BMC Medical Ethics

RESEARCH ARTICLE

Open Access

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

Kathleen Murphy[1], Erica Di Ruggiero[2], Ross Upshur[3,4], Donald J. Willison[5], Neha Malhotra[1], Jia Ce Cai[1], Nakul Malhotra[1], Vincci Lui[6] and Jennifer Gibson[1*]

Abstract

Background: Artificial intelligence (AI) has been described as the "fourth industrial revolution" with transformative and global implications, including in healthcare, public health, and global health. AI approaches hold promise for improving health systems worldwide, as well as individual and population health outcomes. While AI may have potential for advancing health equity within and between countries, we must consider the ethical implications of its deployment in order to mitigate its potential harms, particularly for the most vulnerable. This scoping review addresses the following question: What ethical issues have been identified in relation to AI in the field of health, including from a global health perspective?

Methods: Eight electronic databases were searched for peer reviewed and grey literature published before April 2018 using the concepts of health, ethics, and AI, and their related terms. Records were independently screened by two reviewers and were included if they reported on AI in relation to health and ethics and were written in the English language. Data was charted on a piloted data charting form, and a descriptive and thematic analysis was performed.

Results: Upon reviewing 12,722 articles, 103 met the predetermined inclusion criteria. The literature was primarily focused on the ethics of AI in health care, particularly on carer robots, diagnostics, and precision medicine, but was largely silent on ethics of AI in public and population health. The literature highlighted a number of common ethical concerns related to privacy, trust, accountability and responsibility, and bias. Largely missing from the literature was the ethics of AI in global health, particularly in the context of low- and middle-income countries (LMICs).

Conclusions: The ethical issues surrounding AI in the field of health are both vast and complex. While AI holds the potential to improve health and health systems, our analysis suggests that its introduction should be approached with cautious optimism. The dearth of literature on the ethics of AI within LMICs, as well as in public health, also points to a critical need for further research into the ethical implications of AI within both global and public health, to ensure that its development and implementation is ethical for everyone, everywhere.

Keywords: Artificial intelligence, Ethics, Health care, Public and population health, Global health

Introduction

Rationale

Artificial intelligence (AI) has been described as the "fourth industrial revolution" with transformative and global implications [1]. AI can be generally understood as "a field of study that combines computer science, engineering and related disciplines to build machines capable of behaviour that would be said to require intelligence were it to be observed in humans" [2]. Some such behaviours include the ability to visually perceive images, recognize speech, translate language, and learn from and adapt to new information [2]. To do so, AI as a field of study can employ a number of techniques. Machine

*Correspondence: jennifer.gibson@utoronto.ca
[1] Joint Centre for Bioethics, Dalla Lana School of Public Health, University of Toronto, 155 College Street, Suite 754, Toronto, ON M5T 1P8, Canada
Full list of author information is available at the end of the article

learning, for instance, allows algorithms to make predictions and solve problems based on large amounts of data, without being explicitly programmed [2]. Deep learning is a subset of machine learning, and goes further to use multiple layers of artificial neural networks to solve complex problems from unstructured data, much like the human brain [2–4]. Many countries have developed or are in the process of developing national AI strategies and policies to promote research, development, and adoption of AI methods and technologies [5]. Amongst them, Canada was the first country to release a $125 million Pan-Canadian Artificial Intelligence Strategy to advance new public and private sector collaborations to stimulate research in AI [6]. Investments in AI are rapidly increasing with the potential for economic gains, projected at a $15.7 trillion contribution to the global economy by 2030 [7].

Amidst the nascence of AI, ethics has been identified as a priority concern in the development and deployment of AI across sectors [8–10]. In efforts to address this concern, there has been a proliferation of initiatives, including the establishment of organizations and principles documents [11] to provide guidance to those working within the AI space. Some such initiatives include the Partnership on AI [12], OpenAI [13], the Foundation for Responsible Robotics [14], the Ethics and Governance of Artificial Intelligence Initiative [15], the Montréal Declaration for Responsible Development of Artificial Intelligence [16], and the Principles for Accountable Algorithms [17, 18]. While there is increasing support from funding bodies for research on the social and ethical implications of AI [19–22], to date there has been limited attention by the academic bioethics community on AI within the field of health, particularly within the context of a globalized world. The health sector, however, is a growing area of AI research, development and deployment, with AI holding promise for the promotion of healthy behaviours; the detection and early intervention of infectious illnesses and environmental health threats; and the prevention, diagnosis, and treatment of disease [23–25].

The World Health Organization (WHO), for example, has established the "triple billion" target whereby it aims to have 1 billion more people benefit from universal health coverage, be better protected from health emergencies, and experience better health and wellbeing, and it believes that AI can help it achieve those objectives [26]. The WHO has been advancing the discussion of AI within health through its various Collaborating Centres, the AI for Global Good Summit, the development of the WHO Guideline Recommendations on Digital Interventions for Health System Strengthening [27], and its commitment to supporting countries in realizing the benefits of AI for health. Indeed, AI has been described by former WHO Director General Dr. Margaret Chan as the new frontier for health with transformative implications [28]. Yet amidst its promise, the introduction of AI in all corners of the world is accompanied by ethical questions that need to be uncovered from a global health perspective in order to be adequately addressed.

Global health has been defined as "an area for study, research, and practice that places a priority on improving health and achieving equity in health for all people worldwide" (p.1995), placing particular emphasis on the prevention and treatment of transnational population- and individual-level health issues through interdisciplinary and international collaboration [29]. To the extent that public health concerns the health of populations, global health concerns the health of populations on a global scale that transcends national boundaries and that underpins the interdependencies and interconnectivity of all people within a broader geopolitical, economic, and environmental context [29]. While both are critically important, AI, with its potential impact on research and development, trade, warfare, food systems, education, climate change, and more [30, 31], all of which either directly or indirectly impact the health of individuals, is inherently global.

