



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/183475/>

Version: Accepted Version

Article:

Siala, H. and Wang, Y. (2022) SHIFTing artificial intelligence to be responsible in healthcare: a systematic review. *Social Science & Medicine*, 296. 114782. ISSN: 0277-9536

<https://doi.org/10.1016/j.socscimed.2022.114782>

Article available under the terms of the CC-BY-NC-ND licence
(<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

SHIFTing Artificial Intelligence to be Responsible in Healthcare: A Systematic Review

Dr. Haytham Siala
Department of International Business and Marketing,
Newcastle University London
102 Middlesex Street, London E1 7EZ, United Kingdom
E-mail: haytham.siala@ncl.ac.uk
Tel: +44 7809 667128

Dr. Yichuan Wang
Sheffield University Management School
The University of Sheffield
Conduit Rd, Sheffield S10 1FL, United Kingdom
E-mail: yichuan.wang@sheffield.ac.uk
Tel: +44 114 215 7181

Abstract

A variety of ethical concerns about artificial intelligence (AI) implementation in healthcare have emerged as AI becomes increasingly applicable and technologically advanced. The last decade has witnessed significant endeavors in striking a balance between ethical considerations and health transformation led by AI. Despite a growing interest in AI ethics, implementing AI-related technologies and initiatives responsibly in healthcare settings remains a challenge. In response to this topical challenge, we reviewed 253 articles pertaining to AI ethics in healthcare published between 2000 and 2020, summarizing the coherent themes of responsible AI initiatives. A preferred reporting items for systematic review and meta-analysis (PRISMA) approach was employed to screen and select articles, and a hermeneutic approach was adopted to conduct systematic literature review. By synthesizing relevant knowledge from AI governance and ethics, we propose a responsible AI initiative framework that encompasses five core themes for AI solution developers, healthcare professionals, and policy makers. These themes are summarized in the acronym SHIFT: *Sustainability, Human centeredness, Inclusiveness, Fairness, and Transparency*. In addition, we unravel the key issues and challenges concerning responsible AI use in healthcare, and outline avenues for future research.

Keywords: Systematic literature review, Responsible artificial intelligence (AI), Health-medicine, AI ethics, digital health

1. Introduction

Artificial intelligence (AI), an algorithm-driven computing technology, is programmed to self-learn from data and make intelligent predictions and real-time decisions through the use of artificial neural networks, machine learning, robotic process automation, and data mining (Chen & Asch, 2017; Davenport & Kalakota, 2019). AI in the healthcare milieu is defined as the use of intelligent data-driven technologies that leverage healthcare resources and data more effectively to support and streamline decision-making in healthcare, and to consequently provide better healthcare services that are tailored to individual needs. In order to effectively inform the decision making in healthcare, AI technologies typically apply machine learning algorithms to perform ‘intelligent’ analytical and inferential activities on health data, which are used amongst other things, to detect and predict pandemics and disease (infodemiology), diagnose and manage chronic and neurological conditions, interpret medical scans and radiology images, deliver health services and treatments, drug discovery and matching suitable patients to clinical trials. In addition, AI has the potential to address societal challenges unique to global health (Mehta et al., 2020) and expedite the achievement of the sustainable development goals related to health and well-being (Vinuesa et al., 2020). However, there is increasing concern about AI’s effectiveness in health care due to a variety of ethical issues, including algorithmic bias leading to inconsistent results or discriminatory outcomes, privacy violations, conflicts over data ownership, and a lack of transparency in data use (Vayena et al., 2018).

As a result, Professor Stephen Hawking urged AI proponents to exercise caution, saying: “*Success in creating AI would be the biggest event in human history. Unfortunately, it might also be the last, unless we learn how to avoid the risks.*” The ethical issues can be exemplified by the following recent cases of unethical use of AI in healthcare:

- *IBM’s Watson supercomputer, touted as a revolutionary tool for cancer treatment, makes treatment recommendations based on a handful of hypothetical cancer cases, therefore yielding unsafe and inaccurate medical advice, and posing a health and safety threat to patients (Ross & Swetlitz, 2018). This case demonstrates the negative implications of ambiguous processes in data access and use, as well as algorithmic design and explanation.*
- *The UK Information Commissioner’s Office (ICO) notes that data from 1.6 million patients had been provided to Google’s DeepMind without properly informing patients or obtaining their consent (Powles & Hodson, 2017). In this case, patients’ autonomy has been compromised due to a lack of robustness and agility in the response of institutions and regulators to data policies.*
- *Algorithms using surrogate health status measures to predict future healthcare needs may perpetuate inaccuracies and disparities in care. For example, Black patients spend an average of \$1100-1800 less per year than White patients for a given level of disease burden. The algorithm interpreted Black patients’ healthcare needs as lower, when they needed as much coordinated healthcare as white patients. This algorithm fails to take into consideration some important factors*

including household income, insurance status, hospital access, and employment status. (Obermeyer et al. 2019). In this case, medical decisions can be problematic and potentially inequitable due to what is known as ‘label choice bias’.

To tackle the ethical dilemmas and concerns of AI, the last decade has witnessed several debates on striking a balance between ethical considerations and digital transformation (Culnan & Williams, 2009; Zhang & Hon, 2020). Newell and Marabelli (2015) called for researchers to study the impact of ‘irresponsible’ use of AI-powered analytics (to inform algorithmic decisions) on individuals, organizations, and societies. Two recent studies have responded to this call for research: Wright and Schultz (2018) have proposed an ethics AI framework where stakeholder theory and social contracts theory are integrated to develop a series of best practices such as acknowledging the transition, minimizing disruptions, and reducing social inequalities to address ethical challenges of AI. Later on, Flyverbom et al. (2019) have underscored the interplay between responsible business practices and digital transformations, suggesting that when using digital technologies such as big data analytics, businesses should consider expanding their remit beyond profit-making and should proactively commit to addressing societal challenges.

In health care, there has been a growing attention given to understand responsible approaches to the development, implementation, management, and governance of AI (Morley et al., 2020; Peters et al., 2020). This leads toward the concept of responsible AI, which is defined as “a methodology for the large-scale implementation of AI methods in real organizations with fairness, model explainability and accountability at its core” (Arrieta et al., 2020, p.82). Although there are several AI ethics frameworks and guidelines available (Jobin et al., 2019), they are highly abstract or not tailored towards healthcare (Morley et al., 2021). Moreover, existing reviews on responsible AI focus on exploring the main trends of AI for healthcare considering ethical aspects using bibliometric approach (Wamba & Queiroz, 2021), or identifying ethical concerns specific to the use of AI in healthcare using a thematic review approach (Morley et al., 2020; Trocin et al., 2021). Our review goes a step further by providing practical guidelines on how healthcare organizations can build an AI initiative that is responsible.

For healthcare organizations, building AI systems that are ethical and safe can be a daunting challenge, requiring significant investments in data governance, algorithm management, and a thorough evaluation of the social and environmental impacts of AI. The process of initiating responsible AI necessitates cross-disciplinary and cross-sectoral cooperation by data scientists, healthcare providers, and policy makers who share interests in establishing a sustainable AI ecosystem from which all stakeholders and society can benefit (Alami et al., 2020). A multitude of studies from the fields of social science, medicine, healthcare information systems and computer science have attempted to address these challenges. By reviewing the current literature on responsible AI in healthcare, we attempt to help researchers and practitioners understand what a responsible AI initiative actually entails and what challenges they may encounter when implementing responsible AI.

In summary, we seek to answer the research question: *How can AI-related technologies and initiatives be implemented in a responsible manner in healthcare?* To answer this research question, our review aims to systematically identify key themes of responsible AI initiatives and to enumerate the pressing challenges associated with responsible AI that have not yet been adequately addressed. Drawing on insights from our review, we propose avenues for future research on responsible AI in health care.

2. Theoretical Background

2.1. Understanding responsible AI from the perspective of virtue ethics theory

In recent years, there has been a growing discourse on AI ethics in healthcare AI research. Various ethical principles have been identified as applicable candidates for the design and development of AI systems (Floridi et al., 2018). However, much of the contemporary AI-driven research lacks the ethical, regulatory, and practical considerations for wide implementation (Schwalbe & Wahl, 2020) due to a lack of a unified framework for governing AI (Morley et al., 2020). Although AI ethical frameworks have undergone several revisions to reflect the complexity of AI ethical issues, they still provide little insight into what initiatives should be implemented to foster responsible use of AI (Jobin et al., 2019; Morley et al., 2021). There has been a recurrent call (Morley et al., 2021) in the literature to push for standardizing and defining an ethical framework for AI governance in healthcare that permits the deployment of responsible AI-based health systems.

A virtue ethics perspective can contribute to the development of responsible AI (Chun, 2005; Song & Kim, 2018). Virtue ethics theory emphasizes the virtue or moral character of the person performing an action in a given circumstance (what will a virtuous person do in a certain situation), rather than the appropriateness of an action or its consequences (Chatterjee et al., 2009; Audi, 2012). Virtue ethicists have provided different accounts of the virtues that define a virtuous person. In management practices, virtue ethics can guide managers in making better ethical decisions (Audi, 2012). In addition, virtue ethics can be applied to mitigate risks associated with using a service, guide a firm in its daily activities and operations, and “increase a firm’s reputation and moral standing in the society in which it operates” (Chakrabarty & Bass, 2015, p. 497). This notion can be extended to health service providers who are implementing AI. Thus, virtue ethics is an appropriate theoretical foundation for developing a responsible AI initiative framework in health care because prior research suggests that the presence of virtuous character traits (e.g., fairness and honesty) in an agent (person or organization) can positively impact the actions of that agent (Audi, 2012; Chun, 2005). Therefore, by conducting a systematic literature review, we not only observe healthcare organizations’ responsible AI actions, but also how those actions relate to AI ethical principles that are potentially underpinned by virtue ethics.

While there is no universally accepted ethical framework, we focus on six key ethical characteristics that emerged from AI ethics literature - *fairness*, *transparency*, *trustworthiness*, *accountability*, *privacy*, and *empathy* – that are deemed to be important and most cited in healthcare

AI research for developing our initial understanding of responsible AI (Blobel et al., 2020; Bukowski et al., 2020; Floridi et al., 2019; Reddy et al., 2020). We discuss these six ethical characteristics in the context of AI in healthcare by exploring how these ethical characteristics emerged in previous healthcare research (see Table 1):

Table 1. A summary of the ethical principles and their application to AI in healthcare

Ethical principles	Description	Application of ethical principles contextualized to AI in healthcare
Fairness	AI health systems must ensure access to health care is equitable so as not to contribute to health disparities or discrimination (Bukowski et al., 2020). AI models should be trained with appropriate and representative datasets to reduce biases, make accurate clinical predictions, and reduce discrimination (Reddy et al., 2020).	<ul style="list-style-type: none"> • <i>Governing bodies and healthcare institutions should develop normative standards for AI in healthcare. These standards should inform how AI models will be designed and deployed in the healthcare context and should conform to the requirements of one of the classic biomedical ethical principles, namely justice. AI design should ensure that procedural (fair process) and distributive justice (fair allocation of resources) is applied consistently, with a view to protect against adversarial attack or the introduction of biases or errors through self-learning or malicious intent (Bukowski et al., 2020).</i>
Transparency	Transparency has been frequently cited as a key challenge for acceptance, regulation, and deployment of AI in healthcare. It focuses on the ability to explain and verify the behaviors of AI algorithms and models (Blobel et al., 2020).	<ul style="list-style-type: none"> • <i>The importance of ensuring that AI health systems designed with human attributes (voice or visual) do not deceive humans; they should introduce themselves explicitly as AI agents. They must also allow patients the freedom to make health-related decisions without coercion or undue pressure (Reddy et al., 2020).</i> • <i>Transparency and explanations of clinical decisions are essential for medical imaging analysis and clinical risk prediction (Blobel et al., 2020)</i> • <i>Where patient data may be shared with AI developers, there must be a process to seek fully informed consent from patients and, if it is unfeasible/impractical to seek approval, data must be anonymized so that individual patient details cannot be recognized by the developers (O'Sullivan et al., 2019).</i> • <i>Institutional policies and guidelines are revamped to ensure patients are aware that the treating clinician is drawing support from AI applications, what the limitations of the applications are, and that the patients are in a position, where relevant, to refuse treatment involving AI (Reddy et al., 2020).</i>
Trustworthiness	To build trustworthiness in AI, both clinicians and patients need to be engaged. Clinicians must explain how AI works and what its advantages and limitations are, and patients need to be willing to accept AI and engage in AI-driven healthcare (Bukowski et al., 2020; Reddy et al., 2020).	<ul style="list-style-type: none"> • <i>To address lack of trust from a clinician and patient perspective, Reddy et al. (2020) proposed a governance model with a multipronged approach that includes technical education, health literacy, and clinical audits.</i> • <i>In relation to clinician's trust, AI applications need to be designed to respect the autonomy of patients (Bukowski et al., 2020).</i>
Accountability	Accountability refers to a system's ability to provide an explanation for its actions (O'Sullivan et al., 2019). The principle of accountability comprises safety where consideration is given so that the AI system will not cause harm or danger to its users and third parties (Blobel et al., 2020; Reddy et al., 2020).	<ul style="list-style-type: none"> • <i>US Food and Drug Administration (FDA), which regulates medication and medical devices, has introduced steps for software to be approved for medical use. This is called SaMD (Software as a Medical Device). A developer would have to submit the AI software to the FDA for review and approval when there is a significant medication that affects its safety or effectiveness (Reddy et al., 2020).</i>
Privacy	Privacy is about ensuring that the AI system will not infringe the privacy of users and third parties including internationally	<ul style="list-style-type: none"> • <i>Healthcare providers and health technology companies using AI must always follow current regulations, such as applicable data privacy provisions, informed consent, and ensure due</i>

	recognized human rights (Blobel et al., 2020).	<i>process (Xafis et al., 2019). To ensure that AI programs adhere to contemporary privacy regulations like GDPR, periodic judicial reviews of how and for how long user data is stored should be applied.</i>
Empathy	Empathy is the characteristic of someone who is compassionate, sympathetic, and considerate of the feelings of others (Chun, 2005), which leads to more confident, supportive, and caring relationships.	<ul style="list-style-type: none"> • <i>Participants preferred empathic AI carebots and messages over human doctors when it comes to communicating bad health news or unfavorable information to patients (Blease et al., 2019; Hoorn and Winter 2018).</i> • <i>The role of AI in medical consultations should be complementary rather than that of a surrogate (Wangmo et al., 2019). Additionally, it is important to understand and weave the element of 'humanness' into AI to provide better support for healthcare professionals and patients (Dalton-Brown, 2020).</i>

3. Methodology

To garner more profound understanding about responsible AI applications in healthcare, this paper adopted a systematic literature review (SLR) approach to address topical research question(s) by sifting the literature to identify, select, critically appraise, and collate findings of prior relevant studies. Unlike narrative literature reviews, the SLR approach is an evidence-based approach that involves employing a transparent, rigorous, and replicable literature searching and reviewing process (Cipriani & Geddes, 2003). The SLR aims at providing a profound understanding of the research topic by adopting a critical evaluation of the extant literature that is expatiating on the topic of interest (Galetsi et al., 2019). Our review followed the five steps suggested by Boell and Cecez-Kecmanovic (2014): (1) Search and acquisition; (2) Mapping and classifying; (3) Critical assessment; (4) Argument development; and (5) Research problem/questions. With this approach, we attempt to understand the research landscape of responsible AI use and applications surrounding the healthcare ecosystem to subsequently identify responsible AI initiatives and challenges from previous studies.

3.1. Search and acquisition

The first step of SLR is to identify databases, journals, keywords, and time frame. A preferred reporting items for systematic review and meta-analysis (PRISMA) approach was employed to help us screen and select articles, as illustrated in Figure 1.

We searched for research articles in two key digital bibliographical databases – PubMed and Google Scholar. Since AI is evolving at a fast pace, we have selected articles for our review that were published recently (i.e., between January 1, 2000, and December 31, 2020, including in-press articles). Using three sets of search keywords (see Figure 2 and Appendix A), we looked for relevant articles that have addressed AI ethical issues in health care or demonstrated ways to deploy AI in a responsible way. The first set of search keywords refer to AI related technologies applied in healthcare; thus we develop a list of keywords based on the scope of AI suggested by Toh et al. (2019), Morley et al. (2020), and Davenport & Kalakota (2019). The second set of search keywords referred to the terms pertaining to AI ethics; thus, we developed a list of search keywords based on the AI ethical principles outlined in the AI governance frameworks (Jobin et al., 2019; Reddy et al., 2020). As mentioned earlier, we discuss

the key and most cited ethical characteristics, which emerged from the AI ethics literature that are deemed to be important to develop and nurture our initial understanding of responsible AI. In addition to these most cited ethical principles, we also identify some other ethical principles that have emerged recently (e.g., explicability and sustainability). Different fields may use different words or terminology to describe ethical principles; some are used interchangeably. The aim of our keyword search strategy is to cover as many keywords related to AI ethics as we can, in order to include a wider range of articles for our review. A single search keyword namely, “health” was then used to refine and confine the search to studies in the healthcare context. Since we were trying to retrieve a wide range of articles relevant to health, we used a generic term (i.e. health) rather than specific terms such as “medicine” or “healthcare”. We used these keywords to generate 96 search strings ($C_1^8 * C_1^{12} * C_1^1 = 96$). In total, we reviewed 9,535 articles from PubMed database and 48,000 articles from Google Scholar. Due to the large number of results, only the 500 most relevant articles retrieved by Google Scholar were reviewed (Morley et al., 2020).

Next, we assessed the title and abstract of the articles and after applying the exclusion criteria (see Appendix B), 56,290 citations were precluded from the review process. For example, articles focusing explicitly on technical aspects of AI were eliminated from the review process. At the end of this stage, 1,245 articles remained. The remaining articles were then analyzed for their relevance to AI ethics and governance in healthcare. The relevance assessment criteria can be found in Appendix C. An additional reliability check was also adopted by scanning relevant articles and reaching a consensus on suitability and relevancy of each article. Studies that did not attract a unanimous agreement from the researchers were discarded (Siddaway et al., 2019). Appendix D provides a complete list of the articles in our review.

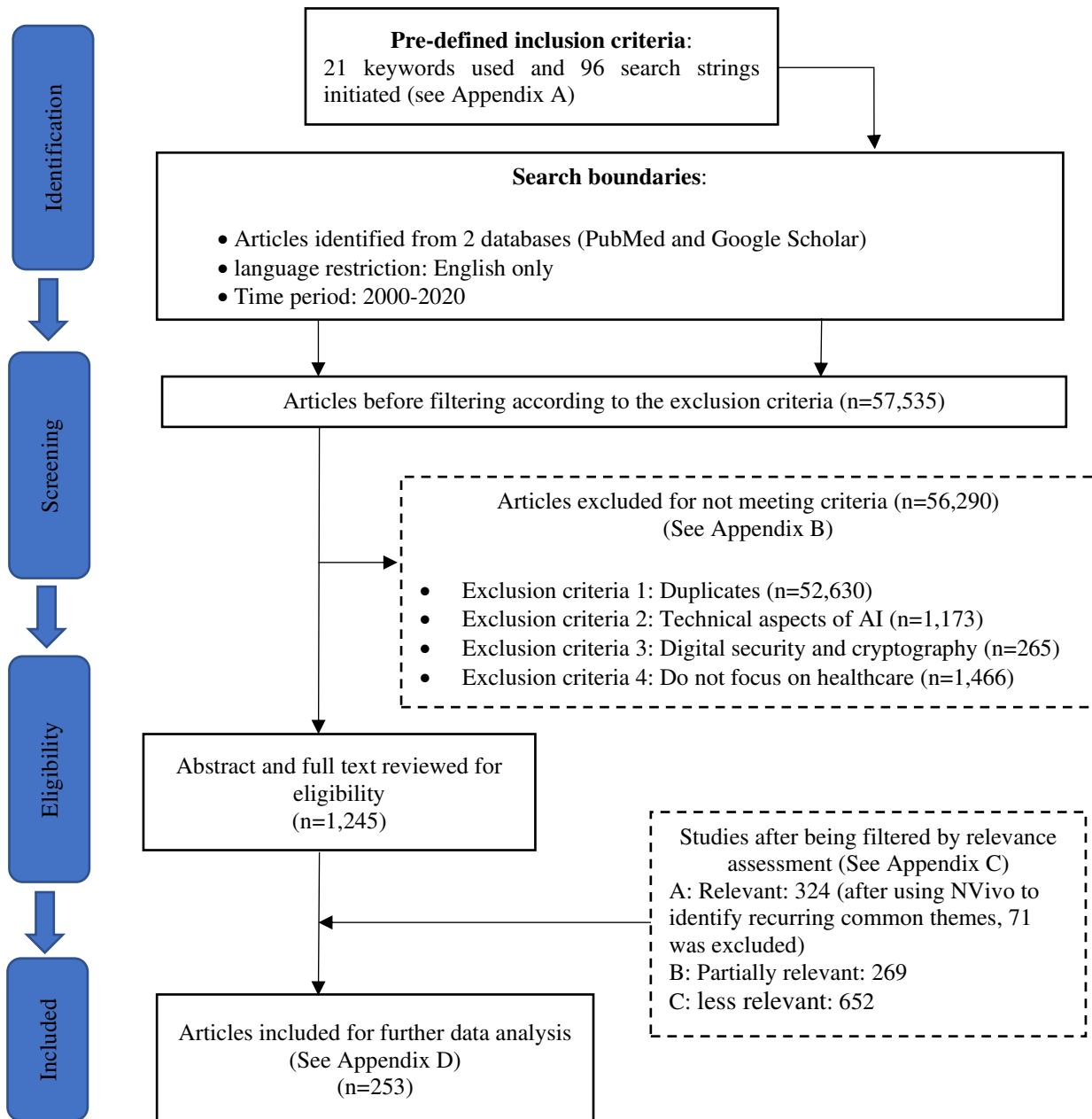


Figure 1. Mapping and classifying process

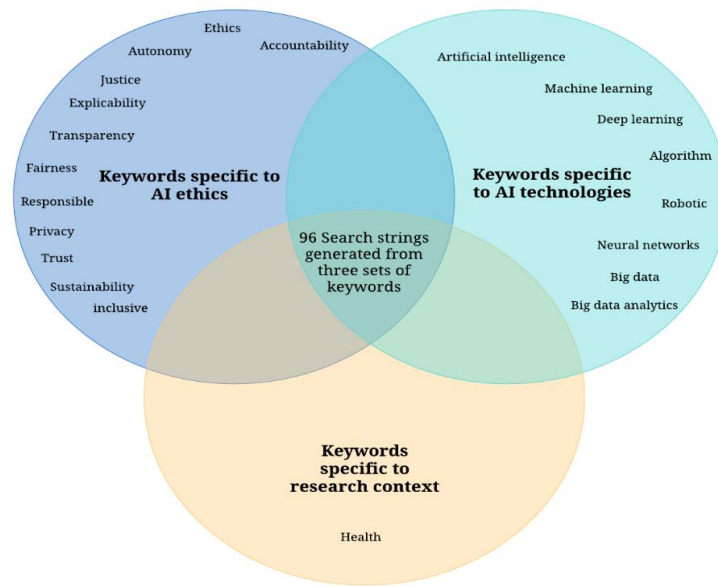


Figure 2. Keywords search strategy

3.2. Mapping and classifying

The mapping and classifying process aims to generate the classification schemes in a meaningful and visual manner. To gain a deeper understanding of the status of research in responsible AI in health care, we classified the 253 articles by adopting an inductive thematic analysis that involved using NVivo to identify the most recurring themes pertaining to responsible AI in health care. Inductive thematic analysis aims to identify the current contributions and knowledge gaps specific to research themes, which can inform future research direction by delineating a research agenda that addresses the current challenges of responsible AI in healthcare. The thematic analysis followed Clarke and Braun’s (2013) guidelines, which involves six phases: familiarization with the data, coding, searching for themes, reviewing themes, defining, and naming themes, and writing results (see Appendix E for more details). Inductive thematic analysis was chosen because of the relative newness of the field, the embryonic nature of some lines of inquiry, and the gradual development of a vocabulary describing responsible AI in healthcare. The iterative process of theme identification and checks for consistency and validity was extensive. Excerpts of an article that mention the chosen keywords were perused and coded. The codes were then transformed into themes and subthemes and subsequently grouped to form overarching themes.

