# AI in Healthcare: Are We Jumping the Gun?

Odin Hoff Gardå | October 19, 2023

With the rapid advancement of machine learning technology recently, such technologies are naturally finding their way into the healthcare systems around the world. A recent example is the announcement by Vestre Viken Hospital Trust in Norway that they will start to use machine learning models to diagnose patients based on radiographic image s (see, for example, Vestre Viken, 2022 and Topdahl et al., 2023). As is the case with most new and disruptive technologies when introduced into society, there are ethical considerations to be made. The Norwegian Directorate of Health has made available some ethical recommendations for the use of AI (Helsedirektoratet, 2022), but these are arguably lacking and too general to serve the intended purpose. This, together with the current AI hype, calls for an open discussion on the topic as this kind of technology has the potential to directly impact people's lives and cause real harm in our society.

In this text, we seek to give an answer to the following question:

> **Is healthcare[1] ready to adopt the use of AI technology?**

By calling attention to numerous non-trivial challenges with the current technology, we argue that the answer to the above question is negative. Despite our somewhat pessimistic conclusion, we also emphasize some of the advantages of the use of AI in healthcare. We note that many of arguments given in this text are also applicable to other uses of AI in society where the output of these algorithms directly, or indirectly, affects people's lives.

## Scope and definitions

We begin by establishing the scope of this text, and by defining some central terms. The term AI[2] is used ambiguously and take various meanings in the literature. We will use terms such as *AI*, *AI model* and *AI (based) systems* as synonyms for machine learning algorithms that inform decision-makers (in our case health professionals), or make independent decisions or recommendations based on input data. We will also use the terms *medical AI* and *AI in healthcare* synonymously. In our setting, input data may include, but is not limited to, radiographic images, MRI scans, ultrasound images, electroencephalography (ECG), electrocardiography (EEG), blood test results, personal information, lifestyle choices, gene data and family disease history. It is not difficult to imagine that AI may play a role in many different parts of the health care system in the near future. To limit the scope of this text, we will mainly consider the activity of diagnosis. Our definition of *diagnosis* will be the one proposed in Maitland, 2010, p. 1, i.e., "the process of determining mechanisms by which the patient's health condition arises and the conclusions reached by doing so." Furthermore, we will use the following definition of *diagnostic error* which was proposed in National Academies of Sciences et al., 2015, Chapter 3:

> [diagnostic error is] the failure to (a) establish an accurate and timely explanation of the patient's health problem(s) or (b) communicate that explanation to the patient.

[1] The term *healthcare* includes both state owned health services and private actors.

[2] *AI* is short for *artificial intelligence*.

## A utilitarian approach to medical AI

There are many proponents of the use of AI in healthcare, and we will begin by presenting some counter-arguments to our claim. It is well-known that AI systems are already capable of achieving better performance than humans on certain tasks. Given the current momentum of AI research and innovation, we can assume that AI will beat us in more and more areas in the years coming. A computer will never have a bad day or feel tired, which will prevent many cognitive mistakes that humans are prone to make. There will be less strain on doctors and other health care professionals as AI systems become responsible for more tasks traditionally delegated to humans. This will enable human experts to do a better job at the task which AI cannot manage (yet). If we take as premise that the ideal goal is to minimize the number of diagnostic errors. It then follows from the above arguments that we should welcome AI technology into healthcare with open arms. Our premise can be directly motivated by the severe consequences of making such errors. As Raffel and Ranji put it: “Diagnostic errors are common and important causes of preventable morbidity and mortality in a variety of medical settings.” Or, as stated in National Academies of Sciences et al., 2015, “The potential harm from diagnostic errors could range from no harm to significant harm, including morbidity or death. Errors can be harmful because they can prevent or delay appropriate treatment, lead to unnecessary or harmful treatment, or result in psychological or financial repercussions.” Furthermore, we will get more affordable diagnostics which in turn makes healthcare more accessible and less dependent on socioeconomic status. If we put too many restrictions on its use, we will only slow the progress towards more explainable and better suited AI models for medical use.