In 2015, the 17 Sustainable Development Goals (SDGs) were unanimously adopted by all United Nations' Member States. Goal 3 aims to achieve "good health and well-being" [32] and Goal 10 targets the reduction of inequalities [33]. While the SDGs are founded on the values of equity, inclusion, global solidarity, and a pledge to leave no one behind [34], the advent of AI could further exacerbate existing patterns of health inequities if the benefits of AI primarily support populations in high-income countries (HICs), or privilege the wealthiest within countries. Vinuesa and colleagues [35] assessed the role of AI in achieving all 17 SDGs (and their 169 targets), and found that while AI may serve predominantly as an enabler for achieving all targets in SDG 3, for SDG 10, it can be almost equally inhibiting as it is enabling. Considering, for instance, that many low- and middle-income countries (LMICs) still face significant challenges in digitizing their health records [36], data from which AI relies, there remains a substantial technological gap to overcome in order for LMICs to harness the potential benefits offered by AI. With increasing scale and diffusion of AI technologies in health worldwide, it is therefore imperative to identify and address the ethical issues systematically in order to realize the potential benefits of AI, and mitigate its potential harms, especially for the most vulnerable.

Objectives

With this pursuit in mind, the purpose of this scoping review was to scope the academic and grey literatures in this emerging field, to better understand the discourse around the ethics of AI in health, and identify where gaps in the literature exist. Our research question was as follows: *What ethical issues have been identified in relation to AI in the field of health, including from a global health perspective?* Results from this scoping review of the academic and grey literatures include: (a) the selection of sources of evidence, (b) a descriptive analysis of the literature reviewed, (c) common ethical issues related to AI technologies in health, (d) ethical issues identified for specific AI applications in health, and (e) gaps in the literature pertaining to health, AI, and ethics.

Methods

Our approach to scoping the literature was informed by the methods outlined by Levac, Colquhoun, and O'Brien [37], and the reporting guidelines established by Tricco, Lillie, Zarin, O'Brien, Colquhoun, Levac, et al. [38]. The core search concepts for the scoping review were AI, health, and ethics. Given the evolving nature of the AI field, both academic and grey literatures were included in the search. To enhance the rigour of our grey literature search specifically, the grey literature search was informed by search methods outlined by Godin, Stapleton, Kirkpatrick, Hanning, and Leatherdale [39].

Eligibility criteria

In keeping with a scoping review methodological approach [37], the inclusion and exclusion criteria were defined a priori and were refined as necessary throughout the iterative screening process involving the full project team at the beginning, middle, and end of the screening process to ensure consistency. Articles were selected during title and abstract screening if they met the following inclusion criteria: [1] records reported on all three core search concepts (AI, ethics, and health), and [2] records were written in the English language. The criterion for articles written in the English language was included because it is the language spoken by the majority of the research team, and thus allowed us to engage in a collaborative analysis process and enhance the rigour of our review. With regard to exclusion criteria, we excluded articles that did not include each of the concepts of AI, ethics and health, as well as those not written in the English language. Although 'big data' is a critical input to AI systems, articles that focused only on ethics and big data without explicit mention of AI methods or applications were excluded. Non-peer-reviewed academic literature was also excluded (e.g. letters, and non-peer reviewed conference proceedings), as were books and book chapters, each of which are categorized as 'irrelevant record type' in Fig. 1. Finally, invalid records (e.g. those that only included a string of code, or a date and no other information) and additional duplicates identified through the title/abstract screening process were excluded as well. No date or study design limits were applied, in order to obtain as extensive a literature base as possible. For the grey literature specifically, media articles, blog posts, and magazine entries were excluded, as we were more interested in documents that were both expert-driven, and which required a degree of methodological rigour (e.g. organization/institution reports). During full-text screening, records were excluded if any of the core search concepts were not engaged in a substantive way (e.g. if a concept was mentioned in passing or treated superficially); if there was an insufficient link made between health, ethics, and AI; if the ethics of AI was not discussed in relation to human health; if the article was not written in the English language; and if it was an irrelevant record type (e.g. a book, news article, etc.).

Information sources

Searches of the peer-reviewed literature were executed in eight electronic databases: OVID MEDLINE (1946-present,includinge-pubaheadofprintandin-processandotherunindexedcitations), OVID Embase, (1947-present), OVID PsycINFO (1806-present), EBSCO CINAHL Plus with Full Text (1937-present), ProQuest Sociological Abstracts (1952-present), ProQuest Philosopher's Index (1940-present), ProQuest Advanced Technologies & Aerospace (1962-present) and Wiley Cochrane Library. The search strategy was translated into each database using combinations of each database platform's command language, controlled vocabulary, and appropriate search fields, using MeSH terms, EMTREE terms, APA's Thesaurus of Psychological Index Terms, CINAHL headings, Sociological Thesaurus, Philosopher's Index subject headings, and Advanced Technologies & Aerospace subject headings in conjunction with keywords. Limits imposed were for English language-only articles; a filter excluding animal studies was applied to searches in MEDLINE, Embase, and PsycINFO, as we were interested in the ethics of AI as it applies to humans; and a filter for health or medicine-related studies was applied to the Advanced Technologies & Aerospace database, to reduce the high volume of solely technical studies. Final searches of the peer-reviewed literature were completed on April 23, 2018.

Grey literature was retrieved between April 25th and September 12th, 2018, from (a) searches of grey literature databases including OAIster, Google Scholar, the Canadian Electronic Library, and the Canadian Institute for