Intercoder reliability was tested by recruiting an independent coder with an academic background who did not participate in the research project (Lombard et al., 2002). The coder was asked to code a subset of a random number of different selected articles to generate an intercoder reliability statistic. The intercoder reliability percentage agreement was 94.85%.

4. Critical Assessment - The SHIFT Framework

Studies exploring AI ethics in health care were classified into five themes: *Sustainable AI*, *Human-centric AI*, *Inclusive AI*, *Fair AI*, *Transparent AI*. Figure 3 presents the evolution in the frequency of themes identified in the articles. Note that an article categorized as, for example, Sustainable AI (Theme A) could overlap with themes from other categories due to commonalities between topics and areas of research. The pattern of the research shows that research about said themes picked up pace after 2015. Transparency issues concerning AI ethics and governance were the most frequently discussed, while issues surrounding inclusiveness have received less attention.

A thematic map was created to illustrate the scope of each theme and subtheme. We summarize the key findings generated from the representative studies as shown in Figure 4. Additionally, Appendix F lists the authors according to thematic areas.

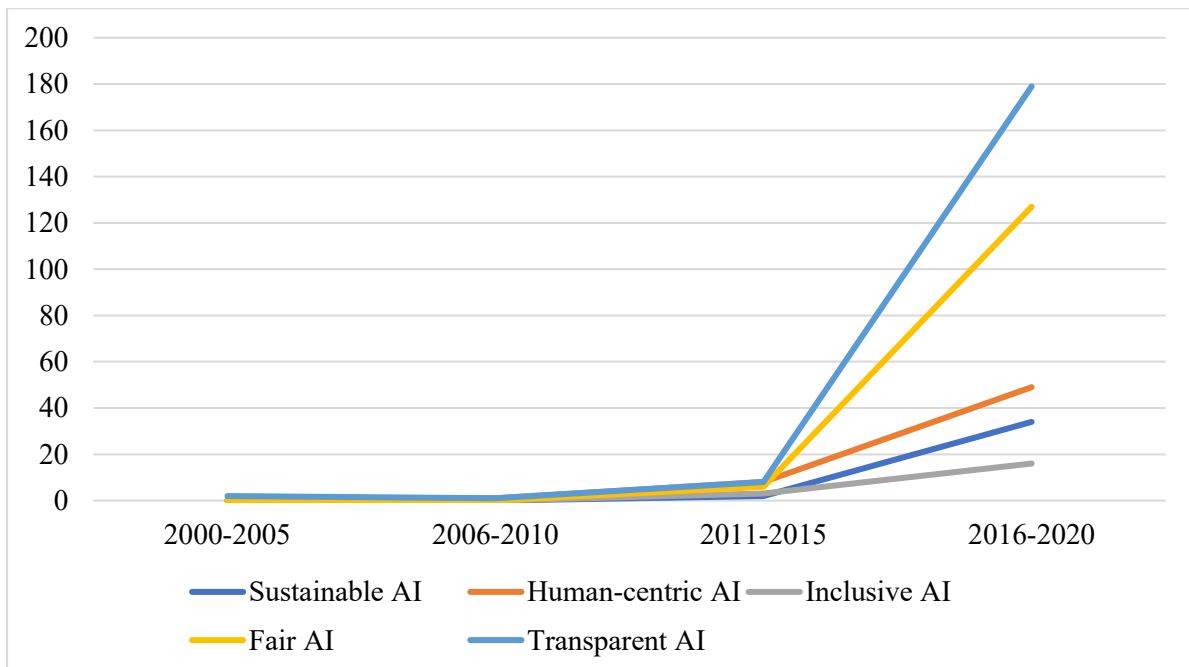


Figure 3. Research trends on each theme over time

Themes and sub-themes

Examples of responsible AI initiatives from representative studies

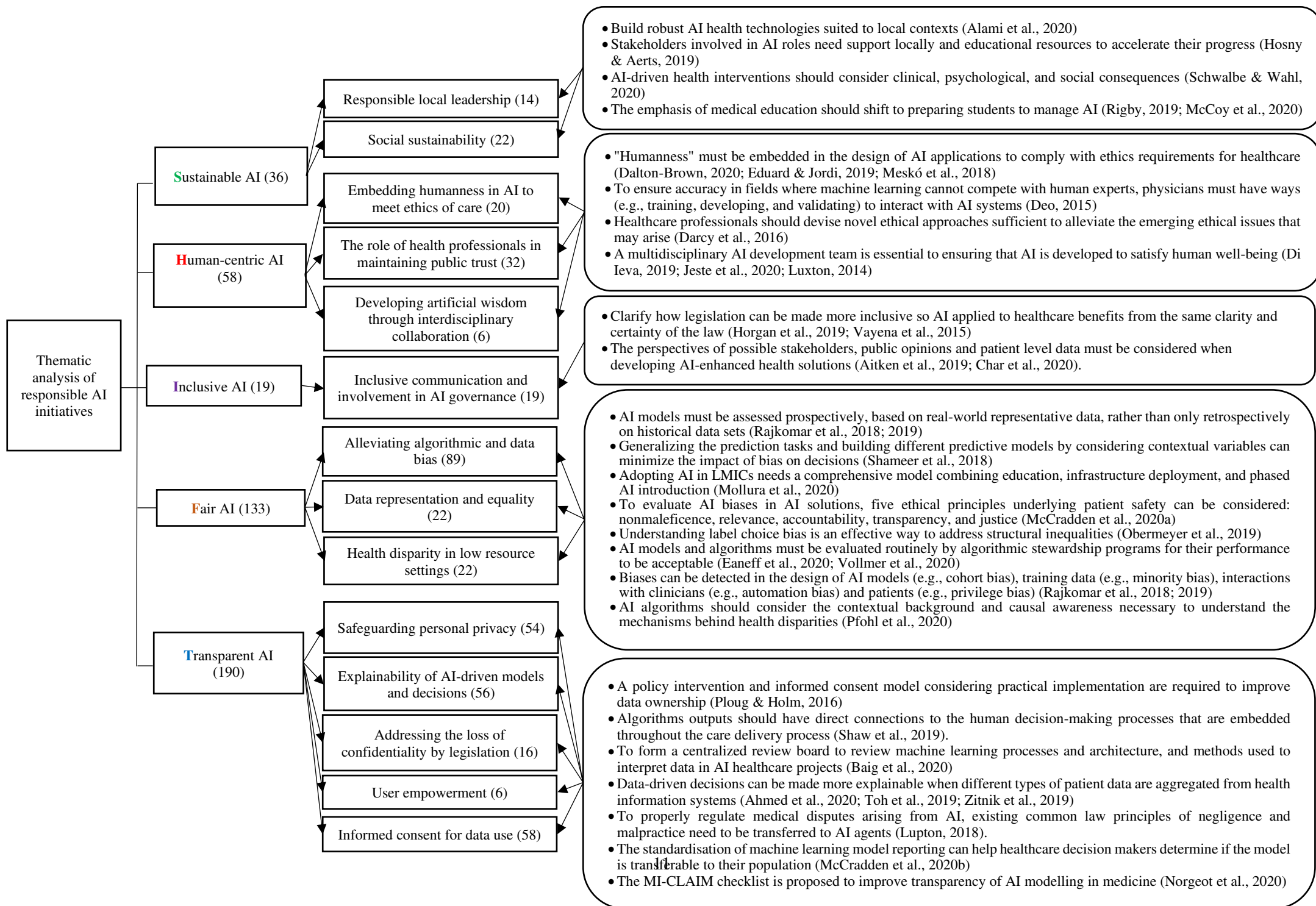


Figure 4. Thematic map of responsible AI initiatives in healthcare

4.1. Sustainable AI

Within the category of sustainable AI, we observed two subthemes: (1) building responsible local leadership to make AI solutions more sustainable and (2) AI for social sustainability.

Responsible local leadership. To make AI more sustainable, it is crucial to have a responsible local leadership to develop AI healthcare technologies that adapt to local contexts and are beneficial to the local population (Alami et al., 2017). Local governments, academic institutions, research centers, international agencies, non-governmental organizations (NGOs), industry, and civil society must be involved in the development and implementation of AI technologies (Alami et al., 2020). Alami et al. (2020) propose a framework comprising four building blocks to guide the implementation of sustainable AI healthcare technologies: (1) training and retention of local expertise, (2) robust monitoring systems, (3) a systems-based approach to implementation, and (4) a responsible and inclusive local leadership that includes all stakeholders. A key component of promoting sustainable AI is empowering local actors and fostering collaboration between stakeholders. To ensure that AI technology meets industry standards, local AI experts must be formed and retained. Local stakeholders should dedicate some resources to offering consulting services to help decision-makers and experts better understand and comply with AI standards (Hosny & Aerts, 2019). In addition, international organizations can also help establish appropriate governance strategies and identify essential areas for investment and interventions to avoid deploying disparate AI-driven solutions that cannot be supported by local infrastructure and expertise (Hosny & Aerts, 2019).

Social sustainability. AI for sustainability has attracted both academic and practitioner interest in recent years. When applying AI, its societal impact on the well-being of humans and the environment should be seriously considered (World Health Organization, 2021). AI must be applied responsibly to healthcare organizations in a way that balances stakeholders' needs, minimizes ethical concerns, and generates long-lasting profits. A healthcare organization's reputation and credibility could be severely tarnished if it develops AI algorithms (advertently or inadvertently) that compromise human rights and wellbeing. For instance, AI misuse such as using smart technology to supplant established health services has been flagged up as an ethical concern (Carter et al., 2020; Powell, 2019) that can potentially exacerbate existing health inequalities (Abramoff et al., 2020). In this sense, healthcare organizations should develop AI solutions that prioritize and support economic and social sustainability. Specifically, they need to establish policies on ethical governance considering socially preferable approaches, address ethical issues both in the initial design and post-launch stage of AI systems, and incorporate AI ethics in their social responsibility strategy (McCall, 2020).

Training of all healthcare staffs on responsible use of AI-based technology in healthcare should go beyond the scope of computer science and information technology. Partnerships between academic institutions and health service providers should be formed to ensure complementary use of skills in AI

technology, pedagogy, ethics, healthcare policy, and clinical practice. It has been suggested that medical schools need to integrate such programs into their curricula so that future medical and healthcare students become aware of the ethical considerations surrounding data collection and AI use (McCoy et al., 2020).

4.2. Human-centric AI

Embedding humanness in AI agents to meet ethics of care requirements. Since the existing ethical guidelines, standards, and regulations are generic and lack a central ethical framework and concrete principles applicable in the AI-based healthcare milieu, Ethics of Care has been proposed as the central ethical framework for designing AI-based health systems. Ethics of Care offers several advantages: it provides sufficiently concrete principles; it embodies values that are sensitive and applicable to the design of AI-based health systems in the caring practice milieu; and it is closely associated with the preservation and maintenance of trusting relationship between patients and AI agents (Dalton-Brown, 2020).

Humanness such as recognition, empathy and human experiences needs to be infused in AI agents to meet the requirements of ethics of care (Dalton-Brown, 2020). Several benefits have been identified for the use of humanized AI in healthcare. For instance, humans are open to interacting with AI agents on an emotional level, rather than just relying on them for practical functionality, even when they are aware that they are interacting with machines (Eduard & Jordi, 2019). Hoorn and Winter (2018) conclude robot doctors are better than human doctors at communicating bad news or unfavorable information to patients, but Blease et al. (2019) emphasize that communication and empathy are exclusively human abilities and clinical reasoning and value-based care are determined by a physician's judgment. Moreover, healthcare workers or caregivers may experience burnout (Luxton, 2014), but machines are not prone to cognitive errors, fatigue, boredom, and human negligence or forgetfulness. This may result in more accurate diagnostics and attentiveness during interactions with patients.

In addition, humanized AI applications such as AI carebots can be immune to the personal biases that human therapists may have and thus their consultations may result in a more detached, objective point of view. Some patients may therefore prefer interacting with an AI carebot rather than a human care provider as care seekers would feel less anxious when discussing intimate private issues with a carebot (Kandalaf et al., 2013). The British Standard's BS8611 for carebots provides AI developers with guidance on how to assess and reduce societal risks such as loss of trust, deception, privacy, security, safety, and confidentiality (ISO, 2014). Although the use of carebots in place of humans may eliminate the risk of ethical pitfalls encountered by human care providers, patients perceive the ability of AI carebots to emulate humans or animals as deception in some circumstances (Wangmo et al., 2019; Yew, 2020). Wangmo et al. (2019) found that some interviewees expressed concerns over AI carebots manifesting in human or pet form, which could implicitly deceive older adults with dementia by being erroneously perceived as real humans or pets.

The role of health professionals in maintaining public trust. AI can never fully replace personal trust and cognitive abilities, and as such healthcare professionals play a crucial role in AI-enabled care delivery (Meskó et al., 2018; Wangmo et al., 2019). In fact, the integrity of AI depends on the ability of health professionals to maintain broad public trust (Deo, 2015). Health professionals using AI technologies have three key ethical roles to play in their practices: (1) as medical domain experts who should provide computer and data scientists with the clinical context they need, (2) as gatekeepers for data quality who ensure that data inputs are relevant, accurate and sourced appropriately, and (3) as interpreters of AI black-box solutions who make real-time and post-hoc recommendations to patients (Miller, 2020). It has been suggested that to ensure that AI technology progresses in a way that upholds the social trust in medicine, healthcare professionals should join forces with industry leaders through collaboration and attempt to devise novel ethical approaches that can alleviate the emerging ethical issues that may arise in the future (Darcy et al., 2016). In the context of AI robot, Poulsen et al. (2020) contend that although AI robots in healthcare help healthcare professionals extend the service they provide, it is not clear how robots shape the codes of conduct, especially when it comes to cybersecurity. In their conclusion, they underscore the importance of including cybersecurity considerations in codes of conduct for robot developers and caregivers since the onus is on humans and not the machine to ensure that an AI system is secure and safe to use.

Developing artificial wisdom through interdisciplinary collaboration. Jeste et al. (2020) contend that the future need in AI technology is for artificial wisdom (AW) not artificial intelligence as the term intelligence does not really represent the technological needs of advancing society; it is wisdom that is associated with well-being, happiness, health, and longevity of individuals and the society. The development of AW necessitates the close and prudent collaborations by computer scientists, neuroscientists, mental health experts, and ethicists that collectively provide the greatest benefits to humanity (Jeste et al., 2020). Powell (2019) reasserts Jeste's point by stating that many medical decisions require not only ethical judgements, but also the doctor-patient relationship (Luxton, 2014), interdisciplinary collaborations (Littmann et al., 2020) and empathy and understanding (Wangmo et al., 2019) to arrive at a shared decision, often handling large areas of uncertainty and balancing competing risks.

4.3. Inclusive AI

Inclusive communication and involvement in AI governance. Digitalization is redefining the patient-provider relationship in terms of communication. Patients are concerned that AI-based health systems may change how they communicate with their providers, which may impact health service costs and quality (Luxton, 2014). To address this, scientific communities and public agencies can be instrumental in ensuring that inclusive communication between healthcare providers and the public is in place (Luxton, 2014; Noorbakhsh-Sabet et al., 2019; Poulsen et al., 2020). In addition, stakeholders (e.g., AI companies, healthcare organizations, regulatory agencies and policy makers, and patients)

from diverse fields and cultures, with different languages and forms of communication, should be involved in the design of AI solutions to mitigate unintended biases (Aitken et al., 2019; Char et al., 2020; World Health Organization, 2021). Horgan et al. (2019) and Vayena et al. (2015) advocate that legislation (e.g., inclusive impact assessment; World Health Organization, 2021) can be implemented to make AI applications more inclusive and ensure legal certainty and clarity.

4.4. Fair AI

Alleviating algorithmic and data bias. In general, biases can be found in AI model design (e.g. label bias and cohort bias), training data (e.g. minority bias, missing data bias), interactions with clinicians (e.g. automation bias and feedback loops bias), and interactions with patients (e.g. privilege bias and agency bias) (Rajkomar et al., 2018; 2019). For example, a study found that computer vision algorithms were used to ascertain sexual orientation of people by scrutinizing thousands of facial images taken from public profiles on a dating website (Wang & Kosinski, 2017). Another study with more far-reaching ethical implications reported various instances of how AI algorithms sometimes discriminate against certain groups, ethnic minorities, and individuals from deprived communities in areas such as credit ratings and health insurance (Ienca & Ignatiadis, 2020; O’Neil, 2016).

To minimize the impact of such bias, several solutions have been suggested by scholars (e.g., Eaneff et al., 2020; Shameer et al., 2018). In clinical settings, generalizing the prediction task and building different predictive models by considering contextual variables can minimize the impact of algorithmic bias on clinical decisions (Shameer et al., 2018). Moreover, AI models and algorithms must be evaluated routinely by algorithmic stewardship programs for their performance to be acceptable (Eaneff et al., 2020; Vollmer et al., 2020). Algorithm stewardship programs are designed to maintain an algorithm inventory overseen by a centralized therapeutics committee to ensure safety and fairness in the development of algorithms (Eaneff et al., 2020). Obermeyer et al. (2019) also suggest that identifying label choice bias in algorithms could potentially address structural inequalities.

Data representation and equality. The limitations and biased character of the data used to inform AI may have far-reaching socio-political and ethical implications (Strydom & Strydom, 2018). Biased AI algorithms trained with unrepresentative and inequitable datasets could lead to inaccurate medical diagnosis and decisions, or worse, discriminatory profiling of citizens in low resource settings (Ienca & Ignatiadis, 2020; Mittelstadt et al. 2016; Powell 2019). Faraj et al. (2018) maintain that AI algorithms are political by design in that they are imbued with the values, choices, beliefs, and norms of their developers and of those who assemble the datasets. For example, an AI solution trained with data biased towards an over-diagnosis of schizophrenia in African Americans can have detrimental consequences if used in some sub-Saharan African populations (Vayena et al., 2018). If algorithms trained with unrepresentative datasets are adopted in healthcare, they have the potential to exacerbate health disparities and may lead to underestimation or overestimation of risks in certain patient populations (Reddy et al., 2020; Vayena et al., 2018).

AI bias should be addressed by establishing a data governance panel made up of a representative target group or patients, clinical experts, and experts in AI, ethics, and law (Char et al., 2018; Reddy et al., 2020). The panel would monitor, and review datasets and algorithms used for training AI to ensure that the data is representative and that the algorithms used are impartial and sufficient to inform requisite model outcomes. To fulfil ethical duties to vulnerable individuals and avoid discrimination in the use of AI, stakeholders must be transparent about which communities and individuals are being monitored. Community leaders should be involved so that they can determine and report any adverse incidents affecting members of their community (Reddy et al., 2020).