## Limitations of AI in healthcare

In this section we discuss some of the possible ethical issues arising with the use of AI in healthcare. We present several arguments demonstrating the arguably immature state of current regulations, the healthcare system itself and the present-day AI technology. In the previous section, we discussed the advantages of medical AI based on the premise of minimizing diagnostic errors being the ultimate goal. We now contrast this view by taking a step back to get a wider perspective.

### The lack of explainability

Explainability[3] of an AI model refers to the degree of which its output can be explained from its input, in a way us humans can understand. For example, if an AI model informs us that a patient has skin cancer, can we actually understand how the model came to this conclusion? Some classical model architectures such as, for example, decision trees and particular regression models enjoys explainability to a certain degree. However, most of the state-of-the-art AI models today rely on deep learning using artificial neural networks and as a consequence of this, exhibit a very low degree of explainability. Such models are often referred to as black-box models for this very reason. We will address two possible issues arising from the use of non-explainable AI in healthcare. Namely, the inability to properly learn from failures, and the unfairness that can potentially be hidden in such systems.

[3] *Explainability* is also sometimes called *interpretability*.

### Learning from mistakes

The use of current AI technology in healthcare causes us to lose the ability to learn from our mistakes. In the case where a human expert make a diagnostic error, we can often trace the error back to its origin. For example, if a physician diagnose a patient based on certain medical findings and the diagnosis turns out to be wrong, we can start by asking the physician for their reasoning. The cause might be a cognitive mistake by the physician, a technical issue with the equipment used, a communication error or any number of other reasons. The point is that we can learn from the mistakes by backtracing the reasoning along the chain of explainability. If a non-explainable AI system is introduced, we break this chain, and we can no longer learn from our mistakes to the same degree as before.

### Fairness

It is difficult to ensure equally good performance of AI systems across different patient groups. In Saltelli et al., 2020, p. 483 the authors claim that "Results from models will at least partly reflect the interests, disciplinary orientations and biases of the developers." The first thing that springs to mind when talking about biased models is often the presence of vested interests. This is a relevant topic in itself, but we will rather be interested in a more hidden form of bias present in current AI technology, namely, the lack of fairness.[4] Let us consider a simple example demonstrating what we mean by fairness, or the lack thereof: Suppose we have an AI system designed for detecting a variety of diseases based on patient data. If the model was trained on data where patients of lower socioeconomic status were under-represented, then the model will likely perform better with respect to some patients and diseases than others. For example, the model may fail to be accurate on diseases influenced by poor nutrition and or unhealthy lifestyle. Such bias is also a problem already present in human-driven healthcare (see, for example, Smith et al., 1990, or Adler and Newman, 2002), but the lack of explainability in AI makes it extremely challenging to both measure and improve the inequalities. Therefore, without achieving a significantly higher level of explainability, we can not guarantee, nor even improve, the equality of quality of health services across different patient groups.

## The problem of accountability

The regulatory framework is not ready to accommodate the use of AI systems in healthcare yet. No system is completely fail-proof, and consequently it is only a matter of time until erroneous decisions will be made by an AI based system. If a doctor makes a diagnostic error, there are protocols in place to help us identify the underlying reasons so that we can learn from our mistakes and take preventive measures. Furthermore, there are laws governing the distribution of responsibility and accountability when errors happen. Who is accountable when an AI system makes an error? Certainly, we cannot hold the algorithm itself accountable. Or as Saltelli et al. put it: "... models tend to be developed with large teams and use such complex feedback loops that no one can be held accountable if the predictions are catastrophically wrong."(Saltelli et al., 2020, p. 483). Should the doctor using the system, or the company providing the system be held accountable? What about the people who approved of its use in the specific hospital? Politicians? Maybe we need to blame all AI researchers for not warning about

[4] The term *fairness* in the context of AI has multiple interpretations. We refer the reader to John-Mathews et al., 2022 where a comprehensive review of the term, extending beyond examples, is given.

the risks earlier. The point being, it is very difficult to give a meaningful answer to this question of accountability when an AI system is taking part in the process. Since there are so many actors involved, and the main actor (the AI system) cannot be held accountable, there is a high risk that accountability vaporizes. Accountability is not only important because of insurance reasons, but it is also fundamental for people's trust in healthcare systems. For example, in the absence of accountability, there is no basis on which licences and authorizations can be revoked. Because of the importance of accountability in healthcare, our society must be prepared to deal with accountability *before* AI systems are widely brought into use in the healthcare sector.

