficient protection of individuals and their interests. We do believe, however, that a risk-based approach to the regulation of AI is necessary. It relieves individuals of some of the burden of self-protection—and it may regulate AI development and use with a view to the wider societal interests in this. Thus, in conclusion, we are committed to the view that a risk-based approach should be supplemented with a strong rights-based regulation including, among others, the right to a second opinion and the right to be offered inexplainable AI diagnostics for that matter.

6.2 | Future research

Much research on these issues is needed both conceptually and empirically.

In relation to the right to a second opinion, there is specifically need for further work on what constitutes an independent AI opinion. As argued above, the acceptability of another AI system

acting as second opinion hinges crucially on the independence of such a system. But what does it take for an AI diagnosis to be substantially independent of the diagnosis of another AI system?

All throughout this article, our arguments have rested on assumptions about the future performance of AI and the dynamic effects of such changes. We have carefully tried to support our assumptions by relevant evidence in accordance with the ideals of evidence-based policymaking. There are, however, obvious uncertainties regarding some of these effects and mechanisms. These uncertainties must be addressed through continued empirical work. Giving priority to the regulation of AI is just as much a matter of securing continued empirical research into such issues.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Thomas Ploug http://orcid.org/0000-0002-3693-0547

AUTHOR BIOGRAPHIES

Thomas Ploug is a Danish philosopher with a special interest in AI- and Dataethics. He is professor of information and communication ethics at the Centre of Applied Ethics and Philosophy of Science, Aalborg University, Denmark.

Søren Holm is a Danish doctor and philosopher. He is professor of bioethics at the Centre for Social Ethics and Policy, University of Manchester, UK, and professor of medical ethics (part-time) at the Centre for Medical Ethics, University of Oslo, UK.

How to cite this article: Ploug, T., & Holm, S. (2023). The right to a second opinion on Artificial Intelligence diagnosis—Remedying the inadequacy of a risk-based regulation. *Bioethics*, 37, 303–311. https://doi.org/10.1111/bioe.13124