Is Healthcare AI Research Engaging Publics In Conversations About Ethics? Protocol For A Scoping Review

Emma Kellie Frost (emmaf@uow.edu.au) University of Wollongong

Rebecca Bosward University of Wollongong

Yves Saint James Aquino University of Wollongong

Annette Braunack-Mayer University of Wollongong

Stacy M Carter University of Wollongong

Protocol

Keywords: Artificial Intelligence, AI ethics, healthcare AI, patients, publics

Posted Date: December 20th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1109170/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

**Background:** In recent years, innovations in artificial intelligence (AI) have led to the development of new healthcare AI (HCAI) technologies. Whilst some of these technologies show promise for improving the patient experience, ethicists have warned that AI can introduce and exacerbate harms and wrongs in healthcare. It is important that HCAI reflects the values that are important to people. However, involving patients and publics in substantive conversations about AI ethics remains challenging due to relatively limited awareness of HCAI technologies. This scoping review aims to map how the existing literature on publics’ attitudes toward HCAI addresses key issues in AI ethics and governance.

**Methods:** We developed a search query to conduct a comprehensive search of PubMed, Scopus, Web of Science, CINAHL, and Academic Search Complete from January 2010 onwards. We will include primary research studies which document publics’ or patients’ attitudes toward HCAI. A coding framework has been designed and will be used capture qualitative and quantitative data from the articles. Two reviewers will code a proportion of the included articles and any discrepancies will be discussed amongst the team, with changes made to the coding framework accordingly. Final results will be reported quantitatively and qualitatively, examining how each AI ethics issue has been addressed by the included studies.

**Discussion:** If HCAI is to be implemented ethically and legitimately, publics and patients must be included in important conversations about HCAI ethics. This review will explore how ethical issues are addressed in literature examining publics and patients’ attitudes toward HCAI. We aim to describe how publics and patients have been successfully consulted on HCAI ethics, and to identify any areas of HCAI ethics where more work is needed to include publics and patients in research and discussions.

Background

Recent years have seen the development and introduction of a number of artificial intelligence (AI) enabled technologies for healthcare. AI is a term which encompasses diverse computational technologies, making it challenging to define: prominent definitions include that AI is ‘a collection of interrelated technologies used to solve problems that would otherwise require human cognition’ (1), or that AIs are technologies with the ability to ‘perform tasks to achieve defined objectives without explicit guidance from a human being’ (2). Broad in application, AI technologies arrive with optimistic promises of transforming the patient experience. So far, advocates of healthcare AI (HCAI) have promised the technology will improve the accuracy of screening and diagnosis (3), increase the availability of care in remote regions (4), and free up physicians’ time so that they can engage more with patients (5).

Alongside innovations in HCAI, there is also a growing field of AI ethics which cautions against uncritical implementation of HCAI and raises questions about its regulation (6). AI technologies pose new risks and challenges to healthcare: AI has been shown to produce biased outcomes (7), many AI technologies are ‘black boxes’ where the reasons behind an algorithm’s output cannot be interpreted (8), and questions remain about how existing liability structures in medicine will effectively manage errors made by HCAI.
AI development also continues to be dominated by large private companies that have been criticised for failing to engage in meaningful conversations about the ethics of their products and research (10). Publics may be both beneficiaries of new HCAI technologies, and the greatest sufferers of AI-related harms (11). Patients and publics are important voices in developing effective and ethical AI governance, but engaging patients and publics in conversations about ethical HCAI is challenging. Most people have no firsthand experience with HCAI, and some are unfamiliar with the concept of AI in general (12). Publics may have limited understanding of how HCAI may be implemented, and limited knowledge about the potential wrongs and harms that could arise from implementing HCAI. To ensure that HCAI has a positive impact on patients, it is crucial that AI ethics reflects the values that are important to people (11,13), but it remains unclear how this should be achieved.

The aim of this review is to determine how common and emerging themes in HCAI ethics are addressed by the existing literature on publics’ and patients’ attitudes toward HCAI. We aim to explore how studies seeking publics’ and patients’ perspectives on HCAI have involved participants in conversations about AI ethics concepts such as explainability, bias, professional liability, and privacy.

**Methods/design**

Scoping reviews are an effective method for exploring the range and extent of literature on a given topic (14). Our work will follow the framework proposed by Arksey and O’Malley (15) with modifications from Levac and colleagues (14). Their six recommended steps include (a) identifying the research question; (b) identifying relevant studies; (c) study selection; (d) charting the data; (e) collating, summarising, and reporting the results and (f) consultation. The following sections will address each of these steps in greater detail. In preparing this protocol, we completed a PRISMA-P checklist to ensure all necessary details have been reported (Additional File 1).