Health disparity in low resource settings. Fair and equal access to low-cost AI health technologies across all socio-economic classes is a requirement to prevent the exacerbation of the socio-economic digital health divide (Alami et al., 2020; Horgan et al., 2019; Mehta et al., 2020). AI-powered health systems could narrow down the inequality in healthcare between developing and developed countries (Panchmatia et al., 2018). However, most AI-based health applications are developed and implemented in high-income countries and their effectiveness in improving healthcare service quality in Low-to-Middle-Income Countries (LMICs) can be questioned (Alami et al., 2020).

The economic challenges in LMICs and their dependency on development assistance impede investments in public health. Furthermore, the lack of governance may enable companies to commercialize solutions in LMICs that would not obtain regulatory approval in high-income countries (Christie, 2018). Prior studies found that most medical equipment donated to LMICs was substandard – they sporadically break down, or the provision of user manuals and training for local staff is lacking (Martinez-Martin & Kreitmair, 2018). In view of the extortionate expense and investment needed for AI, some countries may not be able to adopt these technologies beyond the pilot phase. The demands for responsible innovation for AI-based health technologies extend well-beyond observance and compliance with ethical and value-laden frameworks (Blobel et al., 2020) into ensuring the engagement of a highly diverse group of users. Users may vary in socio-demographics such as race, ethnicity, and socio-economic status as well as the diversity of the condition itself, with variations in behavior, cognition, and emotion. There has been a recurrent call for health professionals to play a more proactive role by engaging in the co-design and development of new and innovative AI health technologies (Miller & Polson, 2019; Panchmatia et al., 2018).

Several scenarios have been presented demonstrating how AI can address fairness in healthcare. For example, interactive AI-driven chatbots could provide better care by helping patients access care and follow-up services in a timely manner, and AI automated translation solutions could improve access to healthcare services in areas with language barriers (Luxton, 2014). Alami et al. (2020) argue that AI could predict and anticipate the spread of pathologies or vulnerabilities within certain groups or communities, and thus allow for more effective interventions in LMICs (Hosny & Aerts, 2019).

There is also the risk that the budget for AI might divert overall health, social budget and resources, and over-reliance on AI may lead to an erosion of clinical critical-thinking and local practice skills

(Alami et al., 2019). There is also a potential risk that AI diagnostic tools developed in high-income countries may recommend treatment plans (e.g., medication, surgery) that are very expensive or not available in LMICs (Hosny & Aerts, 2019; Price & Cohen, 2019). In contrast, AI-based health applications may offer many opportunities for LMICs where resources and expertise are lacking and could become a lever to provide access to universal, high-quality, and affordable healthcare for all. Alami et al. (2020) maintain that public–private partnerships can deliver smart health solutions to improve the health outcomes of those at risk of non-communicable diseases (NCD) by leveraging AI to intervene at many touch points along the patient journey (from health literacy and awareness to diagnosis and treatment). To effectively defeat the spread of pandemics and NCDs in LMICs, a focused, strategic, and collaborative approach across the healthcare value chain needs to be adopted where multilateral organizations, academia, governments, civil society organizations, healthcare providers, and the private sector contribute and collaborate in harmony.

4.5. *Transparent AI*

Safeguarding personal privacy. The collection and utilization of personal health data by AI and analytical algorithms gives rise to serious issues where patients suffer from privacy invasion, fraud, algorithmic bias, information leakage, and identity theft (Toh et al., 2019; Wearn et al., 2019). In fact, 49% of surveyed UK adults are unwilling to share their personal health data for developing algorithms that might improve quality of care (Fenech et al., 2018); this reluctance to share health data is mainly due to the possibility that shared or transferred data may be compromised or inadvertently leaked (McNair & Price, 2019). In particular, data breaches can lead to discrimination or criminal behavior in stigmatized or vulnerable populations (Xafis et al., 2019).

To address privacy concerns, patient data protection needs broader rulemaking authority so it can act more quickly as new privacy and security threats emerge (McGraw & Petersen, 2020). As such, Vayena et al. (2018) suggest that scientific committees and regulatory agencies need to research and propose ethical frameworks to identify and minimize the impact of biased models, as well as guide design choices to form systems that foster trust and understanding while maintaining a person’s privacy. Moreover, Kayaalp (2018) highlights how de-identification on patient sensitive data can result in better privacy protection, while Ma et al. (2019) present a perceptron-based privacy-preserving clinical decision model to eliminate the risk of privacy disclosure.

Limitations of conventional face-to-face psychological interventions could potentially be overcome through interventions from AI carebots. This comes in light of AI advances in machine learning, which creates new possibilities for decoding and analyzing neural data to deliver targeted neurointerventions. As neural data analytics become part of the healthcare ecosystem, it is important to assess the ethical implications, and a roadmap needs to be delineated for responsible innovation in this sector by considering various privacy issues such as mind reading, mental privacy, and issues relating to neurotechnology governance (Drew, 2019). This raises the question on whether a balance can be

struck between public rights to privacy and evidence required for law enforcement, and potential improvements in regulatory control (Benke & Benke, 2018).

Explainability of AI-driven models and decisions. This key issue has resulted in the emergence of a field termed explainable AI (XAI) (Rai, 2020). Two important aspects of XAI are understanding how a specific algorithm works and knowing who is responsible for its implementation (Floridi et al., 2018). Accountability of patient data access, analysis, and interpretation, and AI model development are essential to meeting the requirements of responsible AI usage in healthcare (Norgeot et al., 2020). Specifically, an effective AI-enabled healthcare system should be able to provide meaningful and personalized explanations about the results generated by algorithms (Davenport & Kalakota, 2019) and demonstrate the reliability and robustness of the AI models (Norgeot et al., 2020). Baig et al. (2020) propose to form a centralized review board to review machine learning processes and architecture, as well as methods used to interpret data in AI healthcare projects, while Zitnik et al. (2019) suggest that integrating patient data from various sources can improve explainability of clinical decisions.

Moreover, clinical decisions will be trusted if healthcare professionals can explain why a specific AI-driven treatment is an effective solution for a patient. Shaw et al. (2019) contend that decision support in health care can be useful only if its outputs are able to integrate with the human decision-making processes at the heart of health service delivery. Cabitza et al. (2017) further suggest that by combining machine learning with visualization, physicians may be able to explore the implications of outputs in rich interactive ways, alleviating the tension between accuracy and interpretability.

Addressing the loss of confidentiality by legislation. The rapid and prolific growth of big data collection and storage together with advancements in AI health technologies spawned previously unknown challenges: medical data is now accessible via mobile devices, shared networks, and even sensors attached to the human body (Wang et al., 2018). As information storage and retrieval technology has evolved, the public has become more sensitive to data confidentiality. There is growing concern that breaches of confidentiality may lead to an erosion of public confidence in the healthcare system (Ahmed et al., 2020). The presence of such concerns may impair the actual quality of care provided, since patients may self-medicate, visit another doctor, provide incomplete information, or opt out of seeking treatment (Yüksel et al., 2017). This is underpinned by research which found that the most common objection to sharing data with an outside provider is the potential discrimination by insurance companies (Ienca & Ignatiadis, 2020; Price & Cohen, 2019). Ongoing medical monitoring and privacy violations with medical devices can increase stigma for more disadvantaged citizens and possibly jeopardize access to health insurance and care for those unable to adopt new standards of healthy lifestyle (Briganti & Le Moine, 2020; Mittelstadt, 2017). Thus, legislation should be implemented to protect patients' confidentiality by creating a unique cause of action for those who wish to sue AI agents or healthcare organizations for misuse of their data (Lupton, 2018).

User empowerment. The advent of new technologies designed to capture voluntarily submitted data from patients has explored new ways to facilitate the sourcing of patients' personal health data.

Patients are encouraged to use web-based personal health repositories like Microsoft HealthVault to keep records of symptoms and treatments, document progress towards fitness goals, view medical test results, and facilitate communication with healthcare providers. Public consent about AI-powered digital health technology will manifest when privacy and ethical issues are addressed not only from a technical perspective (e.g., using digital key cryptography) or transparent disclosures on how citizen data is stored and processed (firm-generated reassurances), but by encouraging people to become more proactive in sharing and disseminating data about themselves. By converting citizens to prosumers of AI-powered digital health systems (Peters et al., 2020), end-users will feel empowered in controlling their own data and securing their privacy, which can lead to a more ethically aligned deployment of AI-powered health systems (Benke & Benke, 2018; Chadwick et al., 2014).

Empowerment is a recurrent theme rooted in users' concerns about the adoption of AI-powered digital health systems (Manrique de Lara & Peláez-Ballestas, 2020). The type of health data that users are reluctant to share usually harbors information that is considered sensitive, private, and potentially stigmatizing such as information related to pregnancy, contraception, sexual health, and mental health (Powell et al., 2006). Thus, evaluation mechanisms should be developed to assess if AI models and outcomes will be deemed acceptable by users (e.g., patients) whose data was collected and to clinicians who will use the model to make clinical decisions (Vollmer et al., 2020). In addition, it is imperative that the legislature should step in and deal with AI-related medical disputes by transferring existing common law principles of negligence and malpractice to AI agents (Lupton, 2018).

Informed consent for data use. Data privacy and ethical issues surrounding the use and storage of patient data were highlighted as major hurdles to the ownership of data in AI-powered digital health ecosystems (Bukowski et al., 2020). Healthcare stakeholders have also highlighted patient autonomy and informed consent as ethical priorities (Wangmo et al., 2019). This underlines the existence of diverse and substantive ethical challenges associated with obtaining adequate consent from patients. In terms of clinical trials, Jung and Pfister (2020) stressed the importance of introducing a written informed consent (WIC) procedure as a prerequisite to voluntary participation in a clinical trial and they proposed a secure framework for facilitating the development of ethical AI applications in healthcare that involves managing WIC documentation along the entire data value chain from acquiring consent to academic publication, and (commercially) exploiting the results of a clinical study. Larson et al. (2020) took a counter-argumentative approach by controversially contending that patient consent is not required before the data is used in exceptional circumstances such as an emergency (e.g., acute life-threatening situations). GDPR explicitly enables competent public health authorities to lawfully process personal data for reasons of substantial public interest without informed consent (Holub et al., 2020). For example, the consent from a patient or their legal guardian can be obtained at a later stage for conducting a clinical trial during the COVID-19 pandemic (European Medicines Agency, 2020).

It is expected that patients provide explicit consent if their data is shared, but recent episodes like the Royal Free London NHS Foundation Trust sharing patient data for the development of a clinical

application without explicit patient consent has presented concerns about privacy breaches (Carter et al., 2020; Reddy et al., 2020). Therefore, a careful policy intervention (Panch et al., 2019) and a model of informed consent (e.g., Meta Consent proposed by Ploug & Holm, 2016) that considers practical implementation could help in mitigating the said privacy concerns.

5. Argument development – The most pressing challenges in responsible AI

In contrast to the previous section's exploration of responsible AI initiatives, the argument development section focuses on uncovering the most pressing issues and challenges relating to responsible AI that have not been adequately addressed in existing literature, thereby motivating further research (Boell & Cecez-Kecmanovic, 2014). More specifically, we identify crucial issues and challenges of responsible AI specific to the SHIFT themes as well as enumerating the potential solutions proposed by the prior research, which are summarized in Table 2.

Sustainable AI. There remains an incomplete awareness within the medical community of how emerging AI technology can introduce ethical complexities into actual care taking (Rigby, 2019). Meanwhile, there seems to be a lack of clarity regarding what type of AI ethics training should be included to prepare and educate future healthcare professional in using AI technologies (Combs & Combs, 2019). There have been a few attempts to address ethical concerns raised by the AI revolution in healthcare by integrating ethical decision-making training into clinical training (Combs & Combs, 2019) and combining AI courses into curricula (McCoy et al., 2020; Park et al., 2019). For example, introducing virtual patients (VP) in medical education enables students to learn clinical and ethical decision making through practitioner-patient communication (Combs & Combs, 2019). However, VP has been criticized for its representation of diversity in a population and non-transparent algorithms for providing patient feedback.

Human-centric AI. It has been argued that AI's transformational role in healthcare allows for patients to be treated as commodities and not as individuals (Quinn et al., 2021). For example, driven by economic incentives and the perception that AI-powered health systems are superior to traditional techniques, health insurance companies could coerce care seekers to adopt AI-powered health systems without offering them the choice to seek care from a human alternative. Another example is that mental health services in the UK have been experiencing an Uberization due to the increased accessibility and lower costs of AI-powered health systems (e.g., chatbot) (Cotton, 2021). This may result in a loss of focus on the interpersonal processes of psychotherapy and patient-centric methods of addressing clinical problems that place the needs and preferences of individuals first.

To address this challenge, medical AI should shift its emphasis towards a patient-centered approach rather than a problem-oriented one (Quinn et al., 2021). Agarwal et al. (2020) go one step further by proposing a patient value-centered approach, which takes into account three core dimensions of value: preferences, precision, and process. The former approach is based on the concepts of patient-

centeredness, while the latter is primarily driven by value creation. Both approaches place patient rights at the core of clinical decision-making as a moral imperative.

Inclusive AI. The findings of several studies portend to the gradual impending colonization and commercialization of the domain of healthcare by leading healthcare providers such as the NHS and technology giants such as Google, Apple, and IBM, who leverage the colossal amount of digital data amassing online to generate profit (Downey, 2019; Larson et al., 2020). The commercial motives include selling advertising, goods, and services to users and on-selling of archived data to third parties such as pharmaceutical and health product companies. For example, the financial value of the NHS patients' data is estimated to be a whopping £10 billion a year (Downey, 2019). Commercial models were proposed by management consultants McKinsey, and they vary from the NHS receiving a curated dataset without a fee, to a royalty fee and shared ownership of products or discounts on products developed from the collaboration (Downey, 2019). Another prominent example of patient data commercialization can be illustrated in the reciprocal agreement signed by IBM and the Italian Government in early 2016 where IBM was bound to invest \$150 million in a health center that would be used for building e-health applications. In return, IBM will be granted access to valuable health data of the citizens of Lombardy (Monegain, 2016). In these cases, health data are commercialized for AI solutions without consideration of inclusive policies (Aitken et al., 2019) or diverse patient perspectives (Rickert, 2020). In order to regulate AI commercialization, government authorities, technology giants and healthcare service providers need to communicate and collaborate on a regular basis (Rickert, 2020). Meanwhile, it is equally important to engage diverse patient groups and broad stakeholders in AI governance. Some initial efforts such as All OF Us Research Program (All of Us Research Program Investigators, 2019) have been introduced to address this challenge.

Fair AI. The data used to train AI algorithms may use an unrepresentative dataset, limiting the ability to provide meaningful assessments or predictions (Carter et al., 2020; Ienca & Ignatiadis, 2020). Unrepresentative datasets are created by a lack of understanding of the contextualized factors such as sociodemography, health state, and social culture, which need to be carefully considered when developing AI solutions (Obermeyer et al., 2019). It has been suggested that AI algorithms should be trained on local datasets and should act prudently in the context of assisting or making a medical decision in the face of scientific uncertainty (Vollmer Dahlke & Ory, 2020).

In addition, while some AI technologies can deduce ethnicity, which is relevant in certain clinical cases, this function could be used for racial profiling or discrimination and, if not properly governed and managed, it could be exploited to marginalize individuals, groups, or communities based on their gender, ethnicity, socio-economic group, pathology, or sexual orientation (Alami et al., 2020; Ienca & Ignatiadis, 2020; Luxton, 2014; Price & Cohen, 2019). Some potential solutions have been developed, such as exploring local datasets or local checklists for training AI (Rajkomar et al., 2018; Rajkomar et al., 2019; Vollmer Dahlke & Ory, 2020), and conducting validation checks to mitigate label choice bias (Char et al., 2020; Obermeyer et al., 2019), but this challenge seems to persist. We therefore call for

more research in developing a method to ensure that representative datasets are used to avoid structural inequalities in AI development.

Transparent AI. AI-based methods are increasingly popular in population health research. Much of the data collected for such research is drawn from social media in the public domain or anonymous secondary health data, which makes it exempt from ethics committee scrutiny (Samuel & Derrick, 2020). Governance informs how scholars make ethical decisions and provides assurance to the public that researchers are acting ethically and it mitigates the risk associated with health over-surveillance or inequity. Questions remain about how to manage, process, and interpret data predictions in an ethically responsible manner, what constitutes an ethics governance framework, and how, in some jurisdictions, such a system would prevent exporting data to countries with a lax research ethics scrutiny (Benke & Benke, 2018). Blobel et al. (2020) proposed a potential solution to this ethical challenge by emphasizing that open-science repositories should be managed, curated, and driven by higher education institutions and funding bodies, with clear incentives for compliance. For example, to develop best practices, there should be a requirement for algorithms and affiliated data to be placed in repositories where access to data is restricted to certain stakeholders. Such controlled repositories create a way for other researchers and stakeholders to test the algorithms with their own data, checking for spurious predictions and highlighting any concerns or issues that may be present within the AI prediction models.

In addition, Blobel et al. (2020) suggested adding a second layer to ethical regulation: an ex-post review of innovative prediction algorithms used in specific sectors. In the public health milieu, ex-post reviews could be conducted by a multi-disciplinary committee comprising academics and stakeholders (such as professionals or users of the technology) from various disciplines including health and medicine, artificial intelligence, social science, and ethics. The aim of the committee would be to mitigate the risks of potential harm by reviewing scientific questions relating to the origin and quality of the data, algorithms, and AI; confirming the validation steps to ensure the prediction models work; and requesting further validation to be performed when the need arises. The committee could assume the role of an AI ombudsman or superintendent by assuming a regulatory role. Existing regulatory agencies are likely to be the most suitable candidates for such a role, and European and UK agencies have now started introducing measures to scrutinize medical software, which include providing guidelines and recommendations for the standardization of AI in healthcare. Thus, a rigorous and comprehensive approach needs to be developed to regulate and govern AI-powered healthcare ecosystems with transparency (Baric-Parker & Anderson, 2020).

Table 2. Challenges, issues and proposed solutions on responsible AI implementation

Responsible AI themes	Challenges	Issues challenging responsible AI	Proposed solutions
-----------------------	------------	-----------------------------------	--------------------

Sustainable AI	What kind of AI ethics training should be integrated into medical school curriculums?	<ul style="list-style-type: none"> • Medical professionals are not aware that emerging AI technology may introduce ethical complexities into the actual process of care giving (Rigby, 2019). • Lack of clarity regarding what AI ethics training should be included to prepare future healthcare professional for using AI technologies (Combs & Combs, 2019). 	<ul style="list-style-type: none"> • Integrating ethical decision-making training into clinical-based training through AI-enabled virtual patients (Combs & Combs, 2019) • Combining robust data science-focused courses into the baseline curriculum for health research (McCoy et al., 2020; Park et al., 2019)
Human-centric AI	How can a right balance be struck between delivering individualized care based on AI while attaining long-term profitability for healthcare providers?	<ul style="list-style-type: none"> • Patients being treated as commodities rather than individuals (Quinn et al., 2021) • Loss of interpersonal processes of responding to clinical problems in a way that prioritizes the needs and preferences of patients. 	<ul style="list-style-type: none"> • Using a patient-centered approach in designing medical AI that promotes informed choices aligned with patient values and respects patient autonomy (Sarkar et al., 2021; Quinn et al., 2021) • Delivering value with a patient-centric process (Agarwal et al., 2020)
Inclusive AI	How can the pursuit of commercialization of AI be inclusive to the broad public?	<ul style="list-style-type: none"> • Lack of inclusiveness policy for AI governance in relation to access and use of health data (Aitken et al., 2019) • Health data is commercialized for AI solutions without considering the voices of diverse patient groups (Rickert, 2020) 	<ul style="list-style-type: none"> • A continuous dialogue among authorities, technology giants and healthcare service providers to resolve potential ethical complexities of AI commercialization in healthcare (Rickert, 2020) • Engage diverse patient groups and broad stakeholders in AI governance (Chen et al., 2020)
Fair AI	Challenge 4: How can representative data be created and utilized to address patients' needs fairly?	<ul style="list-style-type: none"> • Contextualized factors such as sociodemography, health state, and social culture are not understood adequately to develop AI-based solutions that can cater for patients' needs (Obermeyer et al., 2019) • Training AI algorithms with unrepresentative dataset limits the ability to provide meaningful assessments or predictions (Carter et al., 2020; Ienca & Ignatiadis, 2020) 	<ul style="list-style-type: none"> • Allow for value pluralism and ensure that no protected characteristics are enforced (Morley et al., 2021) • Explore local datasets or local checklist for AI training to promote equal patient outcome, equal performance and equal allocation (Rajkomar et al., 2018; Rajkomar et al., 2019; Vollmer Dahlke & Ory, 2020) • Avoid label choice bias by conducting further validation studies (Char et al., 2020; Obermeyer et al., 2019)
Transparent AI	Challenge 5: To what extent should AI-led data use need to be transparent to all stakeholders?	<ul style="list-style-type: none"> • The fuzziness of AI governance in addressing the questions of how to interpret predictions properly (Benke & Benke, 2018) • The opaque design of explainable AI results from its ambiguous definition (Jiménez-Luna et al., 2020) 	<ul style="list-style-type: none"> • Higher education institutions and funding bodies should maintain, curate, and promote open-science repositories, with clear incentives for compliance (Blobel et al., 2020) • The ex-post reviews could be conducted by a multi-disciplinary committee to evaluate the quality of AI-driven explanations (Baric-Parker & Anderson, 2020; Blobel et al., 2020)

6. Research agenda – Future potentials of responsible AI

Continuous discussions are needed to comprehensively understand responsible AI use in healthcare. There are two primary future potentials on responsible AI in healthcare. First, we need to understand the individual, organizational, and societal impediments to achieving the SHIFT of AI. Previous medical AI research focuses on the technological understanding of its implementation and exploring the economic value of AI applications. Future research is needed to understand the practices,

mechanisms, and ecosystems that facilitate responsible AI use in healthcare. Second, while the tools that leverage AI are valuable for improving clinical practices, their actual use by healthcare professionals is not without challenges. To address the challenges associated with the use of AI, AI solution providers and developers should aim at designing and implementing ethical, transparent, and accountable AI solutions. This ethical consideration of AI would help healthcare organizations maintain trust and minimize potential risks. Thus, future research is needed to understand the role of responsible AI use from the SHIFT perspective to create value and reduce potential risks in healthcare.

