ScienceDirect

Social Science & Medicine

Volume 260, September 2020, 113172

Review article

The ethics of AI in health care: A mapping review ☆

Jessica Morley [a][1], Caio C.V. Machado [a][1], Christopher Burr [a], Josh Cowls [a][b], Indra Joshi [d], Mariarosaria Taddeo [a][b][c], Luciano Floridi [a][b][c]

Show more

Share Cite

https://doi.org/10.1016/j.socscimed.2020.113172

Get rights and content

Highlights

- This article maps the ethics of artificial intelligence in healthcare.
- Ethical issues can be epistemic, normative, or related to traceability.
- Issues affect individuals, relationships, groups, institutions, sectors, societies.
- An agreed standard for ethical analysis is needed; split by issue and level.

Abstract

This article presents a mapping review of the literature concerning the ethics of artificial intelligence (AI) in health care. The goal of this review is to summarise current debates and identify open questions for future research. Five literature databases were searched to support the following research question: how can the primary ethical risks presented by AI-health be categorised, and what issues must policymakers, regulators and developers consider in order to be ‘ethically mindful? A series of screening stages were carried out—for example, removing articles that focused on digital health in general (e.g. data sharing, data access, data privacy, surveillance/nudging, consent, ownership of health data, evidence of efficacy)—yielding a total of 156 papers that were included in the review.

We find that ethical issues can be (a) epistemic, related to misguided, inconclusive or inscrutable evidence; (b) normative, related to unfair outcomes and transformative effectives; or (c) related to traceability. We further find that these ethical issues arise at six levels of abstraction: individual, interpersonal, group, institutional, and societal or sectoral. Finally, we outline a number of considerations for policymakers and regulators, mapping these to existing literature, and categorising each as epistemic, normative or traceability-related and at the relevant level of abstraction. Our goal is to inform policymakers, regulators and developers of what they must consider if they are to enable health and care systems to capitalise on the dual advantage of ethical AI; maximising the opportunities to cut costs, improve care, and improve the efficiency of health and care systems, whilst proactively avoiding the potential harms. We argue that if action is not swiftly taken in this regard, a new 'AI winter' could occur due to chilling effects related to a loss of public trust in the benefits of AI for health care.

Introduction

Healthcare systems across the globe are struggling with increasing costs and worsening outcomes (Topol, 2019). This presents those responsible for overseeing healthcare systems with a 'wicked problem', meaning that the problem has multiple causes, is hard to understand and define, and hence will have to be tackled from multiple different angles. Against this background, policymakers, politicians, clinical entrepreneurs and computer and data scientists increasingly argue that a key part of the solution will be Artificial Intelligence (AI), particularly Machine Learning (Chin-Yee and Upshur, 2019). The argument stems not from the belief that all healthcare needs will soon be taken care of by "robot doctors" (Chin-Yee and Upshur, 2019). Instead, the argument rests on the classic definition of AI as an umbrella term for a range of techniques that can be used to make machines complete tasks in a way that would be considered intelligent *were* they to be completed by a human. For example, as mapped by (Harerimana et al., 2018), decision tree techniques can be used to diagnose breast cancer tumours (Kuo et al., 2001); Support Vector Machine techniques can be used to classify genes (Brown et al., 2000) and diagnose Diabetes Mellitus (Barakat et al., 2010); ensemble learning methods can predict outcomes for cancer patients (Kourou et al., 2015); and neural networks can be used to diagnose stroke (Jiang et al., 2017). From this perspective, AI represents a growing *resource* of *interactive*, *autonomous*, and often *self-learning* (in the machine learning sense) *agency*, that can be used on demand (Floridi, 2019a, 2019b), presenting the opportunity for potentially transformative cooperation between machines and doctors (Bartoletti, 2019).