## The risk of AI dependence

Even if we do not consider fully autonomous AI systems, there is a risk that medical professionals using AI based tools will start relying too much on such tools. If the tools are highly accurate in their predictions over a long period of time, why should the expert bother to scrutinize the output of the algorithm? This may become a slippery slope where we in practice end up with a higher degree of AI autonomy[5] than we asked for. This can lead to health professionals trusting the algorithms too much and becoming less critical in their work. We cite a relevant point made by Saltelli et al.: "Once a number takes centre-stage with a crisp narrative, other possible explanations and estimates can disappear from view." (Saltelli et al., 2020, p. 484).

## Other issues not discussed

There are various other issues related to the use of AI in healthcare that could have been included in this text but were omitted due to length limitations. These include, but are not limited to:

- **Privacy concerns**: medical data usually contains sensitive information about patients.
- **Overdiagnosis**: low-cost diagnostics as a consequence of AI technology can cause overdiagnosis which can have harmful consequences on its own: "The main consequence of overdiagnosis is overtreatment. Treating an overdiagnosed condition bears no benefit but can cause harms and generates costs. Overtreatment also diverts health professionals from caring for those most severely ill." (Bulliard and Chiolero, 2015, p. 1).
- **Proprietarization of medical knowledge**: medical AI systems will likely be the property of private companies.
- **Loss of expert knowledge**: knowledge present in the medical peer community may deteriorate if AI systems replace human experts. This can also hinder further development of such systems since they depend on high quality training data.

[5] There exists different levels of autonomy for AI systems. Ranging from being used merely as a tool by a human expert to full autonomy. In the context of veterinary radiography, Cohen and Gordon present a detailed categorization of different levels of automations.

## Final discussion

In this essay we have looked at arguments both advocating for and against the use of AI in healthcare. We will now give the answer to our original question whether healthcare is ready to adopt AI technology at the present time. The counter-argument we considered based itself upon the premise that minimizing diagnostic errors is the ultimate goal. It would however be naïve to think that the relationship between diagnostic errors and patient harm exists in a vacuum, unaffected by other factors. Therefore, we are obliged to examine other possible consequences of the use of medical AI such as those arising from the problems of explainability and accountability that we discussed in this text. Based on the EEA's working definition of the precautionary principle as given in Gee, 2013[6], we should absolutely tread carefully with respect to the introduction of medical AI in society. It is clear from our discussion that careless use of medical AI technology can potentially pose a serious threat to health. Moreover, it is very likely that the ratio of knowledge to ignorance (a term used by Gee) is low, prescribing the use of the precaution. It is not necessarily the case that precaution slows progress. By acknowledging the limitations of the technology, we can put more effort into solving these issues by focusing research where it is dearly needed. Or as Gee puts it: "... [society could] use the precautionary principle, to anticipate and minimise many future hazards, whilst stimulating innovation." (Gee, 2013, p. 643). Saltelli et al. propose five principles to help ensure satisfactory quality of models to be used in society. One of these principles is to acknowledge ignorance: "... communicating what is not known is at least as important as communicating what is known. Yet models can hide ignorance." (Saltelli et al., 2020, p. 484). Furthermore, the authors also stress transparency and open discussion: "Models' assumptions and limitations must be appraised openly and honestly." (Saltelli et al., 2020, p. 484).