**Stage 1: Identifying the research questions**

Our review will address the question, to what extent, and how, are HCAI ethics issues addressed in the existing literature on publics’ attitudes toward HCAI. Our objectives are (1) to determine the extent to which studies on publics’ perspectives on HCAI are designed to collect opinions or facilitate discussion about AI ethics, (2) to explore how existing research has been designed to facilitate engagement with publics about HCAI ethics, and (3) to explore study participants’ perspectives on HCAI ethics issues.

**Stage 2: Identifying relevant literature**

We developed a search query using the Population-Intervention-Context-Outcome (PICO) format. An initial search on Google Scholar helped to identify similar terms which were used to develop a comprehensive search query (Table 1).
We will use a systematic search strategy to find relevant articles for inclusion in the study. Databases to be searched are PubMed, Scopus, Web of Science, CINAHL, and Academic Search Complete. We will examine the reference lists of included studies to find any publications that were missed in the initial search. All studies collected through the search project will be imported into a Zotero library.

**Stage 3: Study selection**

After the search is completed, all studies will be screened for eligibility. EF will be responsible for conducting the search and managing the data. First, duplicates will be removed using the deduplication module from the Systematic Review Assistant (16), then the remaining files will be exported to MS Excel for the screening process. MS Excel will allow reviewers to easily indicate a study’s inclusion or exclusion, as well as keep notes about any uncertainties for discussion.

The first stage of screening will exclude irrelevant articles based on their title and abstract. Screening will be conducted based on a set of criteria defined below. The first reviewer (EF) will screen the first 10% of articles, including all articles that are potentially relevant based on their title and abstract, and excluding articles that are clearly irrelevant. Of this 10%, EF will construct a sample of approximately 40 articles marked for inclusion, and 60 articles marked for exclusion. A second reviewer (RB) will screen this sample of 100 articles and compare results with EF. Results will be discussed with the team and inclusion criteria will be modified if necessary. Once any issues have been resolved, EF will conduct the initial screening on the remainder of the studies.

After initial screening is completed, excluded articles will be removed and full article texts will be collected for the remaining studies. The two-reviewer screening process will be repeated on a random sample of 10% of the full texts. Differences will be discussed and resolved, and modifications will be made to the
inclusion criteria accordingly. Inter-rater scores will be generated to quantify agreeance between reviewers. Once the inclusion criteria are finalised, EF will conduct the remainder of the full-text screening.

**Inclusion criteria**

Articles will be screened against a set of inclusion criteria developed by the team. These criteria may be modified throughout the screening process. Initial design of the criteria was guided by the JBI guidelines for scoping reviews (17). This guide states that inclusion criteria should address (i) the types of participants, (ii) concept, (iii) context, and (iv) types of evidence sources.

**Types of participants.** Studies will be included if research participants are recruited as publics, patients (or their unpaid/familial carers), or healthcare consumers. If studies recruit professionals (e.g. physicians, nurses, policymakers, or professional carers) along with publics, they will be included so long as the data related to patients/publics can be extracted.

**Concept.** Studies will be included if the research addresses AI in patient- or (general) public-facing health care or services. An included study may address AI in healthcare or services in a specific field (e.g., patients’ perspectives on AI for breast screening, or publics’ attitudes toward AI-enabled mobile phone apps for skin cancer detection). An included study may also address specific applications or sub-fields of AI (e.g., care robots, machine learning, decision support systems, or chatbots).

Studies will be excluded if they only address AI in non-patient/public-facing health applications. For example, studies addressing AI used only to manage bills and claims processing in hospitals would be excluded. Studies will be excluded if they only address non-health applications of AI.

**Context.** Studies from any geographical location will be included, so long as the manuscript can be assessed in English. Only studies published between 1 January 2010 and 15 September 2021 will be included. This time period has seen the introduction of modern approaches such as deep learning, convolutional neural networks, and natural language processing into HCAI research (5). These new approaches are the source of much of the current interest and investment in HCAI, and introduce a number of new potential challenges and harms (6).

**Types of evidence sources.** Only primary research studies will be included in this review. Studies will not be excluded based on method. There will be no restrictions on study design.

Studies will be excluded if they are only available in a language other than English, if they do not address AI in a patient-facing healthcare context, or if the study participant profile does not include patients or publics.