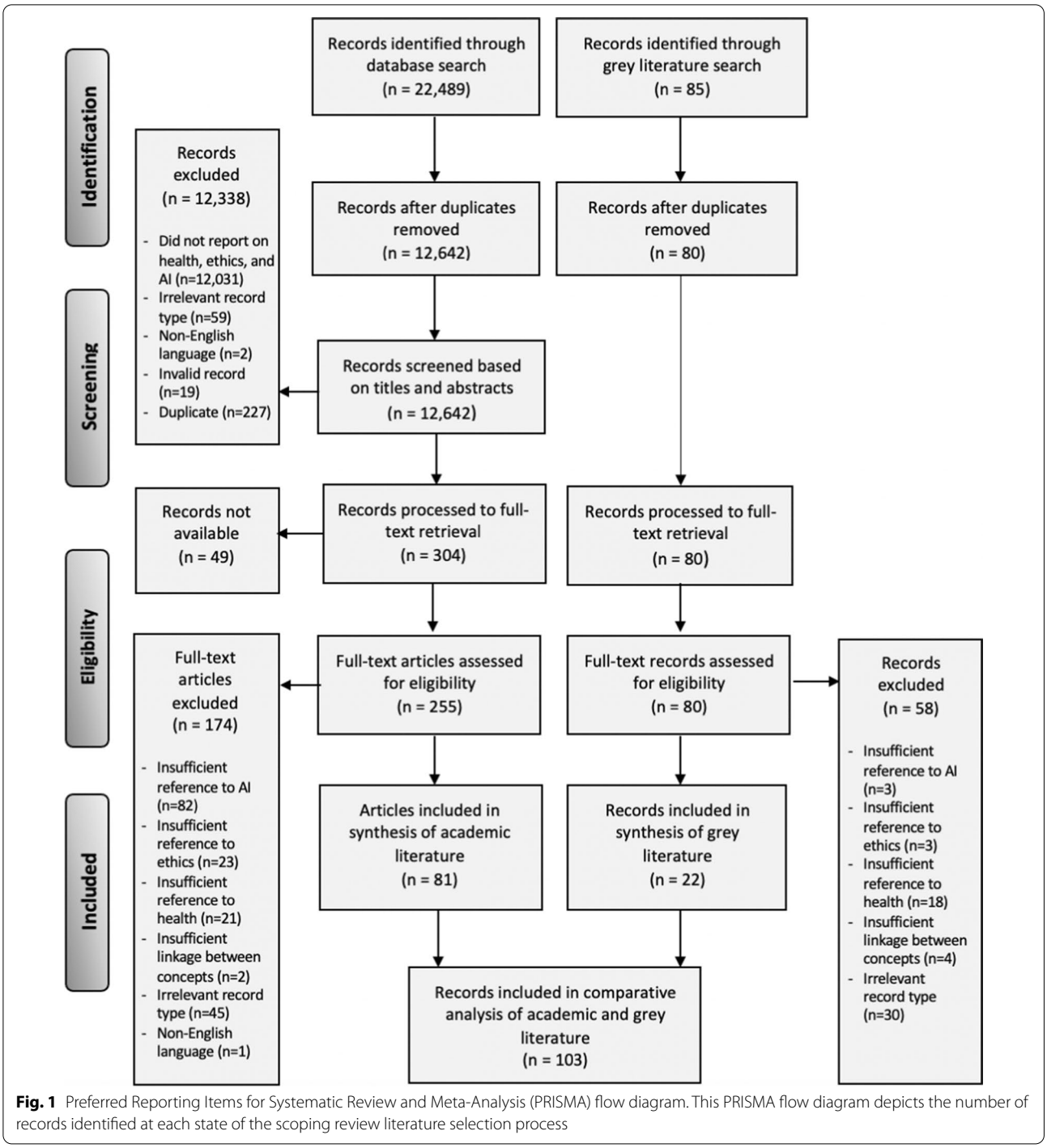

Fig. 1 Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) flow diagram. This PRISMA flow diagram depicts the number of records identified at each state of the scoping review literature selection process

Health Information; (b) a Google search and customized Google search engines which included documents from think tanks, the Canadian government, and non-governmental organizations; (c) 28 targeted website searches of known organizations and institutions; and (d) the results from a prior environmental scan conducted by a member of the project team (J.G.). The targeted website searches were undertaken to identify any grey literature that was not captured in the grey literature databases and customized Google searches. The 28 websites searched were chosen based on the existing knowledge of members of the research team, in addition to input from stakeholders who attended an AI and health symposium in June 2018.

For the purposes of feasibility and relevance, only reports from the year 2015 and beyond were retrieved.

Search

The search strategy for the academic literature was developed by an academic health science librarian (V.L.) based on recommendations from the project leads (J.G., E.DiR., R.U.), and peer-reviewed by a second librarian. The full electronic search of the peer-reviewed literature can be found in Additional file 1, with an example search from OVID MEDLINE (1946-present,includinge-pubaheadofprintandin-processandotherunindexedcitations). The search strategy and results for the grey literature is similarly outlined in Additional file 2.

Selection and sources of evidence

All identified records from the academic and grey literature searches were imported into the reference management software EndNote. After removing duplicate records, screening was conducted in two steps. First, the titles and abstracts of academic records were independently screened by two reviewers based on the inclusion and exclusion criteria established a priori. Reviewers consulted academic record keywords if the title and abstract lacked clarity in relation to the core concepts. Given that the majority of the grey literature did not include abstracts, grey literature records were screened initially on title. So as not to overlook relevant grey literature (given that some grey literature discussed ethical issues of AI more generally, including those pertaining to health), records proceeded to a full-text screening even if the title alluded to two of our three search concepts. A third reviewer assessed any records for which there was uncertainty among the reviewers about fit with the inclusion/exclusion criteria or discrepancy in reviewer assessments, and a final decision was made upon consensus with the research team. All records that passed the first level screening were pulled for full-text review by the two independent reviewers. The independent review and iterative team process were applied. The resulting sample was retained for data charting and analysis.

Data charting process

Draft data charting forms for recording extracted data from the screened articles were created using Microsoft Excel (Version 16.18.(181,014)) based on the scoping review research question. As per the recommendations of Levac et al. [37], the data charting forms were piloted by having two project team members independently chart the first 10 academic and grey literature records [20 in total], with any arising discrepancies or uncertainties being brought to the larger project team for an agreed-upon resolution. The forms were further refined based on discussions with the project team and finalized upon consensus prior to completing the data charting process. For the remaining articles, each record was charted by one member of the research team, and weekly check-in meetings with the research team were held to ensure consistency in data charting, and to verify accuracy.

Data items

We extracted data on the objective of each paper; the institutional affiliations of authors; the publication year; the country of the first and corresponding authors; whether a conflict of interest was stated; the health context of interest; the AI applications or technologies discussed; the ethical concepts, issues or implications raised; any reference to global health; and recommendations for future research, policy, or practice. Data was copy and pasted directly into the data charting form with the corresponding page number, so that no information was lost to paraphrasing. A template of the data charting form can be found in Additional file 3.

Synthesis of results

The analysis comprised two components: descriptive and thematic. The descriptive analysis captured information about global location of primary authorship, dates of publication, and the AI application(s) discussed. Primary authorship was determined by the institutional location of the first author. The academic and grey literatures were compared to identify any notable differences in scope and emphasis. The thematic analysis [40] was conducted inductively. First, open descriptive codes were generated from a random sample of 10 academic records, and 10 grey literature records from which data had been extracted in the data charting form. Upon generating consensus among project team members on the appropriate codes after several attempts at refinement, codes were applied to meaningful data points throughout the entirety of the grey and academic records in the respective data charting forms, with new codes added as necessary. These codes were reorganized into themes and then compared amongst one another to identify commonalities and gaps in the literature, including convergences and divergences between the grey and academic literatures in relation to the original research question. Results are presented below in a narrative format, with complimentary tables and figures to provide visual representation of key findings.