Overall, the postulated benefits of AI-based digital health interventions should be corroborated further with future empirical studies exploring and testing their effectiveness in improving the decision-making and quality of healthcare (Bukowski et al., 2020). It is not just a matter of accentuating the ethical requirements in health care services research, rather it is a matter of anticipating the social consequences (system level) of scientific analysis and evaluation. Responsible AI in healthcare is a promising approach across the field of medical health; however, further research is needed to address the broader ethical and societal concerns of these technologies that are driving an evolution in digital healthcare. Table 3 outlines the future research opportunities for responsible AI in healthcare by proposing research questions based on the SHIFT framework.

Table 3. A research agenda for future research on responsible AI from a to z

Research themes	Sample research questions pertinent to each theme
Sustainable AI	<ul style="list-style-type: none"> a. What effective policies and actions can be taken by healthcare organizations or governments to leverage AI for the purpose of social sustainability? b. What kind of AI ethics training should be integrated into medical school curriculums? c. How can AI tools improve social impacts by reinforcing the regulations in tackling irresponsible medical practices in the digital world? d. What coordination mechanisms can mitigate ethical concerns regarding AI commercialization and ensure a more sustainable AI healthcare ecosystem in the long run?
Human-centric AI	<ul style="list-style-type: none"> e. How can AI systems incorporate and translate human judgements to generate accurate medical knowledge and insights? f. How should ethics education be designed and integrated into the training of AI solution developers? g. What are the effective coordination mechanisms for multi-actor decision making that involves amongst others, AI agents and healthcare professionals? h. How do AI technologies enhance co-ordination among health policy makers, healthcare organizations, and patients? i. Does AI and human coordination impede the effectiveness of AI and clinical efficacy? j. What defines the paradoxical nature of bias from the sociomateriality perspective of algorithmic decision-making?
Inclusive AI	<ul style="list-style-type: none"> k. What are the implications of inclusive communication for AI-based digital health management?

	<ul style="list-style-type: none"> l. What type of individual roles and disciplines in an ethics panel contribute most to a better understanding of inclusive communication in AI-based digital healthcare? m. How can AI-enabled health systems operate in a coordinated manner to deliver inclusive care to patients?
Fair AI	<ul style="list-style-type: none"> n. To what extent do AI algorithms affect healthcare practitioners' efficiency and quality of care? o. What algorithmic attributes and characteristics are required for reducing biases in the data and prediction model? p. What safeguards measures can healthcare organizations take to ensure that patients are treated fairly when a medical decision is delegated to an AI-based health system? q. In what ways can minorities and marginalized groups be involved in consultations to mitigate biases and structural inequalities? r. How to empower local actors and foster local collaboration between stakeholders to develop equitable AI solutions? s. How can AI systems and applications be designed in way to enhance patients' perceptions of fairness and trust? t. What kind of low-cost AI solutions can be deployed to address health disparity in low resource settings (e.g., LMICs)?
Transparent AI	<ul style="list-style-type: none"> u. How can the risks of cybersecurity, data loss and patient identity theft be reduced or mitigated through health data governance? v. To what extent should AI healthcare system be transparent in terms of data use and algorithm management? w. How should AI generate results that are discernible and lucid to users and health practitioners? x. What factors drive patients or healthcare professionals to share data in AI-driven digital health environments? y. How can data quality assurance and programming norm be cultivated in the AI development stages? z. How can health data be utilized to support transparent evidence-based medicine using AI approaches?

7. Discussion and Conclusion

7.1. Academic and practical Implications

The auspicious outlook and promise of better addressing health inequalities through integrating AI into clinical practices comes with an unanticipated challenge: healthcare organizations are now more prone of committing ethical or moral infringements than before. Healthcare service providers and medical-algorithm designers have been asked to act responsibly in response to the expectations of governments, regulators, and wider stakeholders. In the current review, 253 articles across many domains (e.g., medicine, healthcare management, information systems, and bioethics) from the past 20 years were systematically organized and reviewed to devise a responsible AI framework. Underpinned by virtue ethics theory, responsible AI initiatives across 5 main themes and 14 sub-themes were outlined in the framework. As noted above, virtue ethics theory has been used to examine what constitutes a moral action in the context of business management by focusing on the ethical aspects of everyday business operations. (Audi, 2012). This theory offers an excellent anchor to study responsible AI initiatives by emphasizing ethical practices over deontological or utilitarian perspectives of ethics. By

applying this theory to an under researched context- AI use and application in healthcare, our review could inform researchers and practitioners to go beyond the symbolic advocacy of AI ethics and to focus on healthcare practices and actions for governing AI responsibly.

The core of this review highlights to medical-algorithm designers, policymakers, healthcare providers, and patients, the importance of profoundly understanding responsible AI initiatives as this would facilitate the provision of more efficacious and responsible AI-powered healthcare service. Therefore, we provide three practical implications. First, our analysis goes beyond merely exploring AI ethics by identifying in detail how AI is used responsibly in healthcare contexts. For instance, in the sustainable AI theme, our review reveals that building robust education and training programs is one of the core initiatives of responsible AI. In the human-centric AI theme, our review suggests that the most appropriate role for AI-powered health systems is that of an assistant: to support human practitioners in their clinical care decisions. Careful testing and evaluation of these systems will be needed regardless of application, albeit there is a paradoxical argument between AI and humans that a certain degree of imperfection can be conducive to a health treatment and intervention (Luxton, 2014).

Second, we expose the most pressing challenges of responsible AI (see Section 5) and call for addressing these challenges by answering our proposed research questions. For example, in our review, we find that a diverse group of experts and stakeholders is needed to develop equitable medical AI solutions. It has been suggested that the research community, healthcare providers and practitioners, AI health technology providers, and the federal government need to work together to design AI-specific common rules based on the principles of virtue ethics. However, the extant literature did not provide solutions to the question of how minorities and marginalized groups can participate in consultations to mitigate biases and structural inequalities.

Third, this review has provided a detailed analysis of the responsible approaches related to the implementation of AI in healthcare. This review can encourage the development of proportionate ethical policies and regulatory interventions by creating a system of transparency and distributed responsibility that makes not just healthcare practitioners responsible, but all actors involved in the supply chain of AI algorithms. Specifically, developing medical AI responsibly requires incorporating five initiatives: sustainability, human-centeredness, inclusiveness, fairness, and transparency that can be assessed by policymakers and legislators to determine if inherent risks or biases are appropriately mitigated for better adoption of AI.

7.2. Limitations

It is important to acknowledge that the majority of the articles reviewed in our systematic review represent a Western viewpoint (the articles mainly originate from West Europe and North America), despite the fact that our systematic review is grounded in a substantial number of reliable sources. Hence, future research should consider searching and reviewing studies from other languages or

continents to acquire a more nuanced or perhaps, a broader understanding of what constitutes responsible AI in healthcare.

We chose not to conduct bibliometric analyses for one major reason. The bibliometric approach focuses on an author-centric review by tracing back the origins of topic, authorship, and citation over time and presenting the findings in a descriptive manner. Conducting this type of review seems not to be appropriate given that our primary objective is to identify responsible AI initiatives so that healthcare practitioners can take advantage of our findings. However, we encourage future researchers to use bibliometric analyses to visualize descriptive results such as topic development of responsible AI initiatives.

References

- Abràmoff, M. D., Leng, T., Ting, D. S. W., Rhee, K., Horton, M. B., Brady, C. J., & Chiang, M. F. (2020). Automated and computer-assisted detection, classification, and diagnosis of diabetic retinopathy. *Telemedicine and e-Health*, 26(4), 544–550.
- Agarwal, R., Dugas, M., Gao, G. G., & Kannan, P. K. (2020). Emerging technologies and analytics for a new era of value-centered marketing in healthcare. *Journal of the Academy of Marketing Science*, 48(1), 9-23.
- Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database*, <https://doi.org/10.1093/database/baaa010>.
- Aitken, M., Tully, M. P., Porteous, C., Denegri, S., Cunningham-Burley, S., Banner, N., ... & Willison, D. J. (2019). Consensus statement on public involvement and engagement with data intensive health research. *International Journal of Population Data Science*, 4(1), 586.
- Alami, H., Gagnon, M.-P., & Fortin, J.-P. (2017). Digital health and the challenge of health systems transformation. *Mhealth*, 3, 31.
- Alami, H., Gagnon, M.-P., & Fortin, J.-P. (2019). Some multidimensional unintended consequences of telehealth utilization: a multi-project evaluation synthesis. *International Journal of Health Policy and Management*, 8(6), 337–352.
- Alami, H., Rivard, L., Lehoux, P., Hoffman, S. J., Cadeddu, S. B. M., Savoldelli, M., Samri, M. A., Ahmed, M. A. A., Fleet, R., & Fortin, J.-P. (2020). Artificial intelligence in health care: laying the Foundation for Responsible, sustainable, and inclusive innovation in low- and middle-income countries. *Globalization and Health*, 16(1), 52.
- All of Us Research Program Investigators. (2019). The “All of Us” research program. *New England Journal of Medicine*, 381(7), 668-676.

- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115.
- Audi, R. (2012). Virtue ethics as a resource in business. *Business Ethics Quarterly*, 22(2), 273-291.
- Baig, M. A., Almuhaizea, M. A., Alshehri, J., Bazarbashi, M. S., & Al-Shagathrh, F. (2020). Urgent need for developing a framework for the governance of AI in healthcare. *Studies in Health Technology and Informatics*, 272, 253–256.
- Baric-Parker, J., & Anderson, E. E. (2020). Patient data sharing for AI: ethical challenges, Catholic solutions. *The Linacre Quarterly*, 87(4), 471–481.
- Benke, K., & Benke, G. (2018). Artificial intelligence and big data in public health. *International Journal of Environmental Research and Public Health*, 15(12), 2796.
- Blease, C., Kaptchuk, T. J., Bernstein, M. H., Mandl, K. D., Halamka, J. D., & DesRoches, C. M. (2019). Artificial Intelligence and the future of primary care: exploratory qualitative study of UK general practitioners' views. *Journal of Medical Internet Research*, 21(3), e12802.
- Blobel, B., Ruotsalainen, P., Brochhausen, M., Oemig, F., & Uribe, G. A. (2020). Autonomous systems and artificial intelligence in healthcare transformation to 5P medicine – ethical challenges. *Studies in Health Technology and Informatics*, 270, 1089–1093.
- Boell, S. K., & Cecez-Kecmanovic, D. (2014). A hermeneutic approach for conducting literature reviews and literature searches. *Communications of the Association for Information Systems*, 34(1), 12.
- Briganti, G., & Le Moine, O. (2020). Artificial Intelligence in medicine: today and tomorrow. *Frontiers in Medicine*, 7, 27.
- Bukowski, M., Farkas, R., Beyan, O., Moll, L., Hahn, H., Kiessling, F., & Schmitz-Rode, T. (2020). Implementation of eHealth and AI integrated diagnostics with multidisciplinary digitized data: are we ready from an international perspective? *European Radiology*, 30(10), 5510–5524.
- Cabitza, F., Rasoini, R., & Gensini, G. F. (2017). Unintended consequences of machine learning in medicine. *JAMA*, 318(6), 517–518.
- Carter, S. M., Rogers, W., Win, K. T., Frazer, H., Richards, B., & Houssami, N. (2020). The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. *The Breast*, 49, 25–32.
- Chadwick, R., Levitt, M., & Shickle, D. (2014). *The Right to Know and the Right Not to Know: Genetic Privacy and Responsibility*. Cambridge University Press.
- Chakrabarty, S., & Bass, A. E. (2015). Comparing virtue, consequentialist, and deontological ethics-based corporate social responsibility: mitigating microfinance risk in institutional voids. *Journal of Business Ethics*, 126(3), 487–512.
- Char, D. S., Abràmoff, M. D., & Feudtner, C. (2020). Identifying ethical considerations for machine learning healthcare applications. *The American Journal of Bioethics*, 20(11), 7–17.

- Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care – addressing ethical challenges. *The New England Journal of Medicine*, 378(11), 981–983.
- Chatterjee, S., Sarker, S., and Fuller, M. (2009). A deontological approach to designing ethical collaboration. *Journal of the Association for Information Systems*, 10(3), 138–169.
- Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., & Ghassemi, M. (2020). Ethical Machine Learning in Healthcare. *Annual Review of Biomedical Data Science*, 4.
- Chen, J. H., & Asch, S. M. (2017). Machine learning and prediction in medicine—beyond the peak of inflated expectations. *The New England journal of Medicine*, 376(26), 2507–2509.
- Christie, G. (2018). Progressing the health agenda: Responsibly innovating in health technology. *Journal of Responsible Innovation*, 5(1), 143–148.
- Chun, R. (2005). Ethical character and virtue of organizations: an empirical assessment and strategic implications. *Journal of Business Ethics*, 57(3), 269–284.
- Cipriani, A., & Geddes, J. (2003). Comparison of systematic and narrative reviews: the example of the atypical antipsychotics. *Epidemiology and Psychiatric Sciences*, 12(3), 146–153.
- Clarke, V., & Braun, V. (2013). Teaching thematic analysis: overcoming challenges and developing strategies for effective learning. *The Psychologist*, 26(2), 120–123.
- Combs, C. D., & Combs, P. F. (2019). Emerging roles of virtual patients in the age of AI. *AMA Journal of Ethics*, 21(2), E153–159.
- Cotton, E. (2021). Mental health services in England are being ‘Uberised’ – and that’s bad for patients and therapists. Available at: <https://theconversation.com/mental-health-services-in-england-are-being-uberised-and-thats-bad-for-patients-and-therapists-167065>.
- Culnan, M. J., & Williams, C. C. (2009). How ethics can enhance organizational privacy: lessons from the choicepoint and TJX data breaches. *MIS Quarterly*, 33(4), 673–687.
- Dalton-Brown, S. (2020). The ethics of medical AI and the physician-patient relationship. *Cambridge Quarterly of Healthcare Ethics*, 29(1), 115–121.
- Darcy, A. M., Louie, A. K., & Roberts, L. W. (2016). Machine learning and the profession of medicine. *JAMA*, 315(6), 551–552.
- Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, 6(2), 94–98.
- Deo, R. C. (2015). Machine learning in medicine. *Circulation*, 132(20), 1920–1930.
- Di Ieva, A. (2019). AI-augmented multidisciplinary teams: hype or hope?. *The Lancet*, 394(10211), 1801.
- Downey, A. (2019). NHS bosses meet with tech giants to discuss commercial patient database. *DigitalHealth*. available at: <https://www.digitalhealth.net/2019/12/nhs-bosses-meet-with-tech-giants-to-discuss-commercial-patient-database/>.
- Drew, L. (2019). The ethics of brain-computer interfaces, *Nature*, 571, S19.

- Eaneff, S., Obermeyer, Z., & Butte, A. J. (2020). The case for algorithmic stewardship for artificial intelligence and machine learning technologies. *JAMA*, 324(14), 1397–1398.
- Eduard, F.-V., & Jordi, A.-C. (2019). ‘I’ll take care of you,’ said the robot. *Paladyn, Journal of Behavioral Robotics*, 10(1), 77–93.
- European Medicines Agency. (2020). *Guidance on the Management of Clinical Trials during the COVID-19 (Coronavirus) Pandemic*. available at: https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-10/guidanceclinicaltrials_covid19_en.pdf.
- Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, 28(1), 62–70.
- Fenech, M., Strukelj, N., & Buston, O. (2018). Ethical, social and political challenges of artificial intelligence in health. Available at: <https://wellcome.ac.uk/sites/default/files/ai-in-health-ethical-social-political-challenges.pdf>.
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707.
- Floridi, L., Luetge, C., Pagallo, U., Schafer, B., Valcke, P., Vayena, E., ... & Kalra, D. (2019). Key ethical challenges in the European medical information framework. *Minds and Machines*, 29(3), 355–371.
- Flyverbom, M., Deibert, R., & Matten, D. (2019). The governance of digital technology, big data, and the internet: new roles and responsibilities for business. *Business & Society*, 58(1), 3–19.
- Galetsis, P., Katsaliaki, K., & Kumar, S. (2019). Values, challenges and future directions of big data analytics in healthcare: A systematic review. *Social Science & Medicine*, 241, 112533.
- Holub, P., Kozera, L., Florindi, F., van Enckevort, E., Swertz, M., Reihs, R., Wutte, A., Valik, D., & Mayrhofer, M. Th. (2020). BBMRI–ERIC’s contributions to research and knowledge exchange on COVID-19. *European Journal of Human Genetics*, 28(6), 728–731.
- Hoorn, J. F., & Winter, S. D. (2018). Here comes the bad news: doctor robot taking over. *International Journal of Social Robotics*, 10(4), 519–535.
- Horgan, D., Romao, M., Morr e, S. A., & Kalra, D. (2019). Artificial Intelligence: power for civilisation – and for better healthcare. *Public Health Genomics*, 22(5–6), 145–161.
- Hosny, A., & Aerts, H. J. (2019). Artificial intelligence for global health. *Science*, 366(6468), 955–956.
- Ienca, M., & Ignatiadis, K. (2020). Artificial Intelligence in clinical neuroscience: methodological and ethical challenges. *AJOB Neuroscience*, 11(2), 77–87.
- ISO. (2014). Robots and robotic devices – safety requirements for personal care robots. Available at <https://www.iso.org/standard/53820.html>.

- Jeste, D. V., Graham, S. A., Nguyen, T. T., Depp, C. A., Lee, E. E., & Kim, H.-C. (2020). Beyond artificial intelligence: exploring artificial wisdom. *International Psychogeriatrics*, 32(8), 993–1001.
- Jiménez-Luna, J., Grisoni, F., & Schneider, G. (2020). Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence*, 2(10), 573-584.
- Jobin, A., Lenca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389–399.
- Jung, H. H., & Pfister, F. M. (2020). Blockchain-enabled clinical study consent management. *Technology Innovation Management Review*, 10(2), 14–24.
- Kandalaf, M. R., Didehbani, N., Krawczyk, D. C., Allen, T. T., & Chapman, S. B. (2013). Virtual reality social cognition training for young adults with high-functioning autism. *Journal of Autism and Developmental Disorders*, 43(1), 34–44.
- Kayaalp, M. (2018). Patient privacy in the era of big data. *Balkan Medical Journal*, 35(1), 8–17.
- Larson, D. B., Magnus, D. C., Lungren, M. P., Shah, N. H., & Langlotz, C. P. (2020). Ethics of using and sharing clinical imaging data for artificial intelligence: a proposed framework. *Radiology*, 295(3), 192536.
- Littmann, M., Selig, K., Cohen-Lavi, L., Frank, Y., Hönigschmid, P., Kataka, E., ... & Rost, B. (2020). Validity of machine learning in biology and medicine increased through collaborations across fields of expertise. *Nature Machine Intelligence*, 2(1), 18–24.
- Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research*, 28(4), 587–604.
- Lupton, M. (2018). Some ethical and legal consequences of the application of artificial intelligence in the field of medicine. *Trends in Medicine*, 18(4), 1–7.
- Luxton, D. D. (2014). Recommendations for the ethical use and design of artificial intelligent care providers. *Artificial Intelligence in Medicine*, 62(1), 1–10.
- Ma, H., Guo, X., Ping, Y., Wang, B., Yang, Y., Zhang, Z., & Zhou, J. (2019). PPCD: privacy-preserving clinical decision with cloud support. *Plos One*, 14(5), e0217349.
- Manrique de Lara, A., & Peláez-Ballestas, I. (2020). Big data and data processing in rheumatology: bioethical perspectives. *Clinical Rheumatology*, 39(4), 1007–1014.
- Martinez-Martin, N., & Kreitmair, K. (2018). Ethical issues for direct-to-consumer digital psychotherapy apps: Addressing accountability, data protection, and consent. *JMIR Mental Health*, 5(2), e32.
- McCall, B. (2020). COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread. *The Lancet Digital Health*, 2(4), e166–e167.
- McCoy, L. G., Nagaraj, S., Morgado, F., Harish, V., Das, S., & Celi, L. A. (2020). What do medical students actually need to know about artificial intelligence?. *NPJ Digital Medicine*, 3(1), 1–3.