If harnessed effectively, such AI-clinician cooperation, where AI is used to provide comprehensive evidence-based clinical decision-support to the clinician (AI-Health), could offer great opportunities for the improvement of healthcare services and ultimately patients' health (Taddeo and Floridi, 2018) by significantly improving human clinical capabilities in diagnosis (Arieno et al., 2019; De Fauw et al., 2018; Kunapuli et al., 2018), drug discovery (Álvarez-Machancoses and Fernández-Martínez, 2019; Fleming, 2018), epidemiology (Hay et al., 2013), personalised medicine (Barton et al., 2019; Cowie et al., 2018; Dudley et al., 2015) and operational efficiency (Lu and Wang, 2019;

Nelson et al., 2019). However, as Ngiam and Khor (2019) stress, if these AI solutions are to be embedded in clinical practice, then a clear governance framework is needed to protect people from harm, including harm arising from unethical conduct. We use the term 'cooperation' here and suggest that AI will be chiefly used for clinical decision support. This differentiates from arguments often made by the popular press which suggest that AI will be used to 'replace' clinicians.

To support policymakers, the task of the following pages is to classify the ethical risks presented by AI-health, align these with specific questions that must be answered by policymakers, and provide example actions that could be taken by healthcare governing bodies to develop the requisite governance framework. The intention is to ensure that the ethical challenges raised by implementing AI in healthcare settings are tackled *proactively* (Char et al., 2018). We seek to do this because if the ethical risks are not tackled proactively, by encouraging AI-health policymakers, developers and regulators to be ethically mindful, there is a potential risk of incurring significant opportunity costs (Cookson, 2018). For instance, ethical mistakes or misunderstandings may lead to social rejection and/or distorted legislation and policies, which in turn cripple the acceptance and advancement of [the necessary] data science. Encouraging this kind of proactive ethical analysis is essential but also challenging because, although bioethical principles for clinical research and healthcare are well established, and issues related to privacy, effectiveness, accessibility and utility are clear (Nebeker et al., 2019), other issues are less obvious (Char et al., 2018). For example, AI processes may lack transparency, making accountability problematic, or may be biased, leading to unfair, discriminatory behaviour or mistaken decisions (Mittelstadt et al., 2016). Identification of these less obvious concerns requires input from the medical sciences, economics, computer sciences, social sciences, law, and policy-making. Yet, research in these areas is currently happening in siloes, is overly focused on individual level impacts (Morley and Floridi, 2020a), or does not consider the fact that the ethical concerns may vary depending on the stage of the algorithm development pipeline (Morley et al., 2019).

Whilst AI-Health remains in the early stages of development and relatively far away from having a major impact on frontline clinical care (Panch et al., 2019), there is still time to develop this framework. However, this window of opportunity is closing fast, as the pace at which AI-Health solutions are gaining approval for use in clinical care in the US is accelerating (Topol, 2019). Both the Chinese (Zhang et al., 2018) and British governments (Department of Health and Social Care, 2019) have made it very clear that they intend on investing heavily in the spread and adoption of AI-Health technologies. It is for these reasons that the goal of this article is to offer a cross-disciplinary mapping review of the potential ethical implications of the development of AI-Health in order to support policy discussion, which will in turn orient the development of better design practices, and transparent and accountable deployment strategies. We will do this in terms of digital ethics. That is, we will focus on the evaluation of moral problems related to data, algorithms and corresponding practices (Floridi and Taddeo, 2016), with the hope of enabling governments and healthcare systems looking to adopt AI-Health to be ethically mindful (Floridi, 2019a). Specifically, the research question is: how can the primary ethical risks presented by AI-health be categorised, and what issues must policymakers, regulators and developers consider in order to be ethically mindful?