Results

Selection of sources of evidence

Of the 12,722 records identified after de-duplication, 81 peer-reviewed articles and 22 grey literature records

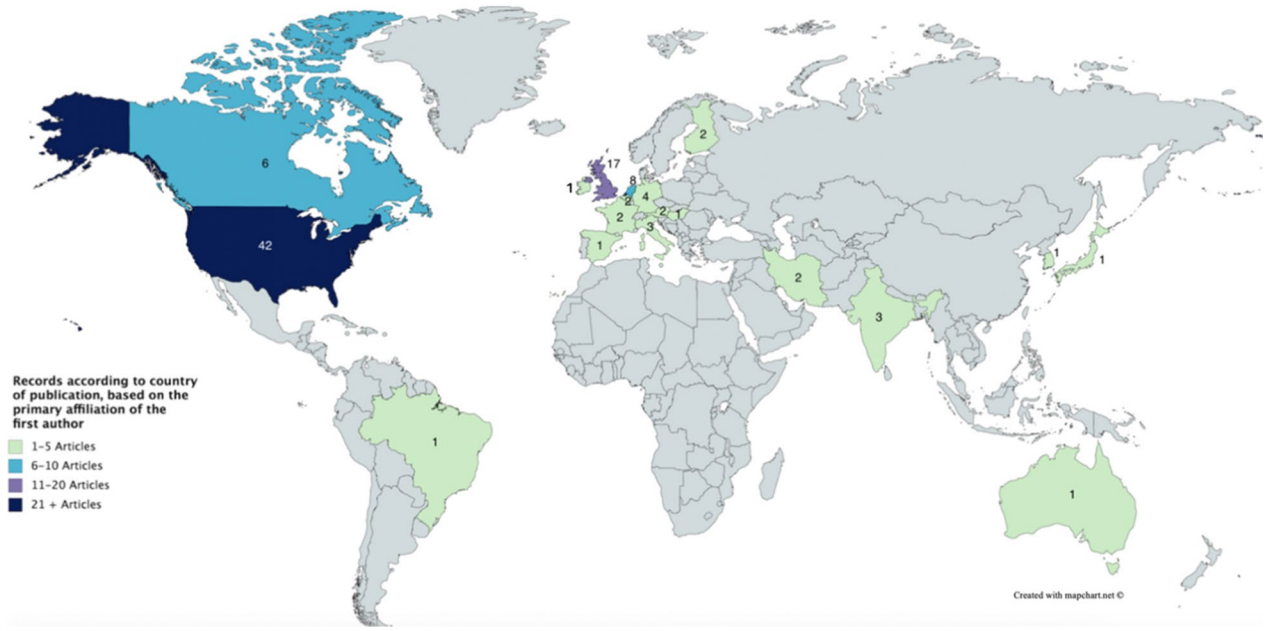

Fig. 2 Number of publications by country, based on first author affiliation. *Note that two records were published by international organizations, and the geographic origin of one record is unknown. These three records are not represented in the above figure. This map was created using mapchart.net

met the inclusion criteria for a total of 103 records in the scoping review sample (Fig. 1).

Synthesis of results

Descriptive analytics

The vast majority of publications had primary authors in the United States (n = 42) or the United Kingdom (n = 17) (Fig. 2) and while our literature search yielded publications between 1989 and 2018, most were published between 2014 and 2018 (Fig. 3). The academic and grey literatures addressed numerous AI-enabled health applications, including in particular, care robots[1] (n = 48), followed by diagnostics (n = 36), and precision medicine (n = 16) (Fig. 4).

There were notable differences between the academic and grey literature sources in terms of authorship, AI health applications addressed, and treatment of ethical implications. The academic literature was written by persons primarily affiliated with academic institutions, whereas the grey literature was written by researchers, industry leaders, and government officials, often collaboratively, with authors frequently affiliated with multiple institutions. The grey literature tended to cover a broader range of AI health applications, issues, and trends, and their associated ethical implications, whereas the academic papers typically centered their discussion on one or at most a few topics or applications. The grey literature was oriented more towards broader health and social policy issues, whereas the academic literature tended to focus on a particular dimension of AI in

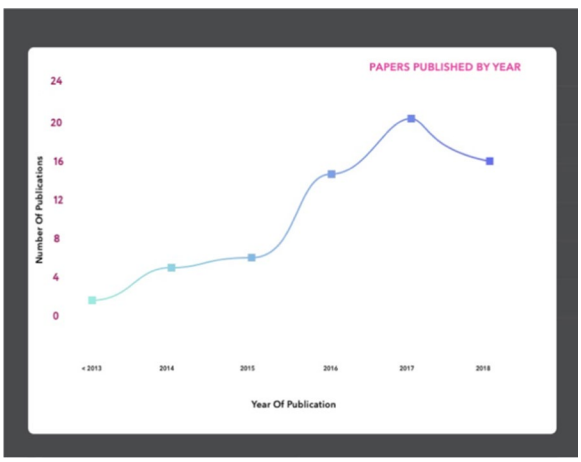

Fig. 3 Number of publications reviewed, categorized by year of publication. *The graph begins in year 2013, after which the majority of articles were published

[1] Robots for the care of the sick, elderly, or disabled bore a number of different labels in the literature, however they will herein be described as 'care robots' in an effort to broadly discuss the associated ethical challenges. 'Care robots' as used in this context are exclusive of surgical robots. Only those care robots that relied on AI are discussed, such as those that can understand commands, can locate and pick up objects, relocate a patient, and other tasks that require machine intelligence.