**Stage 4: Charting the data**

We have designed a coding framework to capture information on whether and how studies address a series of AI ethics concerns. This has been primarily based on Fjeld and colleagues’ (18) analysis of a
series of AI ethics guidelines. Fjeld et al. identified seven domains that were frequently addressed in AI ethics frameworks: (1) privacy, (2) accountability, (3) safety and security, (4) transparency and explainability, (5) fairness and non-discrimination, (6) human control over technology, and (7) professional responsibility. To capture more detailed data on where each of these ethical issues were addressed, we separated the concepts of ‘safety’ and ‘security’, and ‘transparency’ and ‘explainability’ into individual code categories (Table 2).
Table 2
- adaptation of AI ethics frameworks for data extraction

| Concept                        | Reference(s) | Description                                                                                       |
|-------------------------------|--------------|--------------------------------------------------------------------------------------------------|
| Privacy                       | (18)         | Whether study addresses publics’ views on privacy, consent, control over the use of data, and/or right to erasure. |
| Accountability                | (18)         | Whether study addresses publics’ views on legal liability and responsibility for rectification when algorithms perform poorly |
| Safety                        | (18)         | Whether study addresses publics’ views on the consistency and accuracy of algorithms’ performance, or the perceived safety of using AI in healthcare and services. |
| Security                      | (18)         | Whether study addresses publics’ views on algorithms’ vulnerability to nefarious third parties.     |
| Transparency                  | (18)         | Whether study addresses publics’ views on the transparency of AI development and implementation, and/or the importance of disclosing that AI is being used. |
| Explainability                | (18)         | Whether study addresses publics’ views on algorithmic explainability, black box algorithms, and/or the importance of patients’ and physicians’ ability to understand the reasons behind an algorithm’s decision. |
| Fairness and non-discrimination | (18)       | Whether study addresses publics’ views on algorithmic bias, fairness in algorithmic decision-making, and/or inclusivity in AI design. |
| Human control over technology | (18)         | Whether study addresses publics’ perspectives on the extent to which humans should review automated decisions, and whether people should be able to opt out of algorithm-informed decisions. |
| Professional responsibility   | (18)         | Whether study addresses publics’ perspectives on professionals’ roles in ensuring that algorithms are accurate, perform well, and do not cause harms. |
| Power                         | (13,19)      | Whether study addresses publics’ perspectives on the impact of AI on existing power structures in society. E.g. concerns about AI reinforcing existing power structures, the inclusivity of AI governance and regulation, and/or the development of AI technologies which primarily benefit the Global North. |
| Environmental wellbeing       | (20)         | Whether study addresses publics’ views on the environmental impacts of AI, including e-waste, energy consumption, and materials. |
| Societal wellbeing            | (20)         | Whether study addresses publics’ views on whether algorithms are being created and implemented for broader social good. |
| Ethical governance            | (21)         | Whether study addresses publics’ views on the suitability of existing legal structures, or the need for new structures, to manage the ethical issues associated with HCAI. |

We added four additional concepts to the framework. The first, *power*, has become a more common point of discussion in AI ethics frameworks recently, to assess how AI development and governance structures...
are reinforcing existing power dynamics and failing to redistribute power to marginalised groups (13,19). The second, *environmental wellbeing*, addresses the environmental impacts of AI development including energy usage, materials, and e-waste (19,20). *Societal wellbeing*, addresses whether technological development is being implemented for social good (20). Finally, *ethical governance* addresses whether existing governance structures are suitable to manage the ethical issues associated with HCAI.

We will use MS Excel to chart the data from the studies. The initial data extraction tool (Additional File 2) covers the key areas recommended by Arksey and O’Malley (15). In adhering to recommendations from Levac et al (14), we will modify this tool progressively throughout the data collection process. Initially, a random 10% of the included studies will be selected and coded by two coders (EF & RB) and any differences will be resolved in consultation with the research team. We will make changes to the data extraction tool if necessary. The remainder of the charting will be conducted by EF.

**Stage 5: Collating, summarising, and reporting the results**

We will collate results into tabular format for analysis. Guided by Arksey and O’Malley’s (15) recommendations, analysis will begin with descriptive quantitative reporting where it is appropriate (e.g. the number of studies which address each HCAI ethics issue in the framework). Our final reporting will synthesise publics and patients’ perspectives on each HCAI ethics issue in Table 2, and discuss the extent to which the issue is addressed in the literature.

**Discussion**

This review may have some limitations. Firstly, scoping reviews are designed to map the literature in a topic, and are not designed to assess the quality of included studies (14). The quality of the studies included in this review will not be systematically assessed. Secondly, it is possible that relevant studies will not be captured by the search strategy defined in this protocol. We will conduct a systematic pearling process on relevant identified studies to ensure as many relevant articles are identified as possible. Finally, findings will be limited to studies published in English, which may exclude relevant articles published in other languages. We will reflect on the impact of these limitations, as well as discuss any other arising limitations, in the reporting of our results.

The more widespread use of HCAI technologies is often described as an inevitable (5). However, implementation of HCAI may exacerbate certain harms in healthcare (6). Although patients and publics are likely to be the greatest sufferers of HCAI-related harms, involving patients and publics in conversations about AI ethics remains challenging (12).