- McCradden, M. D., Joshi, S., Anderson, J. A., Mazwi, M., Goldenberg, A., & Zlotnik Shaul, R. (2020a). Patient safety and quality improvement: Ethical principles for a regulatory approach to bias in healthcare machine learning. *Journal of the American Medical Informatics Association*, 27(12), 2024–2027.
- McCradden, M. D., Joshi, S., Mazwi, M., & Anderson, J. A. (2020b). Ethical limitations of algorithmic fairness solutions in health care machine learning. *The Lancet Digital Health*, 2(5), e221–e223.
- McGraw, D., & Petersen, C. (2020). From commercialization to accountability: responsible health data collection, use, and disclosure for the 21st century. *Applied Clinical Informatics*, 11(02), 366–373.
- McNair, D., & Price, W. N. (2019). Health care AI: Law, regulation, and policy. Artificial intelligence in health care: The hope, the hype, the promise, the peril. Washington DC: National Academy of Medicine
- Mehta, M. C., Katz, I. T., & Jha, A. K. (2020). Transforming global health with AI. *New England Journal of Medicine*, 382(9), 791–793.
- Meskó, B., Hetényi, G., & Gyórfy, Z. (2018). Will artificial intelligence solve the human resource crisis in healthcare? *BMC Health Services Research*, 18(1), 545. doi:10.1186/s12913-018-3359-4.
- Miller, D. D. (2020). Machine intelligence in cardiovascular medicine. *Cardiology in Review*, 28(2), 53–64..
- Miller, E., & Polson, D. (2019). Apps, avatars, and robots: the future of mental healthcare. *Issues in Mental Health Nursing*, 40(3), 208–214.
- Mittelstadt, B. (2017). Ethics of the health-related internet of things: a narrative review. *Ethics and Information Technology*, 19(3), 157–175.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: mapping the debate. *Big Data & Society*, 3(2), 1–21.
- Mollura, D. J., Culp, M. P., Pollack, E., Battino, G., Scheel, J. R., Mango, V. L., ... & Dako, F. (2020). Artificial intelligence in low-and middle-income countries: innovating global health radiology. *Radiology*, 297(3), 513–520.
- Monegain, B. (2016). IBM Watson takes analytics prowess overseas: Supercomputer to work on big data and genomics in Italy. Available at: <https://www.healthcareitnews.com/news/ibm-watson-takes-analytics-prowess-overseas-supercomputer-work-big-data-and-genomics-italy>.
- Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M., & Floridi, L. (2021). Operationalising AI ethics: barriers, enablers and next steps. *AI & Society*, <https://doi.org/10.1007/s00146-021-01308-8>.
- Morley, J., Machado, C. C., Burr, C., Cows, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: A mapping review. *Social Science & Medicine*, 260, 113172.
- Newell, S., & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of ‘datification’. *The Journal of Strategic Information Systems*, 24(1), 3–14.

- Noorbakhsh-Sabet, N., Zand, R., Zhang, Y., & Abedi, V. (2019). Artificial intelligence transforms the future of health care. *The American Journal of Medicine*, 132, 795–801.
- Norgeot, B., Quer, G., Beaulieu-Jones, B. K., Torkamani, A., Dias, R., Gianfrancesco, M., ... & Butte, A. J. (2020). Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. *Nature Medicine*, 26(9), 1320–1324.
- O’Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Penguin.
- 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.
- O’Sullivan, S., Nevejans, N., Allen, C., Blyth, A., Leonard, S., Pagallo, U., ... & Ashrafian, H. (2019). Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 15(1), e1968.
- Panch, T., Mattie, H., & Celi, L. A. (2019). The “inconvenient truth” about AI in healthcare. *NPJ Digital Medicine*, 2(1), 1–3.
- Panchmatia, J. R., Visenio, M. R., & Panch, T. (2018). The role of artificial intelligence in orthopaedic surgery. *British Journal of Hospital Medicine*, 79(12), 676–681.
- Park, S. H., Do, K. H., Kim, S., Park, J. H., & Lim, Y. S. (2019). What should medical students know about artificial intelligence in medicine?. *Journal of Educational Evaluation for Health Professions*, 16, 18.
- Peters, D., Vold, K., Robinson, D., & Calvo, R. A. (2020). Responsible AI—two frameworks for ethical design practice. *IEEE Transactions on Technology and Society*, 1(1), 34–47.
- Pfohl, S. R., Foryciarz, A., & Shah, N. H. (2020). An empirical characterization of fair machine learning for clinical risk prediction. *Journal of Biomedical Informatics*, <https://doi.org/10.1016/j.jbi.2020.103621>.
- Ploug, T., & Holm, S. (2016). Meta consent—a flexible solution to the problem of secondary use of health data. *Bioethics*, 30(9), 721–732.
- Poulsen, A., Fosch-Villaronga, E., & Burmeister, O. K. (2020). Cybersecurity, value sensing robots for LGBTIQ+ elderly, and the need for revised codes of conduct. *Australasian Journal of Information Systems*, 24
- Powell, J. (2019). Trust me, i’m a chatbot: how artificial intelligence in health care fails the Turing Test. *Journal of Medical Internet Research*, 21(10), e16222.
- Powell, J., Fitton, R., & Fitton, C. (2006). Sharing electronic health records: the patient view. *Journal of Innovation in Health Informatics*, 14(1), 55–57.
- Powles, J., & Hodson, H. (2017). Google DeepMind and healthcare in an age of algorithms. *Health and Technology*, 7(4), 351–367.

- Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature Medicine*, 25(1), 37–43.
- Quinn, T. P., Senadeera, M., Jacobs, S., Coghlan, S., & Le, V. (2021). Trust and medical AI: the challenges we face and the expertise needed to overcome them. *Journal of the American Medical Informatics Association*, 28(4), 890–894.
- Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, 48(1), 137–141.
- Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380(14), 1347–1358.
- Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine*, 169(12), 866–872.
- Reddy, S., Allan, S., Coghlan, S., & Cooper, P. (2020). A governance model for the application of AI in health care. *Journal of the American Medical Informatics Association*, 27(3), 491–497.
- Rickert, J. (2020). On Patient Safety: The Lure of Artificial Intelligence—Are We Jeopardizing Our Patients’ Privacy?. *Clinical Orthopaedics and Related Research*, 478(4), 712–714.
- Rigby, M. J. (2019). Ethical dimensions of using artificial intelligence in health care. *AMA Journal of Ethics*, 21(2), 121–124.
- Ross, C., & Swetlitz, I. (2018). IBM’s Watson supercomputer recommended ‘unsafe and incorrect’ cancer treatments, internal documents show. Stat News. available at: <https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/>.
- Samuel, G., & Derrick, G. (2020). Defining ethical standards for the application of digital tools to population health research. *Bulletin of the World Health Organization*, 98(4), 239–244.
- Schwalbe, N., & Wahl, B. (2020). Artificial intelligence and the future of global health. *The Lancet*, 395(10236), 1579–1586.
- Shameer, K., Johnson, K. W., Glicksberg, B. S., Dudley, J. T., & Sengupta, P. P. (2018). Machine learning in cardiovascular medicine: are we there yet?. *Heart*, 104(14), 1156–1164.
- Shaw, J., Rudzicz, F., Jamieson, T., & Goldfarb, A. (2019). Artificial intelligence and the implementation challenge. *Journal of Medical Internet Research*, 21(7), e13659.
- Siddaway, A., Wood, A., & Hedges, L. (2019). How to do a systematic review: a best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology*, 70, 747–770.
- Song, S. Y., & Kim, Y.-K. (2018). Theory of virtue ethics: do consumers’ good traits predict their socially responsible consumption? *Journal of Business Ethics*, 152(4), 1159–1175.
- Strydom, S. K., & Strydom, M. (2018). *Big Data Governance and Perspectives in Knowledge Management*. IGI Global.
- Toh, T. S., Dondelinger, F., & Wang, D. (2019). Looking beyond the hype: Applied AI and machine learning in translational medicine. *EBioMedicine*, 47, 607–615.

- Trocin, C., Mikalef, P., Papamitsiou, Z., & Conboy, K. (2021). Responsible AI for digital health: a synthesis and a research agenda. *Information Systems Frontiers*, 1-19. <https://doi.org/10.1007/s10796-021-10146-4>.
- Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: addressing ethical challenges. *PLoS Medicine*, 15(11), e1002689.
- Vayena, E., Salathé, M., Madoff, L. C., & Brownstein, J. S. (2015). Ethical challenges of big data in public health. *PLoS Computational Biology*, 11(2), e1003904.
- Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., ... & Nerini, F. F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, 11(1), 1–10.
- Vollmer Dahlke, D., & Ory, M. G. (2020). Emerging issues of intelligent assistive technology use among people with dementia and their caregivers: a US perspective. *Frontiers in Public Health*, 8(191). doi: 10.3389/fpubh.2020.00191.
- Vollmer, S., Mateen, B. A., Bohner, G., Király, F. J., Ghani, R., Jonsson, P., ... & Hemingway, H. (2020). Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *BMJ*, 368, doi: <https://doi.org/10.1136/bmj.16927>.
- Wamba, S. F., & Queiroz, M. M. (2021). Responsible Artificial Intelligence as a Secret Ingredient for Digital Health: Bibliometric Analysis, Insights, and Research Directions. *Information Systems Frontiers*, 1-16, <https://doi.org/10.1007/s10796-021-10142-8>.
- Wang, Y., & Kosinski, M. (2017). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of Personality and Social Psychology*, 114(2), 246–257.
- Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3–13.
- Wangmo, T., Lipps, M., Kressig, R. W., & Ienca, M. (2019). Ethical concerns with the use of intelligent assistive technology: findings from a qualitative study with professional stakeholders. *BMC Medical Ethics*, 20(1), 98.
- Wearn, O. R., Freeman, R., & Jacoby, D. M. (2019). Responsible AI for conservation. *Nature Machine Intelligence*, 1(2), 72–73.
- World Health Organization (2021). Ethics and governance of artificial intelligence for health. Available at: <https://www.who.int/publications/i/item/9789240029200>.
- Wright, S. A., & Schultz, A. E. (2018). The rising tide of artificial intelligence and business automation: Developing an ethical framework. *Business Horizons*, 61(6), 823–832.

- Xafis, V., Schaefer, G. O., Labude, M. K., Brassington, I., Ballantyne, A., Lim, H. Y., Lipworth, W., Lysaght, T., Stewart, C., Sun, S., Laurie, G. T., & Tai, E. S. (2019). An ethics framework for big data in health and research. *Asian Bioethics Review*, *11*(3), 227–254.
- Yew, G. C. K. (2020). Trust in and ethical design of carebots: the case for ethics of care. *International Journal of Social Robotics*, <https://doi.org/10.1007/s12369-020-00653-w>.
- Yüksel, B., Küpçü, A., & Özkasap, Ö. (2017). Research issues for privacy and security of electronic health services. *Future Generation Computer Systems*, *68*, 1–13.
- Zhang, J., & Hon, H. W. (2020). Towards responsible digital transformation. *California Management Review*, *62*(3).
- Zitnik, M., Nguyen, F., Wang, B., Leskovec, J., Goldenberg, A., & Hoffman, M. M. (2019). Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. *Information Fusion*, *50*, 71–91.

Appendices

Appendix A: Keywords and search strings

Using three sets of search keywords, we generated 96 search strings ($C_1^8 * C_1^{12} * C_1^1 = 96$). A spreadsheet was used to record the search results. Below, we provide a brief glimpse into the search strings used and results returned. The full list is available upon request.

NO	Search strings	Returns
1	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND ("artificial intelligence"[MeSH Terms] OR ("artificial"[All Fields] AND "intelligence"[All Fields]) OR "artificial intelligence"[All Fields]) AND "ethics"[All Fields] AND "health"[All Fields] AND (ff[Filter]))	672
2	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND ("artificial intelligence"[MeSH Terms] OR ("artificial"[All Fields] AND "intelligence"[All Fields]) OR "artificial intelligence"[All Fields]) AND "autonomy"[All Fields] AND "health"[All Fields] AND (ff[Filter]))	72
3	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND ("artificial intelligence"[MeSH Terms] OR ("artificial"[All Fields] AND "intelligence"[All Fields]) OR "artificial intelligence"[All Fields]) AND "justice"[All Fields] AND "health"[All Fields] AND (ff[Filter]))	56
4	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND ("artificial intelligence"[MeSH Terms] OR ("artificial"[All Fields] AND "intelligence"[All Fields]) OR "artificial intelligence"[All Fields]) AND "explicability"[All Fields] AND "health"[All Fields] AND (ff[Filter]))	3
5	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND ("artificial intelligence"[MeSH Terms] OR ("artificial"[All Fields] AND "intelligence"[All Fields]) OR "artificial intelligence"[All Fields]) AND "transparency"[All Fields] AND "health"[All Fields] AND (ff[Filter]))	121
6	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND ("artificial intelligence"[MeSH Terms] OR ("artificial"[All Fields] AND "intelligence"[All Fields]) OR "artificial intelligence"[All Fields]) AND "fairness"[All Fields] AND "health"[All Fields] AND (ff[Filter]))	11
7	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND ("artificial intelligence"[MeSH Terms] OR ("artificial"[All Fields] AND "intelligence"[All Fields]) OR "artificial intelligence"[All Fields]) AND "responsible"[All Fields] AND "health"[All Fields] AND (ff[Filter]))	228
8	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND ("artificial intelligence"[MeSH Terms] OR ("artificial"[All Fields] AND "intelligence"[All Fields]) OR "artificial intelligence"[All Fields]) AND "privacy"[All Fields] AND "health"[All Fields] AND (ff[Filter]))	318
9	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND ("artificial intelligence"[MeSH Terms] OR ("artificial"[All Fields] AND "intelligence"[All Fields]) OR "artificial intelligence"[All Fields]) AND "trust"[All Fields] AND "health"[All Fields] AND (ff[Filter]))	831
10	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND ("artificial intelligence"[MeSH Terms] OR ("artificial"[All Fields] AND "intelligence"[All Fields]) OR "artificial intelligence"[All Fields]) AND "sustainability"[All Fields] AND "health"[All Fields] AND (ff[Filter]))	71
11	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND ("artificial intelligence"[MeSH Terms] OR ("artificial"[All Fields] AND "intelligence"[All Fields]) OR "artificial intelligence"[All Fields]) AND "inclusive"[All Fields] AND "health"[All Fields] AND (ff[Filter]))	25
12	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND ("artificial intelligence"[MeSH Terms] OR ("artificial"[All Fields] AND "intelligence"[All Fields]) OR "artificial intelligence"[All Fields]) AND "accountability"[All Fields] AND "health"[All Fields] AND (ff[Filter]))	80
13	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "machine learning"[All Fields] AND "health"[All Fields] AND "ethics"[All Fields] AND (ff[Filter]))	348
14	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "machine learning"[All Fields] AND "health"[All Fields] AND "autonomy"[All Fields] AND (ff[Filter]))	24
15	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "machine learning"[All Fields] AND "health"[All Fields] AND "justice"[All Fields] AND (ff[Filter]))	43
16	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "machine learning"[All Fields] AND "health"[All Fields] AND "explicability"[All Fields] AND (ff[Filter]))	1
17	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "machine learning"[All Fields] AND "health"[All Fields] AND "transparency"[All Fields] AND (ff[Filter]))	76
18	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "machine learning"[All Fields] AND "health"[All Fields] AND "fairness"[All Fields] AND (ff[Filter]))	11
19	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "machine learning"[All Fields] AND "health"[All Fields] AND "responsible"[All Fields] AND (ff[Filter]))	129
20	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "machine learning"[All Fields] AND "health"[All Fields] AND "privacy"[All Fields] AND (ff[Filter]))	218
21	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "machine learning"[All Fields] AND "health"[All Fields] AND "trust"[All Fields] AND (ff[Filter]))	577
22	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "machine learning"[All Fields] AND "health"[All Fields] AND "sustainability"[All Fields] AND (ff[Filter]))	38

78	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "robotic"[All Fields] AND "health"[All Fields] AND "fairness"[All Fields] AND (fft[Filter])	0
79	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "robotic"[All Fields] AND "health"[All Fields] AND "responsible"[All Fields] AND (fft[Filter])	34
80	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "robotic"[All Fields] AND "health"[All Fields] AND "privacy"[All Fields] AND (fft[Filter])	3
81	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "robotic"[All Fields] AND "health"[All Fields] AND "trust"[All Fields] AND (fft[Filter])	224
82	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "robotic"[All Fields] AND "health"[All Fields] AND "sustainability"[All Fields] AND (fft[Filter])	10
83	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "robotic"[All Fields] AND "health"[All Fields] AND "inclusive"[All Fields] AND (fft[Filter])	8
84	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "robotic"[All Fields] AND "health"[All Fields] AND "accountability"[All Fields] AND (fft[Filter])	7
85	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "big data analytics"[All Fields] AND "health"[All Fields] AND "ethics"[All Fields] AND (fft[Filter])	19
86	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "big data analytics"[All Fields] AND "health"[All Fields] AND "autonomy"[All Fields] AND (fft[Filter])	0
87	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "big data analytics"[All Fields] AND "health"[All Fields] AND "justice"[All Fields] AND (fft[Filter])	3
88	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "big data analytics"[All Fields] AND "health"[All Fields] AND "explicability"[All Fields] AND (fft[Filter])	0
89	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "big data analytics"[All Fields] AND "health"[All Fields] AND "transparency"[All Fields] AND (fft[Filter])	5
90	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "big data analytics"[All Fields] AND "health"[All Fields] AND "fairness"[All Fields] AND (fft[Filter])	2
91	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "big data analytics"[All Fields] AND "health"[All Fields] AND "responsible"[All Fields] AND (fft[Filter])	6
92	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "big data analytics"[All Fields] AND "health"[All Fields] AND "privacy"[All Fields] AND (fft[Filter])	25
93	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "big data analytics"[All Fields] AND "health"[All Fields] AND "trust"[All Fields] AND (fft[Filter])	11
94	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "big data analytics"[All Fields] AND "health"[All Fields] AND "sustainability"[All Fields] AND (fft[Filter])	7
95	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "big data analytics"[All Fields] AND "health"[All Fields] AND "inclusive"[All Fields] AND (fft[Filter])	0
96	(2000/01/01:2020/12/31[Date - Publication] AND "english"[Language] AND "big data analytics"[All Fields] AND "health"[All Fields] AND "accountability"[All Fields] AND (fft[Filter])	2
97		
98		
99		
100		
101		
102		
103		
104		

Sheet1

Ready

Type here to search

Figure 1. The record of search strings (Snapshot)

Appendix B: Exclusion criteria

No.	Exclusion criteria	Reason for exclusion
1	duplicates	Duplicated records need to be removed.
2	Focusing on technical aspects of AI in healthcare	In these articles, AI applications are not discussed directly from the point of view of stakeholders (e.g., patients or providers) and there are no explanations offered on how AI can be implemented in healthcare in an ethical manner. Typically, these excluded articles focus explicitly on technical aspects of AI in healthcare such as mathematic formulae, finely grained scientific computing models, software engineering models, programming techniques, specific hardware architectures, etc. Thus, they would be precluded from the review process.
3	Focusing on digital security and cryptography	This is now a pervasively known and clichéd proposed solution to address the privacy and security concerns and issue surrounding system use and including it in the systematic literature review will not add much value to the critical analysis and findings.
4	Do not focus on healthcare sectors	The review process excludes articles that do not focus on healthcare sectors.
5	Source creditability	Articles published as conference papers/proceedings, books, book chapters, thesis were excluded from the review. The preprint version of articles published in open online repositories and articles from unknown sources were also excluded.

Appendix C: Relevance assessment criteria

We included the articles that provide insights and practical implications on AI ethics in health care and demonstrate ways to deploy AI in a responsible and ethical way. After reading the full text of each article, studies were organized into three lists: List A includes studies that were deemed to be relevant, list B comprises studies where the relevance was not clear a priori, and list C includes studies where the research was deemed as either irrelevant or ambiguous. Out of the 1,245 peer-reviewed articles, 652 were less relevant (List C), 269 were partially relevant (List B), and 324 were relevant (List A). Out of the 324 relevant articles, 253 articles were found to be directly relevant to our research scope and context and consequently, they were analyzed using NVivo to identify recurrent themes.

Element	Level				
	0 – Absence	1 – Low	2 – Medium	3 – High	Not applicable
Theory robustness	The article does not provide enough information to assess this criterion.	Weak development of theoretical insights and limited awareness of prevailing literature.	Basic development of theory and use of concepts garnered from existing literature.	Good use of theory, including the novel and provocative development of concepts.	This element is not relevant to the study.
Implication for practice	The article does not provide enough information to assess this criterion.	Hard to use the concepts and ideas in pragmatic problem solving.	The findings and observations have potential use for businesses and policymakers.	The use for practitioners is clear.	This element is not relevant to the study.
Methodology, data supporting arguments	The article does not provide enough information to assess this criterion.	Data incomplete and not related to theory and weak research design.	Data broadly related to the arguments and conveyed through a clear research design.	Data strongly supports arguments. Robust research design.	This element is not relevant to the study.
Relevance of three areas: findings; theories; methods	The article does not provide enough information to assess this criterion.	Only tangentially relevant; provocative but linked to “line of flight”.	Broadly relevant – perhaps in one area or applied in different disciplinary field.	High level of relevance across findings, methods, and theoretical constructs/concepts.	This element is not relevant to the study.
Contribution	The article does not provide enough information to assess this criterion.	Does not make an important contribution. It is not clear the advances it makes.	Although using other ideas, it builds on existing theory.	Further develops existing knowledge, expanding issues.	This element is not relevant to the study.