Section snippets

Methodology

A mapping review methodology (Grant; Booth, 2009) was used to find literature from across disciplinary boundaries that highlighted ethical issues *unique* to the use of AI algorithms in healthcare. This type of review is used to map and categorise existing literature on a particular topic (in this case the ethics of AI) and contextualise the findings within broader literature. The mapping review methodology was developed by the Evidence for Policy and Practice Information and Co-ordinating Centre …

Findings

What follows is a detailed discussion of the issues uncovered. A total of 223 titles were selected, duplicates were removed and, as reading commenced, relevant bibliography references were also added, resulting in approximately 147 papers to be read and included in the review. The flowchart in the appendix illustrates our methodology. Also, a summary map of our findings (Table 2) is provided at the end of the section. …

The need for an ethically-mindful and proportionate approach

The literature surveyed in this review clearly indicates the need for an agreed standard for AI-Health ethical evaluation. While these issues are all connected, they cannot be treated under the blanket discussion of "Ethics of AI" when discussing specific recommendations and solutions. For example, handling privacy at the individual LoA, considering design issues, is different from handling privacy at a group level, where the concern is raised from the ways in which the aggregate data is …

Conclusion

This thematic literature review has sought to map out the ethical issues around the incorporation of data-driven AI technologies into healthcare provision and public health systems. In order to make this overview more useful, the relevant topics have been organised into themes and six different levels of abstraction (LoAs) have been highlighted. The hope is that by encouraging a discussion of the ethical implications of AI-Health at individual, interpersonal, group, institutional and societal …

Funding

Taddeo's and Floridi's work was partially supported by Privacy and Trust Stream – Social lead of the PETRAS Internet of Things research hub – PETRAS is funded by the UK Engineering and Physical Sciences Research Council (EPSRC), grant agreement no. EP/N023013/1 Machado's, Taddeo's and Floridi's work was also partially supported by a Microsoft grant and a Google grant. …

Acknowledgements

We would like to thank the anonymous reviewers for their detailed comments which led to this much improved version of the paper. …

Recommended articles

References (149)

P. Balthazar *et al.*

Protecting your patients' interests in the era of big data, artificial intelligence, and predictive analytics

J. Am. Coll. Radiol. (2018)

C. Barton *et al.*

Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs

Comput. Biol. Med. (2019)

Y. Ding *et al.*

Identification of drug-side effect association via multiple information integration with centered kernel alignment

Neurocomputing (2019)

F. Greaves *et al.*

What is an appropriate level of evidence for a digital health intervention?

Lancet (2018)

H. Kim *et al.*

Health literacy in the eHealth era: a systematic review of the literature

Patient Educ. Counsel. (2017)

Edouard Kleinpeter

Four Ethical Issues of “E-Health”

IRBM (2017)

M. Kohli *et al.*

Ethics, artificial intelligence, and radiology

J. Am. Coll. Radiol. (2018)

K. Kourou *et al.*

Machine learning applications in cancer prognosis and prediction

Comput. Struct. Biotechnol. J. (2015)

F. López-Martínez *et al.*

A neural network approach to predict early neonatal sepsis

Comput. Electr. Eng. (2019)

Nicole Maher *et al.*

Passive data collection and use in healthcare: A systematic review of ethical issues

Int. J. Med. Inform. (2019)

View more references

Cited by (489)

Ethical framework for Artificial Intelligence and Digital technologies

2022, International Journal of Information Management

Show abstract

Fairness of artificial intelligence in healthcare: review and recommendations

2024, Japanese Journal of Radiology

Show abstract

Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility?

2022, Frontiers in Surgery

Show abstract

Artificial intelligence for good health: a scoping review of the ethics literature

2021, BMC Medical Ethics

Show abstract

Patient apprehensions about the use of artificial intelligence in healthcare

2021, Npj Digital Medicine

Show abstract

The Role of Artificial Intelligence in Early Cancer Diagnosis

2022, Cancers

View all citing articles on Scopus

[☆] b Traceability is introduced as an overarching ethical concern by Mittelstadt et al. (2016). It is used to summarise concerns that arise from the fact that potential algorithmic harms result from the actions of multiple actors. This makes it hard to find the 'cause' of the harm and hard to identify who should be held responsible and/or accountable for the harm caused. It is an overarching concern as it encompasses ethical risks that are both normative and epistemic, and can be applied at any LoA.

[1] These authors contributed equally to the writing of this paper.

View full text