To date, the extent to which patients and publics are involved in discussions about HCAI ethics is unclear. This review will examine how existing research has involved patients and publics in conversations about a series of HCAI ethics issues. In doing so, we will describe patients’ and publics’ perspectives on each HCAI ethics issue, and highlight potential gaps, or areas of HCAI ethics where research with patients and publics is limited. The results from this review will be important to understanding where further effort is
required to involve patients and publics in comprehensive conversations about HCAI ethics. Such an effort is crucial to ensuring that HCAI is implemented safely and effectively.

**List Of Abbreviations**

AI – Artificial Intelligence  
HCAI – Healthcare Artificial Intelligence  
PICO – Population-Intervention-Context-Outcome

**Declarations**

**Ethics approval and consent to participate**

Not applicable to this study.

**Consent for publication**

Not applicable to this study.

**Availability of data and materials**

No datasets were generated or used for this paper

**Competing interests**

The authors declare that they have no competing interests

**Funding**

No funding has been sought for this study

**Authors’ contributions**

EKF developed the protocol and drafted the manuscript. RB, YSJA, and ABM consulted on protocol design and reviewed and edited the manuscript. SMC oversaw development of the protocol and reviewed and edited the manuscript.

**Acknowledgements**

Not applicable.

**References**
1. Walsh T, Levy N, Bell G, Elliott A, Maclaurin J, Mareels I, et al. The effective and ethical development of artificial intelligence: an opportunity to improve our wellbeing. The Australian Council of Learned Academies; 2019.

2. Hajkowicz S, Karimi S, Wark T, Chen C, Evans M, Rens N, et al. Artificial Intelligence: Solving problems, growing the economy and improving our quality of life. CSIRO, Data61; 2019 p. 68.

3. Esteva A, Topol E. Can skin cancer diagnosis be transformed by AI? The Lancet. 2019 Nov 16;394(10211):1795.

4. Sechopoulos I, Mann RM. Stand-alone artificial intelligence - The future of breast cancer screening? The Breast. 2020 Feb 1;49:254–60.

5. Topol E. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. Illustrated edition. New York: Basic Books; 2019. 400 p.

6. Carter SM, Rogers W, Win KT, Frazer H, Richards B, Houssami N. The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. The Breast. 2020 Feb;49:25–32.

7. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019 Oct 25;366(6464):447–53.

8. Dilsizian ME, Siegel EL. Machine Meets Biology: a Primer on Artificial Intelligence in Cardiology and Cardiac Imaging. Curr Cardiol Rep. 2018 Oct 18;20(12):139.

9. Grote T, Berens P. On the ethics of algorithmic decision-making in healthcare. Journal of Medical Ethics. 2020 Mar 1;46(3):205–11.

10. Holzmeyer C. Beyond ‘AI for Social Good’ (AI4SG): social transformations—not tech-fixes—for health equity. Interdisciplinary Science Reviews. 2021 Apr 3;46(1–2):94–125.

11. Lai M-C, Brian M, Mamzer M-F. Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France. Journal of Translational Medicine. 2020 Jan 9;18(1):14.

12. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. The Lancet Digital Health. 2021 Sep 1;3(9):e599–611.

13. Hickok M. Lessons learned from AI ethics principles for future actions. AI Ethics. 2021 Feb 1;1(1):41–7.
14. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implementation Sci. 2010 Sep 20;5(1):69.

15. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology. 2005 Feb 1;8(1):19–32.

16. Rathbone J, Carter M, Hoffmann T, Glasziou P. Better duplicate detection for systematic reviewers: evaluation of Systematic Review Assistant-Deduplication Module. Syst Rev. 2015 Jan 14;4:6.

17. Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. JBI Evidence Implementation. 2015 Sep;13(3):141–6.

18. Fjeld J, Achten N, Hilligoss H, Nagy A, Srikumar M. Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. SSRN Journal [Internet]. 2020 [cited 2021 Sep 2]; Available from: https://www.ssrn.com/abstract=3518482

19. Crawford K. Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence [Internet]. Atlas of AI. Yale University Press; 2021 [cited 2021 Sep 15]. Available from: https://www.degruyter.com/document/doi/10.12987/9780300252392/html

20. European Commission. Ethics guidelines for trustworthy AI | Shaping Europe's digital future [Internet]. 2019 [cited 2021 Sep 2]. Available from: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

21. Guan J. Artificial Intelligence in Healthcare and Medicine: Promises, Ethical Challenges and Governance. Chinese Medical Sciences Journal. 2019 Jun 30;34:76–83.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile1.docx
- AdditionalFile2.docx