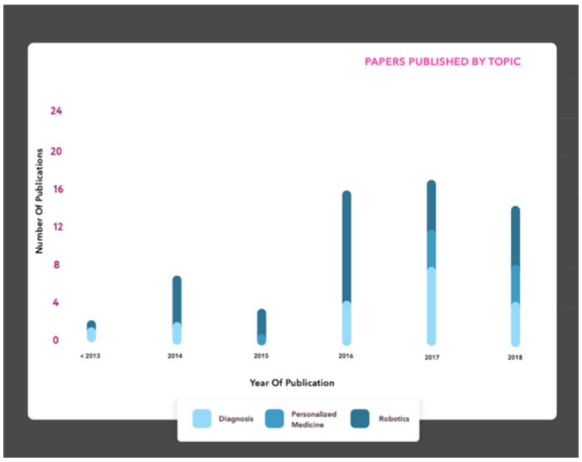

Fig. 4 Publications reviewed according to the most frequently reported AI health applications. *The graph begins in year 2013, after which the majority of articles were published

health. As compared to the grey literature, robotics, particularly care robotics[(a)] were highly represented in the peer-reviewed literature (48% of peer-reviewed literature, n = 39; 18% of the grey literature, n = 4). The academic literature on care robots was most concerned with the ethics of using care robots in health settings (e.g. "How much control, or autonomy, should an elderly person be allowed?"… "Are the safety and health gains great enough to justify the resulting restriction of the individual's liberty?" (41, p.31, p.33), whereas the grey literature tended to emphasize ethical or operational implications of using robots in health settings, such as the potential displacement of human jobs [42].

Common ethical themes

Four ethical themes were common across the health applications of AI addressed in the literature, including data privacy and security, trust in AI, accountability and responsibility, and bias. These issues, while in many ways interconnected, were identified based on how distinctly they were discussed in the literature.

Privacy and security

Issues of privacy and data security were raised about the collection and use of patient data for AI-driven applications, given that these systems must be trained with a sizeable amount of personal health information [43, 44]. Highlighted concerns about the collection and use of patient data were that they may be used in ways unbeknownst to the individual from whom the information was collected [45], and that there is a potential for information collected by and for AI systems to be hacked [45]. One illustrative example of this challenge was that of the diagnostic laboratory database in Mumbai that was hacked in 2016, during which 35,000 patient medical records were leaked, inclusive of patient HIV status, with many patients never informed of the incident [45]. Further noted was that patients may believe that their data are being used for one purpose, yet it can be difficult to predict what the subsequent use may be [46, 47]. For example, ubiquitous surveillance for use by AI systems through personal devices, smart cities, or robotics, introduces the concern that granular data can be re-identified [48, 49], and personal health information can be hacked and shared for profit [49]. Of further concern was that these smart devices are often powered by software that is proprietary, and consequently less subject to scrutiny [48]. The stated implications of these privacy and security concerns were vast, with particular attention given to if ever personal data was leaked to employers and insurance companies [46, 50–54]. A prevailing concern was how population sub-groups may then be discriminated against based on their social, economic, and health statuses by those making employment and insurance decisions [49–51, 53].

Trust in AI applications

The issues of privacy, security, and patient and healthcare professional [HCP] trust of AI were frequently and closely linked in the literature. Attention was given, for instance, to how individuals must be able to trust that their data is used safely, securely, and appropriately if AI technology is to be deployed ethically and effectively [2, 46, 55–57]. Asserted in the literature was that patients must be well enough informed of the use of their data in order to trust the technology and be able to consent or reject its use [52, 56]. One example that highlights these concerns is the data sharing partnership between Google DeepMind, an AI research company, and the Royal Free London NHS Foundation Trust (NHS) [49, 58]. Identifiable data from 1.6 million patients was shared with DeepMind with the stated intention of improving the management of acute kidney injuries with a clinical alert app [58]. However, there was a question of whether the quantity and content of the data shared was proportionate to what was necessary to test the app, and why it was necessary for DeepMind to retain the data indefinitely [49, 58]. Furthermore, this arrangement has come under question for being made in the absence of adequate patient consent, consultations with relevant regulatory bodies, or research approval, threatening patient privacy, and consequently public trust [49, 58].

HCPs have similarly demonstrated a mistrust in AI, resulting in a hesitancy to use the technology [59, 60]. This was exhibited, for instance, by physicians in various

countries halting the uptake of IBM's Watson Oncology, an AI-powered diagnostic support system [61]. These physicians stated that Watson's recommendations were too narrowly focused on American studies and physician expertise, and failed to account for international knowledge and contexts [61]. The distrust amongst HCPs was also raised with regard to machine learning programs being difficult to both understand and explain [62, 63]. In contrast, a fear exists that some HCPs may place too much faith in the outputs of machine learning processes, even if the resulting reports, such as brain mapping results from AI systems, are inconclusive [57]. One suggestion to improve HCP trust in AI technology was to deploy training and education initiatives so HCPs have a greater understanding of how AI operates [43]. A further suggestion was to promote the inclusion of end-users in the design of the technology so that not only will end-users develop a better understanding of how it functions [64], but user trust will also increase through a more transparent development process [47].

Accountability and responsibility for use of AI technology

Frequently mentioned was the question of who ought to assume responsibility for errors in the application of AI technology to clinical and at-home care delivery [41, 45, 58–60, 65–67]. The question often arose in response to the fact that AI processes are often too complex for many individuals to understand and explain, which hinders their ability to scrutinize the output of AI systems [2, 61, 66]. Similarly, grounds for seeking redress for harm experienced as a result of its use were noted to be obstructed by the proprietary nature of AI technology, for under the ownership of private companies, the technology is less publicly accessible for inspection [2, 48, 51, 68]. Further to these questions, a debate remains as to whether or not HCPs ought to be held responsible for the errors of AI in the healthcare setting, particularly with regard to errors in diagnostic and treatment decisions [41, 45, 57, 65]. Several records put forward the view that, because HCPs are legally and professionally responsible for making decisions in their patient's health interests, they bear responsibility for the consequences of decisions aided by AI technology [46, 47, 50, 59, 67, 69, 70]. However, records underlined the responsibility of manufacturers of AI systems for ensuring the quality of AI systems, including safety and effectiveness [47, 59, 71, 72], and for being responsive to the needs and characteristics of specific patient populations [72].