Appendix D: Full List of References in Review

- Aboueid, S., Liu, R. H., Desta, B. N., Chaurasia, A., & Ebrahim, S. (2019). The use of artificially intelligent self-diagnosing digital platforms by the general public: scoping review. *JMIR Medical Informatics*, 7(2), e13445.
- Abràmoff, M. D., Leng, T., Ting, D. S. W., Rhee, K., Horton, M. B., Brady, C. J., & Chiang, M. F. (2020). Automated and computer-assisted detection, classification, and diagnosis of diabetic retinopathy. *Telemedicine and e-Health*, 26(4), 544–550.
- Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database*, <https://doi.org/10.1093/database/baaa010>.
- Aitken, M., Tully, M. P., Porteous, C., Denegri, S., Cunningham-Burley, S., Banner, N., ... & Willison, D. J. (2019). Consensus statement on public involvement and engagement with data intensive health research. *International Journal of Population Data Science*, 4(1), 586.
- Alami, H., Gagnon, M.-P., & Fortin, J.-P. (2017). Digital health and the challenge of health systems transformation. *Mhealth*, 3, 31.
- Alami, H., Gagnon, M.-P., & Fortin, J.-P. (2019). Some multidimensional unintended consequences of telehealth utilization: a multi-project evaluation synthesis. *International Journal of Health Policy and Management*, 8(6), 337–352.
- Alami, H., Lehoux, P., Auclair, Y., de Guise, M., Gagnon, M.-P., Shaw, J., Roy, D., Fleet, R., Ahmed, M. A. A., & Fortin, J.-P. (2020a). Artificial intelligence and health technology assessment: anticipating a new level of complexity. *Journal of Medical Internet Research*, 22(7), e17707.
- Alami, H., Rivard, L., Lehoux, P., Hoffman, S. J., Cadeddu, S. B. M., Savoldelli, M., Samri, M. A., Ahmed, M. A. A., Fleet, R., & Fortin, J.-P. (2020b). Artificial intelligence in health care: laying the Foundation for Responsible, sustainable, and inclusive innovation in low- and middle-income countries. *Globalization and Health*, 16(1), 52.
- Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, 20(1), 1-9.
- Amarasingham, R., Patzer, R. E., Huesch, M., Nguyen, N. Q., & Xie, B. (2014). Implementing electronic health care predictive analytics: considerations and challenges. *Health Affairs*, 33(7), 1148-1154.
- Angehrn, Z., Haldna, L., Zandvliet, A. S., Gil Berglund, E., Zeeuw, J., Amzal, B., ... & Heckman, N. M. (2020). Artificial intelligence and machine learning applied at the point of care. *Frontiers in Pharmacology*, 11, 759.
- Ashrafian, H., Darzi, A., & Athanasiou, T. (2015). A novel modification of the Turing test for artificial intelligence and robotics in healthcare. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 11(1), 38-43.
- Astromskė, K., Peičius, E., & Astromskis, P. (2020). Ethical and legal challenges of informed consent applying artificial intelligence in medical diagnostic consultations. *AI & SOCIETY*, <https://doi.org/10.1007/s00146-020-01008-9>.
- Azencott, C. A. (2018). Machine learning and genomics: precision medicine versus patient privacy. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2128), 20170350.
- Bærøe, K., Miyata-Sturm, A., & Henden, E. (2020). How to achieve trustworthy artificial intelligence for health. *Bulletin of the World Health Organization*, 98(4), 257-262.
- Baig, M. A., Almuhaizea, M. A., Alshehri, J., Bazarbashi, M. S., & Al-Shagathrh, F. (2020). Urgent need for developing a framework for the governance of AI in healthcare. *Studies in Health Technology and Informatics*, 272, 253–256.

- Balthazar, P., Harri, P., Prater, A., & Safdar, N. M. (2018). Protecting your patients' interests in the era of big data, artificial intelligence, and predictive analytics. *Journal of the American College of Radiology*, 15(3), 580-586.
- Banks, J. (2018). The human touch: practical and ethical implications of putting AI and robotics to work for patients. *IEEE Pulse*, 9(3), 15-18.
- Baric-Parker, J., & Anderson, E. E. (2020). Patient data sharing for AI: ethical challenges, Catholic solutions. *The Linacre Quarterly*, 87(4), 471–481.
- Basu, T., Engel-Wolf, S., & Menzer, O. (2020). The ethics of machine learning in medical sciences: Where do we stand today?. *Indian Journal of Dermatology*, 65(5), 358.
- Bauer, M., Glenn, T., Monteith, S., Bauer, R., Whybrow, P. C., & Geddes, J. (2017). Ethical perspectives on recommending digital technology for patients with mental illness. *International Journal of Bipolar Disorders*, 5(1), 1-14.
- Becker, A. (2019). Artificial intelligence in medicine: What is it doing for us today?. *Health Policy and Technology*, 8(2), 198-205.
- Beil, M., Proft, I., van Heerden, D., Sviri, S., & van Heerden, P. V. (2019). Ethical considerations about artificial intelligence for prognostication in intensive care. *Intensive Care Medicine Experimental*, 7(1), 1-13.
- Benke, K., & Benke, G. (2018). Artificial intelligence and big data in public health. *International Journal of Environmental Research and Public Health*, 15(12), 2796.
- Bickmore, T. W., Mitchell, S. E., Jack, B. W., Paasche-Orlow, M. K., Pfeifer, L. M., & O'Donnell, J. (2010). Response to a relational agent by hospital patients with depressive symptoms. *Interacting with Computers*, 22(4), 289–298.
- Bjerring, J. C., & Busch, J. (2020). Artificial intelligence and patient-centered decision-making. *Philosophy & Technology*, <https://doi.org/10.1007/s13347-019-00391-6>.
- Blease, C., Kaptchuk, T. J., Bernstein, M. H., Mandl, K. D., Halamka, J. D., & DesRoches, C. M. (2019). Artificial Intelligence and the future of primary care: exploratory qualitative study of UK general practitioners' views. *Journal of Medical Internet Research*, 21(3), e12802. doi:10.2196/12802.
- Blobel, B., Ruotsalainen, P., Brochhausen, M., Oemig, F., & Uribe, G. A. (2020). Autonomous systems and artificial intelligence in healthcare transformation to 5P medicine – ethical challenges. *Studies in Health Technology and Informatics*, 270, 1089–1093.
- Bonifazi, F., Volpe, E., Digregorio, G., Giannuzzi, V., & Ceci, A. (2020). Machine Learning Systems Applied to Health Data and System. *European Journal of Health Law*, 27(3), 242-258.
- Brady, A. P., & Neri, E. (2020). Artificial intelligence in radiology—ethical considerations. *Diagnostics*, 10(4), 231.
- Briganti, G., & Le Moine, O. (2020). Artificial Intelligence in medicine: today and tomorrow. *Frontiers in Medicine*, 7, 27.
- Bukowski, M., Farkas, R., Beyan, O., Moll, L., Hahn, H., Kiessling, F., & Schmitz-Rode, T. (2020). Implementation of eHealth and AI integrated diagnostics with multidisciplinary digitized data: are we ready from an international perspective? *European Radiology*, 30(10), 5510–5524.
- Burnside, M., Crocket, H., Mayo, M., Pickering, J., Tappe, A., & de Bock, M. (2019). Do-it-yourself automated insulin delivery: a leading example of the democratization of medicine. *Journal of Diabetes Science and Technology*, 14(5), 878–882.
- Buruk, B., Ekmekci, P. E., & Arda, B. (2020). A critical perspective on guidelines for responsible and trustworthy artificial intelligence. *Medicine, Health Care and Philosophy*, 23(3), 387-399.
- Cabitza, F., Rasoini, R., & Gensini, G. F. (2017). Unintended consequences of machine learning in medicine. *JAMA*, 318(6), 517-518.
- Carter, S. M., Rogers, W., Win, K. T., Frazer, H., Richards, B., & Houssami, N. (2020). The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. *The Breast*, 49, 25–32.
- Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T., & Tsaneva-Atanasova, K. (2019). Artificial intelligence, bias and clinical safety. *BMJ Quality & Safety*, 28(3), 231-237.
- Char, D. S., Abràmoff, M. D., & Feudtner, C. (2020). Identifying ethical considerations for machine learning healthcare applications. *The American Journal of Bioethics*, 20(11), 7-17.

- Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care – addressing ethical challenges. *The New England Journal of Medicine*, 378(11), 981–983.
- Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., & Ghassemi, M. (2020). Ethical Machine Learning in Healthcare. *Annual Review of Biomedical Data Science*, 4.
- Chen, I. Y., Szolovits, P., & Ghassemi, M. (2019). Can AI help reduce disparities in general medical and mental health care?. *AMA Journal of Ethics*, 21(2), 167-179.
- Chen, J. H., & Asch, S. M. (2017). Machine learning and prediction in medicine — beyond the peak of inflated expectations. *New England Journal of Medicine*, 376(26), 2507–2509.
- Christie, G. (2018). Progressing the health agenda: responsibly innovating in health technology. *Journal of Responsible Innovation*, 5(1), 143–148.
- Coeckelbergh, M., Pop, C., Simut, R., Peca, A., Pintea, S., David, D., & Vanderborcht, B. (2016). A survey of expectations about the role of robots in robot-assisted therapy for children with ASD: ethical acceptability, trust, sociability, appearance, and attachment. *Science and Engineering Ethics*, 22(1), 47–65.
- Cohen, I. G., Amarasingham, R., Shah, A., Xie, B., & Lo, B. (2014). The legal and ethical concerns that arise from using complex predictive analytics in health care. *Health Affairs*, 33(7), 1139–1147.
- Coiera, E. (2019). On algorithms, machines, and medicine. *The Lancet Oncology*, 20(2), 166-167.
- Combs, C. D., & Combs, P. F. (2019). Emerging roles of virtual patients in the age of AI. *AMA Journal of Ethics*, 21(2), E153-159.
- Cornet, G. (2013). Robot companions and ethics: A pragmatic approach of ethical design. *Journal International De Bioéthique*, 24(4), 49-58.
- Corsico, P. (2019). The risks of risk. Regulating the use of machine learning for psychosis prediction. *International Journal of Law and Psychiatry*, 66, 101479.
- Cresswell, K., Cunningham-Burley, S., & Sheikh, A. (2018). Health care robotics: qualitative exploration of key challenges and future directions. *Journal of Medical Internet Research*, 20(7), e10410
- Currie, G., Hawk, K. E., & Rohren, E. M. (2020). Ethical principles for the application of artificial intelligence (AI) in nuclear medicine. *European Journal of Nuclear Medicine and Molecular Imaging*, 47, 748-752.
- Cuzzolin, F., Morelli, A., Cirstea, B., & Sahakian, B. J. (2020). Knowing me, knowing you: theory of mind in AI. *Psychological Medicine*, 50(7), 1057-1061.
- D’Antonoli, T. A. (2020). Ethical considerations for artificial intelligence: an overview of the current radiology landscape. *Diagnostic and Interventional Radiology*, 26(5), 504-511.
- Dalton-Brown, S. (2020). The ethics of medical AI and the physician-patient relationship. *Cambridge Quarterly of Healthcare Ethics*, 29(1), 115-121.
- Darcy, A. M., Louie, A. K., & Roberts, L. W. (2016). Machine learning and the profession of medicine. *JAMA*, 315(6), 551-552.
- Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, 6(2), 94-98.
- de Miguel, I., Sanz, B., & Lazcoz, G. (2020). Machine learning in the EU health care context: exploring the ethical, legal and social issues. *Information, Communication & Society*, 23(8), 1139-1153.
- Deo, R. C. (2015). Machine learning in medicine. *Circulation*, 132(20), 1920-1930.
- Di Ieva, A. (2019). AI-augmented multidisciplinary teams: hype or hope?. *The Lancet*, 394(10211), 1801.
- Di Nucci, E. (2019). Should we be afraid of medical AI? *Journal of Medical Ethics*, 45, 556-558.
- Diprose, W., & Buist, N. (2016). Artificial intelligence in medicine: humans need not apply? *The New Zealand Medical Journal*, 129(1434), 73.
- Drew, L. (2019). The ethics of brain-computer interfaces. *Nature*, 571, S19.
- Dzobo, K., Adotey, S., Thomford, N. E., & Dzobo, W. (2020). Integrating artificial and human intelligence: a partnership for responsible innovation in biomedical engineering and medicine. *Omics: A Journal of Integrative Biology*, 24(5), 247-263.
- Eaneff, S., Obermeyer, Z., & Butte, A. J. (2020). The case for algorithmic stewardship for artificial intelligence and machine learning technologies. *JAMA*, 324(14), 1397-1398.

- Eduard, F.-V., & Jordi, A.-C. (2019). 'I'll take care of you,' said the robot. *Paladyn Journal of Behavioral Robotics*, 10(1), 77–93.
- Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, 28(1), 62–70.
- Ferretti, A., Schneider, M., & Blasimme, A. (2018). Machine learning in medicine: opening the new data protection black box. *European Data Protection Law Review*, 4, 320.
- Fiske, A., Buyx, A., & Prainsack, B. (2019a). Health information counselors: A new profession for the age of big data. *Academic Medicine: Journal of the Association of American Medical Colleges*, 94(1), 37–41.
- Fiske, A., Henningsen, P., & Buyx, A. (2019b). Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medical Internet Research*, 21(5), e13216.
- Flaxman, A. D., & Vos, T. (2018). Machine learning in population health: opportunities and threats. *PLoS Medicine*, 15(11), e1002702.
- Fletcher, R. R., Nakeshimana, A., & Olubeko, O. (2020). Addressing Fairness, Bias, and Appropriate Use of Artificial Intelligence and Machine Learning in Global Health. *Frontiers in Artificial Intelligence*, doi: 10.3389/frai.2020.561802.
- Floridi, L., Luetge, C., Pagallo, U., Schafer, B., Valcke, P., Vayena, E., ... & Kalra, D. (2019). Key ethical challenges in the European medical information framework. *Minds and Machines*, 29(3), 355-371.
- Forsting, M. (2017). Machine learning will change medicine. *Journal of Nuclear Medicine*, 58(3), 357-358.
- Garattini, C., Raffle, J., Aisyah, D. N., Sartain, F., & Kozlakidis, Z. (2019). Big data analytics, infectious diseases and associated ethical impacts. *Philosophy & Technology*, 32(1), 69-85.
- Garbuio, M., & Lin, N. (2019). Artificial intelligence as a growth engine for health care startups: Emerging business models. *California Management Review*, 61(2), 59-83.
- Garden, H., Bowman, D. M., Haesler, S., & Winickoff, D. E. (2016). Neurotechnology and society: strengthening responsible innovation in brain science. *Neuron*, 92(3), 642–646.
- Gauld, C., Micoulaud-Franchi, J. A., & Dumas, G. (2020). Comment on Starke et al.: 'Computing schizophrenia: ethical challenges for machine learning in psychiatry': from machine learning to student learning: pedagogical challenges for psychiatry. *Psychological Medicine*, 1-3.
- Geis, J. R., Brady, A. P., Wu, C. C., Spencer, J., Ranschaert, E., Jaremko, J. L., ... & Kohli, M. (2019). Ethics of artificial intelligence in radiology: summary of the joint European and North American multisociety statement. *Canadian Association of Radiologists Journal*, 70(4), 329-334.
- Gianfrancesco, M. A., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential biases in machine learning algorithms using electronic health record data. *JAMA Internal Medicine*, 178(11), 1544-1547.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012–1014.
- Gray, E. A., & Thorpe, J. H. (2015). Comparative effectiveness research and big data: balancing potential with legal and ethical considerations. *Journal of Comparative Effectiveness Research*, 4(1), 61-74.
- Grote, T., & Berens, P. (2020). On the ethics of algorithmic decision-making in healthcare. *Journal of Medical Ethics*, 46(3), 205-211.
- Gruson, D. (2020). Positive ethical regulation of digital technology in care plan. *Soins; la revue de reference infirmiere*, 65(842), 41–45.
- Gruson, D., Helleputte, T., Rousseau, P., & Gruson, D. (2019). Data science, artificial intelligence, and machine learning: opportunities for laboratory medicine and the value of positive regulation. *Clinical Biochemistry*, 69, 1-7.
- Guan, J. (2019). Artificial Intelligence in healthcare and medicine: promises, ethical challenges and governance. *Chinese Medical Sciences Journal*, 34(2), 76–83.
- Hadley, T. D., Pettit, R. W., Malik, T., Khoei, A. A., & Salihu, H. M. (2020). Artificial intelligence in global health—a framework and strategy for adoption and sustainability. *International Journal of Maternal and Child Health and AIDS*, 9(1), 121.

- Hallowell, N., Parker, M., & Nellåker, C. (2019). Big data phenotyping in rare diseases: some ethical issues. *Genetics in Medicine*, 21(2), 272-274.
- Handelman, G. S., Kok, H. K., Chandra, R. V., Razavi, A. H., Huang, S., Brooks, M., ... & Asadi, H. (2019). Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods. *American Journal of Roentgenology*, 212(1), 38-43.
- He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., & Zhang, K. (2019). The practical implementation of artificial intelligence technologies in medicine. *Nature Medicine*, 25(1), 30-36.
- Henriksen, A., & Bechmann, A. (2020). Building truths in AI: Making predictive algorithms doable in healthcare. *Information, Communication & Society*, 23(6), 802-816.
- Ho, A. (2019). Deep ethical learning: Taking the interplay of human and artificial intelligence seriously. *Hastings Center Report*, 49(1), 36-39.
- Ho, A. (2020a). Are we ready for artificial intelligence health monitoring in elder care?. *BMC Geriatrics*, 20(1), 1-7.
- Ho, C. W. L. (2020b). Deepening the normative evaluation of machine learning healthcare application by complementing ethical considerations with regulatory governance. *The American Journal of Bioethics*, 20(11), 43-45.
- Hofer, I. S., Burns, M., Kendale, S., & Wanderer, J. P. (2020). Realistically integrating machine learning into clinical practice: a road map of opportunities, challenges, and a potential future. *Anesthesia and Analgesia*, 130(5), 1115-1118.
- Holzinger, A. (2014). Trends in interactive knowledge discovery for personalized medicine: cognitive science meets machine learning. *The IEEE Intelligent Informatics Bulletin*, 15(1), 6-14.
- Holzinger, A., Haibe-Kains, B., & Jurisica, I. (2019). Why imaging data alone is not enough: AI-based integration of imaging, omics, and clinical data. *European Journal of Nuclear Medicine and Molecular Imaging*, 46(13), 2722-2730.
- Hoorn, J. F., & Winter, S. D. (2018). Here comes the bad news: doctor robot taking over. *International Journal of Social Robotics*, 10(4), 519-535.
- Horgan, D., Romao, M., Morr e, S. A., & Kalra, D. (2019). Artificial Intelligence: power for civilisation – and for better healthcare. *Public Health Genomics*, 22(5-6), 145-161.
- Hosny, A., & Aerts, H. J. (2019). Artificial intelligence for global health. *Science*, 366(6468), 955-956.
- Hummel, P., & Braun, M. (2020). Just data? Solidarity and justice in data-driven medicine. *Life Sciences, Society and Policy*, 16(1), 1-18.
- Ibrahim, S. A., Charlson, M. E., & Neill, D. B. (2020). Big data analytics and the struggle for equity in health care: the promise and perils. *Health Equity*. <https://doi.org/10.1089/heap.2019.0112>.
- Ienca, M., & Ignatiadis, K. (2020). Artificial Intelligence in clinical neuroscience: methodological and ethical challenges. *AJOB Neuroscience*, 11(2), 77-87. doi:10.1080/21507740.2020.1740352.
- Jacobson, N. C., Bentley, K. H., Walton, A., Wang, S. B., Fortgang, R. G., Millner, A. J., ... & Coppersmith, D. D. (2020). Ethical dilemmas posed by mobile health and machine learning in psychiatry research. *Bulletin of the World Health Organization*, 98(4), 270-276.
- Jeste, D. V., Graham, S. A., Nguyen, T. T., Depp, C. A., Lee, E. E., & Kim, H.-C. (2020). Beyond artificial intelligence: exploring artificial wisdom. *International Psychogeriatrics*, 32(8), 993-1001.
- Jim enez-Luna, J., Grisoni, F., & Schneider, G. (2020). Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence*, 2(10), 573-584.
- Joerin, A., Rauws, M., Fulmer, R., & Black, V. (2020). Ethical artificial intelligence for digital health organizations. *Cureus*, 12(3), e7202.
- Johnson, S. L. (2019). AI, machine learning, and ethics in health care. *Journal of Legal Medicine*, 39(4), 427-441.
- Johnson, S. L. (2020). AI and machine learning in medicine: ethical considerations. *Current Allergy & Clinical Immunology*, 33(2), 114-117.
- Jung, H. H., & Pfister, F. M. (2020). Blockchain-enabled clinical study consent management. *Technology Innovation Management Review*, 10(2), 14-24.
- Kaissis, G. A., Makowski, M. R., R uckert, D., & Braren, R. F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, 2(6), 305-311.