## Conclusion

Based on the above discussion, we conclude that great care must be taken with respect to the adoption of AI in healthcare. Hasted adoption of medical AI poses a serious threat to health. Despite its advantages, we must assume that many consequences of such technologies still remains unknown. More AI research, especially related to the problems presented in this text, is necessary, but not sufficient in itself. Both regulators and healthcare specialists needs time to adapt before medical AI can be allowed to play a major role in healthcare. Specialized ethical and practical guidelines covering the vast array of use cases must be easily accessible to everyone involved. Moreover, there is a crucial need for impact assessment, transparency and continuous open cross-disciplinary discussions. To get the time and resources required for continued research we are dependent on the public's trust in AI research. Adopting the technology prematurely, can paradoxically harm the progress of AI research caused by the loss of public trust. Or as Saltelli et al. put it: "Opacity about uncertainty damages trust ... Full explanations are crucial." (Saltelli et al., 2020, p. 484). On a more optimistic note, we should not forget that achieving explainable AI can possibly solve some of the problems related to fairness and accountability. If we can understand the exact reasons behind biases and diagnostic errors made, we can also continuously improve the models over time.

[6] *The precautionary principle* provides justification for public policy and other actions in situations of scientific complexity, uncertainty and ignorance, where there may be a need to act in order to avoid, or reduce, potentially serious or irreversible threats to health and/or the environment, using an appropriate strength of scientific evidence, and taking into account the pros and cons of action and inaction and their distribution. (Gee, 2013, p. 649).

# References

Adler, N. E., & Newman, K. (2002). Socioeconomic disparities in health: Pathways and policies. *Health affairs*, *21*(2), 60–76.

Bulliard, J.-L., & Chiolero, A. (2015). Screening and overdiagnosis: Public health implications. *Public health reviews*, *36*(1), 1–5.

Cohen, E. B., & Gordon, I. K. (2022). First, do no harm. ethical and legal issues of artificial intelligence and machine learning in veterinary radiology and radiation oncology. *Veterinary Radiology & Ultrasound*, *63*(S1), 840–850.

Gee, D. (2013). 27 more or less precaution? *Late lessons from early warnings: science, precaution, innovation*, 37.

Helsedirektoratet. (2022, March). Etiske anbefalinger ved bruk av kunstig intelligens. Retrieved October 16, 2023, from https://www.helsedirektoratet.no/tema/kunstig-intelligens/etikk/etiske-anbefalinger-ved-bruk-av-kunstig-intelligens

John-Mathews, J.-M., Cardon, D., & Balagué, C. (2022). From reality to world. a critical perspective on ai fairness. *Journal of Business Ethics*, *178*(4), 945–959.

Maitland, M. E. (2010). A transdisciplinary definition of diagnosis. *Journal of allied health*, *39*(4), 306–313.

National Academies of Sciences, E., Medicine, I., Services, B., Care, C., Ball, J., Miller, B., & Balogh, E. (2015). *Improving diagnosis in health care*. National Academies Press.

Raffel, K., & Ranji, S. (2023, September). Retrieved October 12, 2023, from https://www.uptodate.com/contents/diagnostic-errors

Saltelli, A., Bammer, G., Bruno, I., Charters, E., Di Fiore, M., Didier, E., Nelson Espeland, W., Kay, J., Lo Piano, S., Mayo, D., Pielke, R., Jr, Portaluri, T., Porter, T. M., Puy, A., Rafols, I., Ravetz, J. R., Reinert, E., Sarewitz, D., Stark, P. B., Stirling, A., van der Sluijs, J., & Vineis, P. (2020). Five ways to ensure that models serve society: A manifesto. *Nature*, *582*(7813), 482–484.

Smith, G. D., Bartley, M., & Blane, D. (1990). The black report on socioeconomic inequalities in health 10 years on. *BMJ: British Medical Journal*, *301*(6748), 373.

Topdahl, R. C., Mullis, M. E., & Nøkling, A. (2023, September). Snart vil kunstig intelligens analysere kroppen din: – vi er for dårlig forberedt. Retrieved October 12, 2023, from https://www.nrk.no/rogaland/xl/snart-vil-kunstig-intelligens-analysere-kroppen-din_-_-vi-er-for-darlig-forberedt-1.16553955

Vestre Viken, K. (2022, September). Kunstig intelligens i røntgenavdelingen. Retrieved October 15, 2023, from https://vestreviken.no/om-oss/nyheter/kunstig-intelligens-i-rontgenavdelingen