Beyond the clinical environment, issues of accountability arose in the context of using care robots. Related questions revolved around the burden of responsibility if an AI-enabled robotic care receiver is, for example, harmed by a robotic care provider [2, 73]. Is the burden of responsibility for such harm on the robot manufacturer who wrote the learning algorithm [73]? Similarly, the question arose of who is to be held accountable if a care receiver takes their own life or the life of another under the watch of a care robot [46]. If a care robot is considered an autonomous agent, should this incident then be the responsibility of the robot [46]? While proposed solutions to accountability challenges were few, one suggestion offered included building in a machine learning accountability mechanism into AI algorithms that could themselves perform black box audits to ensure they are privacy neutral (45, p.18). Also suggested was appropriate training of engineers and developers on issues of accountability, privacy, and ethics, and the introduction of national regulatory bodies to ensure AI systems have appropriate transparency and accountability mechanisms [45].

Where the above findings on accountability relate more to the "answerability" of AI's potentially adverse impacts, responsibility was also present in the literature with regard to AI design and governance, albeit far less so. To promote responsible AI, governments were described as holding responsibility for developing policy to address ethical, social, and legal issues, including research and development of AI technologies, and for regulatory oversight [60, 74, 75]. Records also suggested that policymakers seek to understand public perceptions of the use of AI in health [75] and to ensure that AI technologies are distributed equally [74]. One article drew attention to the risk of exacerbating health inequities as a result of the unequal distribution of AI, particularly where AI applications are increasingly being used by patients for the self-management of their health [76]. While there was little mention of corporate responsibility, a small number of articles alluded to commercial strategies for responsible innovation [54, 55, 77]. Some such strategies included identifying where bias manifests and how and by whom it is managed; and being transparent in how the algorithm has been used (e.g. using a training dataset or in a real-world setting) and what type of learning the model is built for (e.g. supervised or unsupervised learning, etc.) [55]. Other suggestions included having AI manufacturing companies monitor the use of their systems in various contexts after being deployed [77], and to have AI research and development involve 'human participation' to ensure its conscientious development (54, p.10).

Adverse consequences of bias

Bias was yet another transcending ethical theme within the literature, notably the potential bias embedded within algorithms [43, 54, 59, 64, 68, 71, 77–79], and within the data used to train algorithms [43, 45, 49, 51,

55, 59–61, 63, 64, 68, 73, 77, 77, 78, 80–84]. The prevailing concern with algorithms was that they are developed by humans, who are by nature fallible, and subverted by their own values and implicit biases [68, 79]. These values have been noted to often reflect those that are societally endemic, and if carried into the design of AI algorithms, could consequently produce outputs that advantage certain population groups over others [43, 51, 54, 59, 63, 68, 71, 77, 77, 81]. Bias was indicated to similarly manifest in the data relied upon to train AI algorithms, by way of inaccurate and incomplete datasets [48, 51, 63, 81, 84], or by unrepresentative data sets [43, 82, 83], thus rendering AI outputs ungeneralizable to the population unto which it is applied [51, 68, 81].

Not only have biased data sets been noted to potentially perpetuate systemic inequities based on race, gender identity, and other demographic characteristics [48, 51, 59, 63, 68, 78], they may limit the performance of AI as a diagnostic and treatment tool due to the lack of generalizability highlighted above [43, 48, 83]. In contrast, some noted the potential for AI to mitigate existing bias within healthcare systems. Examples of this potential include reducing human error [50]; mitigating the cognitive biases of HCPs in determining treatment decisions, such as recency, anchoring, or availability biases [45, 51]; and reducing biases that may be present within healthcare research and public health databases [48]. Suggestions to address the issue of bias included building AI systems to reflect current ethical healthcare standards [78], and ensuring a multidisciplinary and participatory approach to AI design and deployment [79].

Specific ethical themes by AI application in health

Three health applications were emphasized in the reviewed literature: care robots, diagnostics, and precision medicine. Each health application raised unique ethical issues and considerations.

Care robotics

A notable concern for the use of care robots was the social isolation of care recipients, with care robots potentially replacing the provision of human care [41, 61, 72, 85–89]. Some asserted that the introduction of care robots would reduce the amount of human contact care recipients would receive from family, friends, and human care providers [41, 61, 72, 85, 87–89]. Implications of this included increased stress, higher likelihood of dementia, and other such impacts on the well-being of care recipients [41]. Others, in contrast, viewed robots as an opportunity to increase the "social" interaction that already isolated individuals may experience [41, 85, 90, 91]. Care robots could, for example, offer opportunities for care recipients to maintain interactive skills [91], and increase the amount of time human care providers spend having meaningful interactions with those they are caring for [85] as opposed to being preoccupied with routine tasks. Yet despite these opportunities, of note was the idea that care robots risk deceiving care recipients into having them believe that the robots are 'real' care providers and companions [41, 46, 72, 85, 87, 88, 92–94], which could undermine the preservation and promotion of human dignity [41, 92].

The issue of deception often linked to the question of 'good care', what the criteria for good care are, and whether robots are capable of providing it. In the context of deceit, some considered it justified as long as the care robot allows recipients to achieve and enhance their human capabilities [93, 95]. Also challenged was the assumption that good care is contingent upon humans providing it [46, 93, 96], for while robots may not be able to provide reciprocal emotional support [93], humans similarly may fail to do so [96]. A further illustrated aspect of good care was the preservation and advancement of human dignity [93], support for which can be offered by robots insofar as they promote individual autonomy [41, 61, 73, 85, 87, 88]. Some, however, contested this, arguing that care robots may in fact reduce a person's autonomy if the technology is too difficult to use [87]; if the robot supersedes one's right to make decisions based on calculations of what it thinks is best [61]; and because the implementation of robots may lead to the infantilization of care recipients, making them feel as though they are being treated like children [88]. The promotion of autonomy also appeared controversial, acknowledged at times as the pre-eminent value for which robots ought to promote [73, 91], where at others, autonomy was in tension with the safety of the care recipient [41, 91]. For example, with the introduction of care robots, care recipients might choose to engage in unsafe behaviours in pursuit of, and as a result of, their new independence [41, 91]. A comparable tension existed in the literature between the safety of care recipients, which some believe care robots protect, and the infringement on the recipient's physical, and information privacy [41, 46, 88, 91, 97, 98].