- Kandalaft, M. R., Didehbani, N., Krawczyk, D. C., Allen, T. T., & Chapman, S. B. (2013). Virtual reality social cognition training for young adults with high-functioning autism. *Journal of Autism and Developmental Disorders*, 43(1), 34–44.
- Kaplan, B. (2016). How should health data be used?: Privacy, secondary use, and big data sales. *Cambridge Quarterly of Healthcare Ethics*, 25(2), 312-329.
- Kayaalp, M. (2018). Patient privacy in the era of big data. *Balkan Medical Journal*, 35(1), 8-17.
- Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, 17(1), 1-9.
- Keskinbora, K. H., & Keskinbora, K. (2018). Ethical considerations on novel neuronal interfaces. *Neurological Sciences*, 39(4), 607–613.
- Khatoun, S., Rahman, S. M. M., Alrubaian, M., & Alamri, A. (2019). Privacy-preserved, provable secure, mutually authenticated key agreement protocol for healthcare in a smart city environment. *IEEE Access*, 7, 47962–47971.
- Kim, E., Caraballo, P. J., Castro, M. R., Pieczkiewicz, D. S., & Simon, G. J. (2019). Towards more accessible precision medicine: building a more transferable machine learning model to support prognostic decisions for micro-and macrovascular complications of type 2 diabetes mellitus. *Journal of Medical Systems*, 43(7), 1-12.
- Kohli, M., & Geis, R. (2018). Ethics, Artificial Intelligence, and Radiology, *Journal of the American College of Radiology*, 15(9), 1317-1319.
- Kolachalama, V. B., & Garg, P. S. (2018). Machine learning and medical education. *NPJ Digital Medicine*, 1(1), 1-3.
- Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, 23(1), 89-109.
- Krutzinna, J., Taddeo, M., & Floridi, L. (2019). Enabling posthumous medical data donation: an appeal for the ethical utilisation of personal health data. *Science and Engineering Ethics*, 25(5), 1357-1387.
- Larson, D. B., Magnus, D. C., Lungren, M. P., Shah, N. H., & Langlotz, C. P. (2020). Ethics of using and sharing clinical imaging data for artificial intelligence: a proposed framework. *Radiology*, 295(3), 192536.
- Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *Science*, 343(6176), 1203–1205.
- Li, D., Kulasegaram, K., & Hodges, B. D. (2019). Why we needn't fear the machines: opportunities for medicine in a machine learning world. *Academic Medicine*, 94(5), 623-625.
- Liaw, S. T., Liyanage, H., Kuziemy, C., Terry, A. L., Schreiber, R., Jonnagaddala, J., & de Lusignan, S. (2020). Ethical use of electronic health record data and artificial intelligence: recommendations of the primary care informatics working group of the international medical informatics association. *Yearbook of Medical Informatics*, 29(1), 51-57.
- Licitra, L., Trama, A., & Hosni, H. (2017). Benefits and risks of machine learning decision support systems. *JAMA*, 318(23), 2354-2354.
- Liew, C. (2018). The future of radiology augmented with artificial intelligence: a strategy for success. *European Journal of Radiology*, 102, 152-156.
- Lisboa, P. J. (2002). A review of evidence of health benefit from artificial neural networks in medical intervention. *Neural Networks*, 15(1), 11-39.
- Littmann, M., Selig, K., Cohen-Lavi, L., Frank, Y., Hönigschmid, P., Kataka, E., ... & Rost, B. (2020). Validity of machine learning in biology and medicine increased through collaborations across fields of expertise. *Nature Machine Intelligence*, 2(1), 18-24.
- Loh, E. (2018). Medicine and the rise of the robots: a qualitative review of recent advances of artificial intelligence in health. *BMJ Leader*, leader-2018-000071.
- Lupton, M. (2018). Some ethical and legal consequences of the application of artificial intelligence in the field of medicine. *Trends in Medicine*, 18(4), 1-7.
- Luxton, D. D. (2014). Recommendations for the ethical use and design of artificial intelligent care providers. *Artificial Intelligence in Medicine*, 62(1), 1–10.
- Ma, H., Guo, X., Ping, Y., Wang, B., Yang, Y., Zhang, Z., & Zhou, J. (2019). PPCD: privacy-preserving clinical decision with cloud support. *Plos One*, 14(5), e0217349.

- Mahomed, S. (2018). Healthcare, artificial intelligence and the Fourth Industrial Revolution: Ethical, social and legal considerations. *South African Journal of Bioethics and Law*, 11(2), 93-95.
- Malkin, R., & von Oldenburg Beer, K. (2013). Diffusion of novel healthcare technologies to resource poor settings. *Annals of Biomedical Engineering*, 41(9), 1841–1850.
- Manrique de Lara, A., & Peláez-Ballestas, I. (2020). Big data and data processing in rheumatology: bioethical perspectives. *Clinical Rheumatology*, 39(4), 1007–1014.
- Martinez-Martin, N. (2019). What are important ethical implications of using facial recognition technology in health care?. *AMA Journal of Ethics*, 21(2), E180-E187.
- Martinez-Martin, N., & Kreitmair, K. (2018). Ethical issues for direct-to-consumer digital psychotherapy apps: addressing accountability, data protection, and consent. *JMIR Mental Health*, 5(2), e32.
- Martinez-Martin, N., Dunn, L. B., & Roberts, L. W. (2018). Is it ethical to use prognostic estimates from machine learning to treat psychosis?. *AMA Journal of Ethics*, 20(9), E804-E811.
- Martinez-Martin, N., Luo, Z., Kaushal, A., Adeli, E., Haque, A., Kelly, S. S., ... & Milstein, A. (2020). Ethical issues in using ambient intelligence in health-care settings. *The Lancet Digital Health*. [https://doi.org/10.1016/S2589-7500\(20\)30275-2](https://doi.org/10.1016/S2589-7500(20)30275-2).
- Matheny, M. E., Whicher, D., & Israni, S. T. (2020). Artificial intelligence in health care: a report from the National Academy of Medicine. *JAMA*, 323(6), 509–510.
- Mathur, P., Srivastava, S., Xu, X., & Mehta, J. L. (2020). Artificial intelligence, machine learning, and cardiovascular disease. *Clinical Medicine Insights: Cardiology*, 14, 1179546820927404.
- Mazurowski, M. A. (2020). Artificial intelligence in radiology: some ethical considerations for radiologists and algorithm developers. *Academic Radiology*, 27(1), 127-129.
- McCoy, L. G., Nagaraj, S., Morgado, F., Harish, V., Das, S., & Celi, L. A. (2020). What do medical students actually need to know about artificial intelligence?. *NPJ Digital Medicine*, 3(1), 1-3.
- McCraden, M. D., Baba, A., Saha, A., Ahmad, S., Boparai, K., Fadaiefard, P., & Cusimano, M. D. (2020a). Ethical concerns around use of artificial intelligence in health care research from the perspective of patients with meningioma, caregivers and health care providers: a qualitative study. *CMAJ Open*, 8(1), E90-E95.
- McCraden, M. D., Joshi, S., Anderson, J. A., Mazwi, M., Goldenberg, A., & Zlotnik Shaul, R. (2020b). Patient safety and quality improvement: Ethical principles for a regulatory approach to bias in healthcare machine learning. *Journal of the American Medical Informatics Association*, 27(12), 2024-2027.
- McCraden, M. D., Joshi, S., Mazwi, M., & Anderson, J. A. (2020c). Ethical limitations of algorithmic fairness solutions in health care machine learning. *The Lancet Digital Health*, 2(5), e221-e223.
- McGraw, D., & Petersen, C. (2020). From commercialization to accountability: responsible health data collection, use, and disclosure for the 21st century. *Applied Clinical Informatics*, 11(02), 366-373.
- Mehta, M. C., Katz, I. T., & Jha, A. K. (2020). Transforming global health with AI. *New England Journal of Medicine*, 382(9), 791-793.
- Meskó, B., Hetényi, G., & Györffy, Z. (2018). Will artificial intelligence solve the human resource crisis in healthcare? *BMC Health Services Research*, 18(1), 545. doi:10.1186/s12913-018-3359-4.
- Miller, D. D. (2020). Machine intelligence in cardiovascular medicine. *Cardiology in Review*, 28(2), 53–64.
- Miller, E., & Polson, D. (2019). Apps, avatars, and robots: the future of mental healthcare. *Issues in Mental Health Nursing*, 40(3), 208–214.
- Minssen, T., Gerke, S., Aboy, M., Price, N., & Cohen, G. (2020). Regulatory responses to medical machine learning. *Journal of Law and the Biosciences*, 7(1), lsa002.
- Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics*, 19(6), 1236-1246.
- Mittelman, M., Markham, S., & Taylor, M. (2018). Patient commentary: stop hyping artificial intelligence – patients will always need human doctors. *British Medical Journal*, 363, k4669.
- Mittelstadt, B. (2017a). Designing the health-related internet of things: ethical principles and guidelines. *Information*, 8(3), 77.
- Mittelstadt, B. (2017b). Ethics of the health-related internet of things: a narrative review. *Ethics and Information Technology*, 19(3), 157–175.

- Mittelstadt, B. D., & Floridi, L. (2016). The ethics of big data: current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics*, 22(2): 303–341.
- Mollura, D. J., Culp, M. P., Pollack, E., Battino, G., Scheel, J. R., Mango, V. L., ... & Dako, F. (2020). Artificial intelligence in low-and middle-income countries: innovating global health radiology. *Radiology*, 297(3), 513-520.
- Moore, J. H., Boland, M. R., Camara, P. G., Chervitz, H., Gonzalez, G., Himes, B. E., ... & Holmes, J. H. (2019). Preparing next-generation scientists for biomedical big data: artificial intelligence approaches. *Personalized Medicine*, 16(3), 247-257.
- Morley, J., Machado, C. C. V., Burr, C., Cows, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: a mapping review. *Social Science & Medicine*, 260, 113172.
- Mudgal, K. S., & Das, N. (2020). The ethical adoption of artificial intelligence in radiology. *BJR| Open*, 2(1), 20190020.
- Mullainathan, S., & Obermeyer, Z. (2019). Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care (No. w26168). National Bureau of Economic Research.
- Nagaraj, S., Harish, V., McCoy, L. G., Morgado, F., Stedman, I., Lu, S., ... & Singh, D. (2020). From Clinic to Computer and Back Again: Practical Considerations When Designing and Implementing Machine Learning Solutions for Pediatrics. *Current Treatment Options in Pediatrics*, 1-14.
- Nagendran, M., Chen, Y., Lovejoy, C. A., Gordon, A. C., Komorowski, M., Harvey, H., Topol, E. J., Ioannidis, J. P. A., Collins, G. S., & Maruthappu, M. (2020). Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ*, 368, m689.
- Nebeker, C., Torous, J., & Bartlett Ellis, R. J. (2019). Building the case for actionable ethics in digital health research supported by artificial intelligence. *BMC Medicine*, 17(1), 137.
- Neri, E., Coppola, F., Miele, V., Bibbolino, C., & Grassi, R. (2020). Artificial intelligence: Who is responsible for the diagnosis?. *La radiologia medica*, 125, 517-521.
- Ngiam, K. Y., & Khor, W. (2019). Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*, 20(5), e262-e273.
- Noorbakhsh-Sabet, N., Zand, R., Zhang, Y., & Abedi, V. (2019). Artificial intelligence transforms the future of health care. *The American Journal of Medicine*, 132, 795–801.
- Norgeot, B., Quer, G., Beaulieu-Jones, B. K., Torkamani, A., Dias, R., Gianfrancesco, M., ... & Butte, A. J. (2020). Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. *Nature Medicine*, 26(9), 1320-1324.
- Noseworthy, P. A., Attia, Z. I., Brewer, L. C., Hayes, S. N., Yao, X., Kapa, S., Friedman, P. A., & Lopez-Jimenez, F. (2020). Assessing and mitigating bias in medical artificial intelligence. *Circulation: Arrhythmia and Electrophysiology*, 13(3), e007988.
- Nwosu, A. C., Sturgeon, B., McGlinchey, T., Goodwin, C. D. G., Behera, A., Mason, S., Stanley, S., & Payne, T. R. (2019). Robotic technology for palliative and supportive care: strengths, weaknesses, opportunities and threats. *Palliative Medicine*, 33(8), 1106–1113.
- O'Reilly-Shah, V. N., Gentry, K. R., Walters, A. M., Zivot, J., Anderson, C. T., & Tighe, P. J. (2020). Bias and ethical considerations in machine learning and the automation of perioperative risk assessment. *British Journal of Anaesthesia*, 125(6), 843-846.
- O'Sullivan, S., Nevejans, N., Allen, C., Blyth, A., Leonard, S., Pagallo, U., ... & Ashrafian, H. (2019). Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 15(1), e1968.
- Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. *The New England Journal of Medicine*, 375(13), 1216.
- 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.
- Panch, T., Mattie, H., & Celi, L. A. (2019). The “inconvenient truth” about AI in healthcare. *NPJ Digital Medicine*, 2(1), 1-3.
- Panchmatia, J. R., Visenio, M. R., & Panch, T. (2018). The role of artificial intelligence in orthopaedic surgery. *British Journal of Hospital Medicine*, 79(12), 676–681.
- Parikh, R. B., Obermeyer, Z., & Navathe, A. S. (2019). Regulation of predictive analytics in medicine. *Science*, 363(6429), 810-812.

- Park, S. H., Do, K. H., Kim, S., Park, J. H., & Lim, Y. S. (2019a). What should medical students know about artificial intelligence in medicine?. *Journal of Educational Evaluation for Health Professions*, 16, 18.
- Park, S. H., Kim, Y. H., Lee, J. Y., Yoo, S., & Kim, C. J. (2019b). Ethical challenges regarding artificial intelligence in medicine from the perspective of scientific editing and peer review. *Science Editing*, 6(2), 91-98.
- Paulus, J. K., & Kent, D. M. (2020). Predictably unequal: understanding and addressing concerns that algorithmic clinical prediction may increase health disparities. *NPJ Digital Medicine*, 3(1), 1-8.
- Pesapane, F., Volonté, C., Codari, M., & Sardanelli, F. (2018). Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. *Insights into Imaging*, 9(5), 745-753.
- Peters, D., Vold, K., Robinson, D., & Calvo, R. A. (2020). Responsible AI—two frameworks for ethical design practice. *IEEE Transactions on Technology and Society*, 1(1), 34-47.
- Pfohl, S. R., Foryciarz, A., & Shah, N. H. (2020). An empirical characterization of fair machine learning for clinical risk prediction. *Journal of Biomedical Informatics*, <https://doi.org/10.1016/j.jbi.2020.103621>.
- Pierce, R. (2018). Machine learning for diagnosis and treatment: Gymnastics for the GDPR. *European Data Protection Law Review*, 4, 333.
- Pirracchio, R., Cohen, M. J., Malenica, I., Cohen, J., Chambaz, A., Cannesson, M., ... & ACTERREA Research Group. (2019). Big data and targeted machine learning in action to assist medical decision in the ICU. *Anaesthesia Critical Care & Pain Medicine*, 38(4), 377-384.
- Ploug, T., & Holm, S. (2016). Meta consent—a flexible solution to the problem of secondary use of health data. *Bioethics*, 30(9), 721-732.
- Portacolone, E., Halpern, J., Luxenberg, J., Harrison, K. L., & Covinsky, K. E. (2020). Ethical issues raised by the introduction of artificial companions to older adults with cognitive impairment: A call for interdisciplinary collaborations. *Journal of Alzheimer's Disease*, 76(2), 445-455.
- Poulsen, A., Fosch-Villaronga, E., & Burmeister, O. K. (2020). Cybersecurity, value sensing robots for LGBTIQ+ elderly, and the need for revised codes of conduct. *Australasian Journal of Information Systems*, 24.
- Powell, J. (2019). Trust me, i'm a chatbot: how artificial intelligence in health care fails the Turing Test. *Journal of Medical Internet Research*, 21(10), e16222.
- Powles, J., & Hodson, H. (2017). Google DeepMind and healthcare in an age of algorithms. *Health and Technology*, 7(4), 351-367.
- Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature Medicine*, 25(1), 37-43.
- Price, W. N., Kaminski, M. E., Minssen, T., & Spector-Bagdady, K. (2019). Shadow health records meet new data privacy laws. *Science*, 363(6426), 448-450.
- Prosperi, M., Guo, Y., Sperrin, M., Koopman, J. S., Min, J. S., He, X., ... & Bian, J. (2020). Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nature Machine Intelligence*, 2(7), 369-375.
- Prosperi, M., Min, J. S., Bian, J., & Modave, F. (2018). Big data hurdles in precision medicine and precision public health. *BMC Medical Informatics and Decision Making*, 18(1), 1-15.
- Racine, E., Boehlen, W., & Sample, M. (2019). Healthcare uses of artificial intelligence: Challenges and opportunities for growth. *Healthcare Management Forum*, 32(5), 272-275.
- Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380(14), 1347-1358.
- Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine*, 169(12), 866-872.
- Reddy, S., Allan, S., Coghlan, S., & Cooper, P. (2020). A governance model for the application of AI in health care. *Journal of the American Medical Informatics Association*, 27(3), 491-497.
- Reiz, A. N., de la Hoz, M. A., & García, M. S. (2019). Big data analysis and machine learning in intensive care units. *Medicina Intensiva (English Edition)*, 43(7), 416-426.
- Rickert, J. (2020). On Patient Safety: The Lure of Artificial Intelligence—Are We Jeopardizing Our Patients' Privacy?. *Clinical Orthopaedics and Related Research*, 478(4), 712-714.

- Rigby, M. J. (2019). Ethical dimensions of using artificial intelligence in health care. *AMA Journal of Ethics*, 21(2), 121-124.
- Rubeis, G. (2020). The disruptive power of artificial intelligence. Ethical aspects of gerontechnology in elderly care. *Archives of Gerontology and Geriatrics*, 91, 104186.
- Rush, B., Celi, L. A., & Stone, D. J. (2019). Applying machine learning to continuously monitored physiological data. *Journal of Clinical Monitoring and Computing*, 33(5), 887-893.
- Safdar, N. M., Banja, J. D., & Meltzer, C. C. (2020). Ethical considerations in artificial intelligence. *European Journal of Radiology*, 122, 108768.
- Salerno, J., Knoppers, B. M., Lee, L. M., Hlaing, W. M., & Goodman, K. W. (2017). Ethics, big data and computing in epidemiology and public health. *Annals of Epidemiology*, 27(5), 297-301.
- Sallstrom, L., Morris, O., & Mehta, H. (2019). Artificial Intelligence in Africa's Healthcare: Ethical Considerations. *Observer Research Foundation Issue Brief*, (312).
- Samuel, G., & Derrick, G. (2020). Defining ethical standards for the application of digital tools to population health research. *Bulletin of the World Health Organization*, 98(4), 239-244.
- Sandhu, S., Lin, A. L., Brajer, N., Sperling, J., Ratliff, W., Bedoya, A. D., ... & Sendak, M. P. (2020). Integrating a Machine Learning System Into Clinical Workflows: Qualitative Study. *Journal of Medical Internet Research*, 22(11), e22421.
- Schneider, D. F. (2019). Making machine learning accessible and actionable for clinicians. *JAMA Network Open*, 2(12), e1917362-e1917362.
- Schönberger, D. (2019). Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications. *International Journal of Law and Information Technology*, 27(2), 171-203.
- Schwalbe, N., & Wahl, B. (2020). Artificial intelligence and the future of global health. *The Lancet*, 395(10236), 1579-1586.
- Sendak, M. P., D'Arcy, J., Kashyap, S., Gao, M., Nichols, M., Corey, K., ... & Balu, S. (2020). A path for translation of machine learning products into healthcare delivery. *EMJ Innovation*, 10, 19-172.
- Senders, J. T., Arnaout, O., Karhade, A. V., Dasenbrock, H. H., Gormley, W. B., Broekman, M. L., & Smith, T. R. (2018). Natural and artificial intelligence in neurosurgery: a systematic review. *Neurosurgery*, 83(2), 181-192.
- Séroussi, B., Hollis, K. F., & Soualmia, L. F. (2020). Transparency of Health Informatics Processes as the Condition of Healthcare Professionals' and Patients' Trust and Adoption: the Rise of Ethical Requirements. *Yearbook of Medical Informatics*, 29(01), 7-10.
- Shachar, C., Gerke, S., & Adashi, E. Y. (2020). AI Surveillance during Pandemics: Ethical Implementation Imperatives. *Hastings Center Report*, 50(3), 18-21.
- Shah, N. H., Milstein, A., & Bagley, S. C. (2019). Making machine learning models clinically useful. *JAMA*, 322(14), 1351-1352.
- Shameer, K., Johnson, K. W., Glicksberg, B. S., Dudley, J. T., & Sengupta, P. P. (2018). Machine learning in cardiovascular medicine: are we there yet?. *Heart*, 104(14), 1156-1164.
- Shaw, J., Rudzicz, F., Jamieson, T., & Goldfarb, A. (2019). Artificial intelligence and the implementation challenge. *Journal of Medical Internet Research*, 21(7), e13659.
- Shilo, S., Rossman, H., & Segal, E. (2020). Axes of a revolution: challenges and promises of big data in healthcare. *Nature Medicine*, 26(1), 29-38.
- Six Dijkstra, M. W., Siebrand, E., Dorrestijn, S., Salomons, E. L., Reneman, M. F., Oosterveld, F. G., ... & Bieleman, H. J. (2020). Ethical considerations of using machine learning for decision support in occupational health: An example involving periodic workers' health assessments. *Journal of Occupational Rehabilitation*, 30, 343-353.
- Starke, G., De Clercq, E., Borgwardt, S., & Elger, B. S. (2020). Computing schizophrenia: ethical challenges for machine learning in psychiatry. *Psychological Medicine*, 1-7.
- Subramaniam, K. (2020). Public private partnerships delivering smart health to combat the tsunami of noncommunicable diseases. *Studies in Health Technology and Informatics*, 272, 374-378.
- Sung, J. J., & Poon, N. C. (2020). Artificial intelligence in gastroenterology: where are we heading?. *Frontiers of Medicine*, 14, 511-517.
- Tekin, Ş. (2020). Is big data the new stethoscope? Perils of digital phenotyping to address mental illness. *Philosophy & Technology*, 1-15.

- Thesmar, D., Sraer, D., Pinheiro, L., Dadson, N., Veliche, R., & Greenberg, P. (2019). Combining the power of artificial intelligence with the richness of healthcare claims data: opportunities and challenges. *Pharmacoeconomics*, 37(6), 745-752.
- Toh, T. S., Dondelinger, F., & Wang, D. (2019). Looking beyond the hype: Applied AI and machine learning in translational medicine. *EBioMedicine*, 47, 607-615.
- Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44-56.
- Tran, B. X., Nghiem, S., Sahin, O., Vu, T. M., Ha, G. H., Vu, G. T., Pham, H. Q., Do, H. T., Latkin, C. A., Tam, W., Ho, C. S. H., & Ho, R. C. M. (2019). Modeling research topics for artificial intelligence applications in medicine: Latent Dirichlet Allocation Application Study. *Journal of Medical Internet Research*, 21(11), e15511.
- Van Biesen, W., Decruyenaere, J., Sideri, K., Cockbain, J., & Sterckx, S. (2019). Remote digital monitoring of medication intake: methodological, medical, ethical and legal reflections. *Acta Clinica Belgica*, 1-8.
- Vayena, E., Blasimme, A., & Cohen, I. G. (2018a). Machine learning in medicine: addressing ethical challenges. *PLoS Medicine*, 15(11), e1002689.
- Vayena, E., Dzenowagis, J., Brownstein, J. S., & Sheikh, A. (2018b). Policy implications of big data in the health sector. *Bulletin of the World Health Organization*, 96(1), 66-68.
- Vayena, E., Salathé, M., Madoff, L. C., & Brownstein, J. S. (2015). Ethical challenges of big data in public health. *PLoS Computational Biology*, 11(2), e1003904.
- Vellido, A. (2020). The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural Computing and Applications*, 32(24), 18069-18083.
- Vollmer Dahlke, D., & Ory, M. G. (2020). Emerging issues of intelligent assistive technology use among people with dementia and their caregivers: a US perspective. *Frontiers in Public Health*, 8(191).
- Vollmer, S., Mateen, B. A., Bohner, G., Király, F. J., Ghani, R., Jonsson, P., ... & Hemingway, H. (2020). Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *BMJ*, 368.
- Wahl, B., Cossy-Gantner, A., Germann, S., & Schwalbe, N. R. (2018). Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? *BMJ Global Health*, 3(4), e000798.
- Waithira, N., Mutinda, B., & Cheah, P. Y. (2019). Data management and sharing policy: the first step towards promoting data sharing. *BMC Medicine*, 17(1), 1-5.
- Wangmo, T., Lipps, M., Kressig, R. W., & Ienca, M. (2019). Ethical concerns with the use of intelligent assistive technology: findings from a qualitative study with professional stakeholders. *BMC Medical Ethics*, 20(1), 98.
- Wartman, S. A., & Combs, C. D. (2019). Reimagining medical education in the age of AI. *AMA Journal of Ethics*, 21(2), 146-152.
- Watson, D. S., Krutzinna, J., Bruce, I. N., Griffiths, C. E., McInnes, I. B., Barnes, M. R., & Floridi, L. (2019). Clinical applications of machine learning algorithms: beyond the black box. *BMJ*, 364, 1886.
- Weintraub, W. S., Fahed, A. C., & Rumsfeld, J. S. (2018). Translational medicine in the era of big data and machine learning. *Circulation Research*, 123(11), 1202-1204.
- Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., ... & Goldenberg, A. (2019). Do no harm: a roadmap for responsible machine learning for health care. *Nature Medicine*, 25(9), 1337-1340.
- Wilkinson, J., Arnold, K. F., Murray, E. J., van Smeden, M., Carr, K., Sippy, R., ... & Tennant, P. W. (2020). Time to reality check the promises of machine learning-powered precision medicine. *The Lancet Digital Health*, 2(12), e677-e680.
- Willeminck, M. J., Koszek, W. A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., ... & Lungren, M. P. (2020). Preparing medical imaging data for machine learning. *Radiology*, 295(1), 4-15.
- Xafis, V., Schaefer, G. O., Labude, M. K., Brassington, I., Ballantyne, A., Lim, H. Y., Lipworth, W., Lysaght, T., Stewart, C., Sun, S., Laurie, G. T., & Tai, E. S. (2019). An ethics framework for big data in health and research. *Asian Bioethics Review*, 11(3), 227-254.

- Xie, Q., Faust, K., Van Ommeren, R., Sheikh, A., Djuric, U., & Diamandis, P. (2019). Deep learning for image analysis: Personalizing medicine closer to the point of care. *Critical Reviews in Clinical Laboratory Sciences*, 56(1), 61-73.
- Yew, G. C. K. (2020). Trust in and ethical design of carebots: the case for ethics of care. *International Journal of Social Robotics*, 1–17.
- Yu, K. H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, 2(10), 719-731.
- Zhang, X., Pérez-Stable, E. J., Bourne, P. E., Pehrah, E., Duru, O. K., Breen, N., ... & Denny, J. (2017). Big data science: opportunities and challenges to address minority health and health disparities in the 21st century. *Ethnicity & Disease*, 27(2), 95.
- Zitnik, M., Nguyen, F., Wang, B., Leskovec, J., Goldenberg, A., & Hoffman, M. M. (2019). Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. *Information Fusion*, 50, 71-91.
- Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., & Telenti, A. (2019). A primer on deep learning in genomics. *Nature Genetics*, 51(1), 12-18.

Appendix E: Procedures for thematic analysis and ontological organisation (Jones et al., 2011)

The following procedures were adapted from Braun and Clarke's (2006) and Clarke and Braun's (2013) guidance for the use of thematic analysis in qualitative research in psychology and framed within Tranfield et al.'s (2003) guidance for systematic reviews in management.

A. Data organization	<ol style="list-style-type: none"> 1. Arrange papers in chronological order from 2000 to 2020. 2. Prepare NVivo for recording and comparing coding by researchers.
B. Theme identification and coding	<ol style="list-style-type: none"> 1. Two researchers independently read each paper for meaning, taking into account the author(s) focus and stated purpose, research questions, key arguments, methodology, and main constructs to determine the phenomena(on) with which it is concerned. 2. A descriptive statement outlining its primary focus is assigned to each paper in turn taking care to echo its conceptual terminology and vocabulary. 3. After reading and describing the first 20 papers, the researchers compare descriptive statements and resolve discrepancies through discussion. 4. Initial thematic names are generated for each paper, with names derived from their descriptive statements. 5. New thematic names are generated as required and applied to similar papers published later, thus preserving the chronology of the dataset. 6. After every 20 papers, the two researchers compare results for consistency.
C. Ontological organization	<ol style="list-style-type: none"> 1. The thematic names for each paper are ranked and agreed, resulting in the first-ranked theme becoming the first order thematic name for each paper. 2. Following an ontological process, thematic names are grouped by similarity to form second-order and major thematic areas, thus forming a taxonomic hierarchy. 3. Themes are reviewed for redundancy or duplication.
D. Thematic and ontological interpretation and validation	<ol style="list-style-type: none"> 1. Descriptors and themes are summarized in tables and sorted in chronological order by thematic area, followed by second order and first-order themes and corresponding descriptive statements as a domain ontology. 2. The thematic structure of the field is mapped (Figure 4 in main text) and compared for consistency. 3. An interpretive account of each theme is written by returning to the papers, pattern-matching against theme descriptors and ontological fit.
E. Quality checking against established practice	<ol style="list-style-type: none"> 1. Each data item (paper) was given equal attention and coded independently by two researchers. 2. The process was thorough, inclusive, and comprehensive (three thematic descriptors). 3. Themes were checked against each other and back to the original dataset. 4. Themes were checked for internal coherence, consistency, and distinctiveness. 5. Data were analyzed (interpreted) for meaning and common vocabulary preserved. 6. Themes were iteratively pattern-matched with the data, and the ontology tables and thematic map checked for consistency. 7. The active and reflective role of the researchers is fully acknowledged.

References

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Clarke, V., & Braun, V. (2013). Teaching thematic analysis: overcoming challenges and developing strategies for effective learning. *The Psychologist*, 26(2), 120–123.
- Jones, M. V., Coviello, N., & Tang, Y. K. (2011). International entrepreneurship research (1989–2009): a domain ontology and thematic analysis. *Journal of Business Venturing*, 26(6), 632–659.
- Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14(3), 207–222.

Appendix F: Authors by thematic area

These thematic areas represent our interpretation and grouping of the primary first-order theme of each article reviewed.

Thematic area	Studies
Theme A: Sustainable AI	
<i>Responsible local leadership</i>	Aboueid et al. (2019); Alami et al. (2017); Alami et al. (2020b); Blobel et al. (2020); Burnside et al. (2019); Cresswell et al. (2018); Hosny & Aerts (2019); Keskinbora & Keskinbora (2018); Luxton (2014); Martinez-Martin & Kreitmair (2018); Matheny et al. (2020); Nebeker et al. (2019); Price & Cohen (2019); Wahl et al. (2018)
<i>Social sustainability</i>	Alami et al. (2020b); Ahmed et al. (2020); Bauer et al. (2017); Carter et al. (2020); Combs & Combs (2019); Garden et al. (2016); Hadley et al. (2020); He et al. (2019); Ho et al. (2020a); D'Antonoli (2020); Kolachalama & Garg (2018); Schwalbe & Wahl (2020); Taddeo & Floridi (2018); McCall (2020); McCoy et al. (2020); Malkin & von Oldenburg Beer (2013); Mittelstadt (2017b); Moore et al. (2019); Park et al. (2019a); Poulsen et al. (2020); Rigby (2019); Wartman & Combs (2019)
Theme B: Human-centered AI	
<i>Embedding “humanness” in AI agents to meet ethics of care requirements</i>	Ashrafian et al. (2015); Banks (2018) Bickmore et al. (2010); Blease et al. (2019); Coeckelbergh et al. (2016); Cornet (2013); Cuzzolin et al. (2020); Dalton-Brown (2020); Eduard & Jordi (2019); Guan (2019); Ho (2019); Hoorn & Winter (2018); Kandalaf et al. (2013); Lupton (2018); Luxton (2014); Ngiam & Khor (2019); Nwosu et al. (2019); Schwalbe & Wahl (2020); Topol (2019); Wangmo et al. (2019); Yew (2020)
<i>The role of health professionals in maintaining public trust</i>	Abràmoff et al. (2020); Brady & Neri (2020); Buruk et al. (2020); Cohen et al. (2014); Darcy et al. (2016); Deo (2015); Di Ieva (2019); Diprose & Buist (2016); Fiske et al. (2019a); Forsting (2017); Garbuio & Lin (2019); Gruson et al. (2019); Holzinger (2014); Horgan et al. (2019); Johnson (2019); Johnson (2020); Li et al. (2019); Mazurowski (2020); Meskó et al. (2018); Miller (2020); Mullainathan & Obermeyer (2019); Neri et al. (2020); Obermeyer & Emanuel (2016); Powell (2019); Poulsen et al. (2020); Price & Cohen (2019); Sandhu et al. (2020); Senders et al. (2018); Sung & Poon (2020); Wangmo et al. (2019); Yu et al. (2018); Zou et al. (2019)
<i>Developing artificial wisdom through interdisciplinary collaboration</i>	Jeste et al. (2020); Littmann et al. (2020); Luxton (2014); Powell (2019); Wangmo et al. (2019); Weintraub et al. (2018)
Theme C: Inclusive AI	

<i>Inclusive communication and involvement in AI governance</i>	Aitken et al. (2019); Char et al. (2020); Chen et al. (2020); Cohen et al. (2014); Di Nucci (2019); Ho et al. (2020a); Ho et al. (2020b); Hofer et al. (2020); Horgan et al. (2019); Joerin et al. (2020); Loh (2018); Luxton (2014); McCradden et al. (2020a); Mittelstadt (2017a); Mittelman et al. (2018); Noorbakhsh-Sabet et al. (2019); Poulsen et al. (2020); Rubeis (2020); Vayena et al. (2015)
Theme D: Fair AI	
<i>Alleviating algorithmic and data bias</i>	Aboueid et al. (2019); Abràmoff et al. (2020); Alami et al. (2020a); Alami et al. (2020b); Angehrn et al. (2020); Ashrafian et al. (2015); Basu et al. (2020); Beil et al. (2019); Benke & Benke (2018); Blobel et al. (2020); Bonifazi et al. (2020); Brady & Neri (2020); Burnside et al. (2019); Bukowski et al. (2020); Carter et al. (2020); Challen et al. (2019); Char et al. (2018); Chen & Asch (2017); Chen et al. (2019); Corsico (2019); Cresswell et al. (2018); Currie et al. (2020); D'Antonoli (2020); de Miguel et al. (2020); Eaneff et al. (2020); Faraj et al. (2018); Fletcher et al. (2020); Gianfrancesco et al. (2018); Geis et al. (2019); Grote & Berens (2020); Hallowell et al. (2020); Handelman et al. (2019); Henriksen & Bechmann (2020); Horgan et al. (2019); Hummel & Braun (2020); Ienca & Ignatiadis (2020); Jacobson et al. (2020); Joerin et al. (2020); Johnson (2019); Kaissis et al. (2020); Kelly et al. (2019); Keskinbora & Keskinbora (2018); Khatoon et al. (2019); Kim et al. (2019); Kohli & Geis (2018); Loh (2018); Luxton (2014); Mahomed (2018); Martinez-Martin (2019); Martinez-Martin et al. (2018); Martinez-Martin & Kreitmair (2018); Martinez-Martin et al. (2020); Matheny et al. (2020); Mathur et al. (2020); Mazurowski (2020); McCradden et al. (2020b); McCradden et al. (2020c); Miller (2020); Minssen et al. (2020); Morley et al. (2020); Nagendran et al. (2020); Neri et al. (2020); Noseworthy et al. (2020); Obermeyer et al. (2019); O'Reilly-Shah et al. (2020); Powell (2019); Price & Cohen (2019); Prospero et al. (2020); Racine et al. (2019); Rajkomar et al. (2018); Rajkomar et al. (2019); Reddy et al. (2020); Rush et al. (2019); Schönberger (2019); Sendak et al. (2020); Shameer et al. (2018); Shaw et al. (2019); Six Dijkstra et al. (2020); Starke et al. (2020); Thesmar et al. (2019); Toh et al. (2019); Tran et al. (2019); Pirracchio et al. (2019); Vayena et al. (2015); Vayena et al. (2018a); Vollmer et al. (2020); Wahl et al. (2018); Wiens et al. (2019); Willeminck et al. (2020); Zou et al. (2019)
<i>Data representation and equality</i>	Alami et al. (2020b); Amarasingham et al. (2014); Char et al. (2018); Cohen et al. (2014); Faraj et al. (2018); Garattini et al. (2019); Ienca & Ignatiadis (2020); Johnson (2019); Krutzinna et al. (2019); Mittelstadt, & Floridi (2016); Nagaraj et al. (2020) ; Powell (2019); Price & Cohen (2019); Rajkomar et al. (2019); Reddy et al. (2020); Shaw et al. (2019); Shilo et al. (2020); Subramaniam (2020); Watson et al. (2019); Wilkinson et al. (2020); Vayena et al. (2018a); Vollmer Dahlke & Ory (2020)
<i>Health disparity in low resource settings</i>	Alami et al. (2017); Alami et al. (2019); Alami et al., (2020b); Bærøe et al. (2020); Blobel et al. (2020); Christie (2018); Horgan et al. (2019); Hosny & Aerts (2019); Ibrahim et al. (2020); Malkin & von Oldenburg Beer (2013); Martinez-Martin & Kreitmair (2018); Mehta et al. (2020); Miller & Polson (2019); Mollura et al. (2020); Obermeyer et al. (2019); Panchmatia et al. (2018); Paulus & Kent (2020); Pfohl et al. (2020); Sallstrom et al. (2019); Schwalbe & Wahl (2020); Wahl et al. (2018); Zhang et al. (2017)
Theme E: Transparent AI	

<p><i>Safeguarding personal privacy</i></p>	<p>Amarasingham et al. (2014); Azencott (2018); Basu et al. (2020); Beil et al. (2019); Benke & Benke (2018); Briganti & Le Moine (2020); Bukowski et al. (2020); Cohen et al. (2014); Currie et al. (2020); Drew (2019); Fiske et al. (2019b); Floridi et al. (2020); Garattini et al. (2019); Gray & Thorpe (2015); Gruson et al. (2019); Ienca & Ignatiadis (2020); Jacobson et al. (2020); Joerin et al. (2020); Kaissis et al. (2020); Kaplan (2016); Kayaalp (2018); Keskinbora & Keskinbora (2018); Khatoun et al. (2019); Krutzinna et al. (2019); Liew (2018); Martinez-Martin (2019); Martinez-Martin et al. (2020); Mathur et al. (2020); McCradden et al. (2020b); McGraw & Petersen (2020); Miller & Polson (2019); Minssen et al. (2020); Mittelstadt (2017a); Mittelstadt (2017b); Mittelstadt, & Floridi (2016); Pierce (2018); Powles & Hodson (2017); Price et al. (2019); Price & Cohen (2019); Racine et al. (2019); Reiz et al. (2019); Rickert (2020); Safdar et al. (2020); Salerno et al. (2017); Sendak et al. (2020); Séroussi et al. (2020); Shachar et al. (2020); Six Dijkstra et al. (2020); Tekin (2020); Thesmar et al. (2019); Van Biesen et al. (2019); Vayena et al. (2018a); Xafis et al. (2019); Yu et al. (2018)</p>
<p><i>Explainability of AI-driven models and decisions</i></p>	<p>Ahmed et al. (2020); Amann et al. (2020); Baig et al. (2020); Basu et al. (2020); Balthazar et al. (2018); Beil et al. (2019); Bjerring & Busch (2020); Brady & Neri (2020); Buruk et al. (2020); Cabitza et al. (2017); Coiera (2019); D'Antonoli (2020); Davenport & Kalakota (2019); de Miguel et al. (2020); Dzobo et al. (2020); Ferretti et al. (2018); Flaxman & Vos (2018); Gauld et al. (2020); Gruson (2020); He et al. (2019); Holzinger et al. (2019); Jacobson et al. (2020); Jiménez-Luna et al. (2020); Joerin et al. (2020); Johnson (2020); Licitra et al. (2017); Liew (2018); Kelly et al. (2019); Kononenko (2001); Lazer et al. (2014); Lisboa (2002); Ma et al. (2019); McCradden et al. (2020b); Miotto et al. (2018); Mittelstadt (2017a); Morley et al. (2020); Norgeot et al. (2020); O'Sullivan et al. (2019); Parikh et al. (2019); Park et al. (2019b); Peters et al. (2020); Rush et al. (2019); Samuel & Derrick (2020); Schneider (2019); Schwalbe & Wahl (2020); Shah et al. (2019); Shaw et al. (2019); Toh et al. (2019); Prospero et al. (2018); Vayena et al. (2018b); Vellido (2020); Watson et al. (2019); Wiens et al. (2019); Xie et al. (2019); Zitnik et al. (2019); Zou et al. (2019)</p>
<p><i>Addressing the loss of confidentiality by legislation</i></p>	<p>Ahmed et al. (2020); Balthazar et al. (2018); Briganti & Le Moine (2020); Ferretti et al. (2018); Ienca & Ignatiadis (2020); Jacobson et al. (2020); Johnson (2019); Kaplan (2016); Lupton (2018); McCradden et al. (2020b); Mittelstadt (2017a); Mittelstadt (2017b); Pesapane et al. (2018); Pierce (2018); Price & Cohen, (2019); Poulsen et al. (2020)</p>
<p><i>User empowerment</i></p>	<p>Ginsberg et al. (2009); Liaw et al. (2020); Lupton (2018); Manrique de Lara & Peláez-Ballestas (2020); Peters et al. (2020); Vollmer et al. (2020)</p>
<p><i>Informed consent for data use</i></p>	<p>Aboueid et al. (2019); Amarasingham et al. (2014); Astromské et al. (2020); Balthazar et al. (2018); Baric-Parker & Anderson (2020); Becker (2019); Brady & Neri (2020); Bukowski et al. (2020); Carter et al. (2020); Cohen et al. (2014); Corsico (2019); Cornet (2013); D'Antonoli (2020); de Miguel et al. (2020); Ferretti et al. (2018); Floridi et al. (2019); Garattini et al. (2019); Gray & Thorpe (2015); Grote & Berens (2020); Gruson et al. (2019); He et al. (2019); Hummel & Braun (2020); Jacobson et al. (2020); Johnson (2019); Johnson (2020); Jung & Pfister (2020); Kaissis et al. (2020); Kaplan (2016); Krutzinna et al. (2019); Larson et al. (2020); Liaw et al. (2020); Martinez-Martin (2019); Martinez-Martin et al. (2018); Mathur et al. (2020); McCradden et al. (2020b); McGraw & Petersen (2020); Mittelstadt (2017a); Mittelstadt,</p>

	& Floridi (2016); Mudgal & Das (2020); Nagaraj et al. (2020); Neri et al. (2020); O'Sullivan et al. (2019); Panch et al. (2019); Pesapane et al. (2018); Pierce (2018); Ploug & Holm (2016); Portacolone et al. (2020); Powles & Hodson (2017); Reddy et al. (2020); Rigby (2019); Safdar et al. (2020); Salerno et al. (2017); Schönberger (2019); Six Dijkstra et al. (2020); Waithira et al. (2019); Wangmo et al. (2019); Willemink et al. (2020); Xie et al. (2019)
--	--