Diagnostics

Diagnostics was an area that also garnered significant attention with regard to ethics. Of note was the 'black box' nature of machine learning processes (36, 45, 51, 63, 74, 80, 99, 100), frequently mentioned with a HCP's inability to scrutinize the output [44, 51, 63, 74]. Acknowledging that the more advanced the AI system, the more difficult it is to discern its functioning [99], there was also a concern that due to the difficulty in understanding how and why a machine learning program produces

an output, there is a risk of encountering biased outputs [80]. Thus, despite the challenge of navigating these opaque AI systems, there was a call for said systems to be explainable in order to ensure responsible AI [45, 80]. Also a pervasive theme was the replacement and augmentation of the health workforce, particularly physicians, as a result of AI's role in diagnostics [44, 59, 63, 100, 101]. While few feared the full replacement of physicians in diagnostics [2, 63, 100], some expected its presence to actually enhance the effectiveness and efficiency of their work [63, 100]. There were expressed concerns, however, about how the roles and interactions of physicians may change with its introduction, such as the ethical dilemma encountered if a machine learning algorithm is inconsistent with the HCP's recommendation, if it contradicts a patient's account of their own condition, or if it fails to consider patients' non-verbal communication and social context [59].

Precision medicine

Issues of bias persisted in discussions of precision medicine, with the recognition that biased data sets, such as those that exclude certain patient populations, can produce inaccurate predictions that in turn can have unfair consequences for patients [81]. While precision medicine was a less prominent theme than the aforementioned AI applications, questions of the accuracy of predictive health information from the intersection of AI and genomics arose, as did an uncertainty of where and by whom that data may then be used [102]. In the case of AI-assisted gene editing, deep learning holds potential for directing experts where in the human genome to use gene editing technologies such as CRISPR, to reduce an individual's risk of contracting a genetic disease or disorder [25]. However, deep learning models cannot discern the moral difference between gene editing for health optimization, and gene editing for human enhancement more generally, which may blur ethical lines [25]. A further tension existed in how the technology is deployed to support human choices; for example if a person not only seeks gene editing to reduce their risk of inheriting a particular genetic disease, but to also increase their muscle mass, obtain a particular personality trait, or enhance their musical ability [25]. Also illuminated was the implications of AI-enabled precision medicine in the global north versus the global south [103]. First is the possibility that this technology, given its high associated costs and greater accessibility in the developed world, might leave LMICs behind [103]. Second was the awareness that the introduction of genetic testing may undermine low cost, scalable and effective public health measures, which should remain central to global health [103].

Gaps in the literature

Healthcare was the predominant focus in the ethics literature on AI applications in health, with the ethics of AI in public health largely absent from the literature reviewed. One article that did illuminate ethical considerations for AI in public health highlighted the use of AI in environmental monitoring, motor vehicle crash prediction, fall detection, spatial profiling, and infectious disease outbreak detection, among other purposes, with the dominant ethical themes linking to data privacy, bias, and 'black box' machine learning models [82]. Other articles that mentioned public health similarly illustrated infectious disease outbreak predictions and monitoring [61, 84, 104], tracking communicable diseases [104], mental health research [105], and health behaviour promotion and management [59, 104]. However, these applications were only briefly mentioned in the broader context of primary healthcare, and few spoke to the ethics of these applications [59, 105, 106].

In the literature reviewed, there were also evident gaps in the area of global health, with few considerations of the unique ethical challenges AI poses for LMICs. Though there was mention of utilizing AI for screening in rural India [45]; genomics research in China [25]; facial recognition to detect malnutrition in Kenya [80]; and precision medicine in LMICs more broadly [103], among others, there was a significant gap in the literature commenting on the ethics of these practices in the global south. Furthermore, there was little discussion of health equity, including how the use of AI may perpetuate or exacerbate current gaps in health outcomes between and within countries. Instead, references to "global" health were often limited to global investments in AI research and development (R&D), and a number of innovations currently underway in HICs [25, 41, 49, 59, 73, 90, 107–109]. The lack of focus on global health was further reflected in the primary authorship of the literature, with a mere 5.8% (n = 6) of the reviewed literature authored by individuals from LMICs. Furthermore, 33% (n = 34) of articles had primary authorship from non-English speaking countries, which indicates that while the discourse of AI is indeed global in scope, it may only be reaching an Anglo-Saxon readership, or at the very least, an educated readership.

Discussion

Summary of evidence

Cross-cutting themes and asymmetries

In this scoping review we identified 103 records (81 academic articles and 22 grey literature articles) that addressed the ethics of AI within health, up to April 2018. Illustrated in the literature reviewed were overarching ethical concerns about privacy, trust, accountability, and bias, each of which were both interdependent and mutually reinforcing. Accountability, for instance, was a noted

concern when considering who ought to bear responsibility for AI errors in patient diagnoses [63, 65, 66], while also a recognized issue in protecting patient privacy within data sharing partnerships [59]. The security of confidential patient data, in turn, was identified as critical for eliciting patient trust in the use of AI technology for health [2]. One suggestion offered to combat the threat to citizen trust in AI is through an inclusive development process [64], a process which has also been proposed to mitigate bias integrated into algorithm development [79]. It is therefore clear from our review that the aforementioned ethical themes cannot be considered in isolation, but rather must be viewed in relation to one another when considering the ethics of AI in health.

These broad ethical themes of privacy and security, accountability and responsibility, bias, and trust have also been revealed in other reviews. In a mapping review by Morley et al. [110] on AI in healthcare, for instance, concerns of trust, 'traceability' (aligning with what we have labelled 'accountability'), and bias emerged. While privacy and security were explicitly excluded from their review [110], these very issues were a significant finding in a systematic review by Stahl et al. [111], both with regard to data privacy and personal (or physical) privacy. Issues of the autonomy and agency of AI machines, the challenge of trusting algorithms (linked with their lack of transparency), as well as others that were more closely associated with non-AI computing technologies were also discussed [111]. While the precise labels of ethical themes differed across these reviews based on the authors' analytic approach, the general challenges were common across them, and indeed, intimately interconnected. It is clear also that these broad ethical themes are not unique to health, but rather transcend multiple sectors, including policing, transportation, military operations, media, and journalism [112, 113].

An asymmetry in the literature was the predominant focus on the ethics of AI in healthcare, with less attention granted to public health, including its core functions of health promotion, disease prevention, public health surveillance, and health system planning from a population health perspective. Yet in the age of ubiquitous computing, data privacy for use in public health surveillance and interventions will be all the more critical to secure, as will ensuring that individuals and communities without access to the latest technologies are not absent from these initiatives. In a recent article, Blasimme and Vayena [114] touched upon issues of consent when employing AI-driven social media analysis for digital epidemiology; the ethics of 'nudging' people towards healthier behaviours using AI technology; and developing paternalistic interventions tailored to marginalized populations. These public health issues and others merit further exploration within the ethics literature, particularly given how powerful such AI applications can be when applied at a population level. From an alternative perspective, the increasing presence of AI within healthcare may in some respects pose a risk to public health, with an expressed concern that the 'hype' around AI in healthcare may redirect attention and resources away from proven public health interventions [103, 115]. Similarly absent in the literature was a public health lens to the issues presented, a lens which rests on a foundation of social justice to "enable all people to lead fulfilling lives" [116]. With respect to jobs, for example, the pervasive discourse around care robots in the literature suggests that there may be a wave of robots soon to replace human caregivers of the sick, elderly, and disabled. Despite this recognition, however, the focus was solely on the impact on patients, and there was little mention given to those caregivers whose jobs may soon be threatened. This is true also for other low-wage workers within health systems at large, despite the fact that unemployment is frequently accompanied by adverse health effects.

A second asymmetry in the literature was the focus on HICs, and a notable gap in discourse at the intersection of ethics, AI, and health within LMICs. Some articles mentioned the challenges of implementing the technology in low-resource settings [25, 45, 80, 102, 103, 106], and whether its introduction will further widen the development gaps between HICs and LMICs [102], however absent in most was the integration of ethics and/or health. Yet AI is increasingly being deployed in the global south; to predict dengue fever hotspots in Malaysia [59], to predict birth asphyxia in LMICs at large [36], and to increase access to primary screening in remote communities in India [45], to name a few examples. Despite these advancements, in LMIC contexts there are challenges around collecting data from individuals without financial or geographic access to health services, data upon which AI systems rely [36, 80], and a further challenge of storing data electronically [80]. The United States Agency for International Development (USAID) and the Rockefeller Foundation [117] have recently illuminated some additional considerations for the deployment of AI in LMICs, one in particular being the hesitancy of governments and health practitioners to share digital health data for concern that it could be used against them, as digitizing health data is often quite politicized for actors on the ground. Given the infancy of these discussions, however, there is far more work to be done in order to critically and collaboratively examine the ethical implications of AI for health in all corners of the world, to ensure that AI contributes to improving, rather than exacerbating health and social inequities.

Towards ethical AI for health: what is needed?

Inclusive and participatory discourse and development of ethical AI for health was commonly recommended in the literature to mitigate bias [79], ensure the benefits of AI are shared widely [59, 74, 79, 80], and to increase citizens' understanding and trust in the technology [47, 59, 64]. However, those leading the discussion on the ethics of AI in health seldom mentioned engagement with the end users and beneficiaries whose voices they were representing. While much attention was given to the impacts of AI health applications on underserved populations, only a handful of records actually included primary accounts from the people for whom they were raising concerns [2, 59, 75, 94, 118, 119]. Yet without better understanding the perspectives of end users, we risk confining the ethics discourse to the hypothetical, devoid of the realities of everyday life. This was illustrated, for instance, when participants in aged care challenged the ethical issue of care robots being considered deceptive, by stating that despite these concerns, they preferred a care robot over a human caregiver [94]. We therefore cannot rely on our predictions of the ethical challenges around AI in health without hearing from a broader mosaic of voices. In echoing recommendations from the literature, there is an evident need to gain greater clarity on public perceptions of AI applications for health, what ethical concerns end-users and beneficiaries have, and how best they can be addressed with the input of these individuals and communities. This recommendation is well aligned with the current discourse on the responsible innovation of AI, an important dimension of which involves the inclusion of new voices in discussions of the process and outcomes of AI [120].

In addition to taking a participatory approach to AI development, there is a responsibility for all parties to ensure its ethical deployment. For instance, it should be the responsibility of the producers of AI technology to advise end users, such as HCPs, as to the limits of its generalizability, just as should be done with any other diagnostic or similar technology. There is a similar responsibility for the end user to apply discretion with regards to the ethical and social implications of the technology they are using. This viewpoint is shared by Bonderman [121], who asserts that when physicians deploy AI during patient diagnoses, for instance, it is important that they remain in control, and retain the authority to override algorithms when they have certainty the algorithm outputs are incorrect [122]. Ahuja [122] compliments this assertion by stating how, since machine learning and deep learning require large quantities of data, said systems can underperform when presented with novel cases, such as atypical side effects or resistance to treatment. Simply stated, we must be critical and discretionary with regards to the application of AI in scenarios where human health and wellbeing are concerned, and we must not simply defer to AI outputs.

Also in need of critical reflection, as it remains unresolved in the literature, is how to appropriately and responsibly govern this technology [25, 45, 49, 52, 57, 102]. While there were hints in the literature regarding how to promote responsible AI, such as equal distribution of the technology, corporate transparency, and participatory development, there was little on how these recommendations could be optimally secured through regulatory mechanisms and infrastructure. The infusion of AI into health systems appears inevitable, and as such, we need to reconsider our existing regulatory frameworks for disruptive health technologies, and perhaps deliberate something new entirely. Given the challenge that many have termed the 'black box', illustrative of the fact that, on the one hand, AI processes operate at a level of complexity beyond the comprehension of many end-users, and on the other, neural networks are by nature opaque, the issue of governance is particularly salient. Never before has the world encountered technology that can learn from the information it is exposed to, and in theory, become entirely autonomous. Even the concept of AI is somewhat nebulous [2, 59, 123, 124], which threatens to cloud our ability to govern its use. These challenges are compounded by those of jurisdictional boundaries for AI governance, an ever-increasing issue given the global 'race' towards international leadership in AI development [125]. Thirty-eight national and international governing bodies have established or are developing AI strategies, with no two the same [125, 126]. Given that the pursuit of AI for development is a global endeavour, this calls for governance mechanisms that are global in scope. However, such mechanisms require careful consideration in order for countries to comply, especially considering differences in national data frameworks that pre-empt AI [49]. These types of jurisdictional differences will impact the ethical development of AI for health, and it is thus important that academic researchers contribute to the discussion on how a global governance mechanism can address ethical, legal, cultural, and regulatory discrepancies between countries involved in the AI race.

Limitations

One potential limitation to this study is that given the field of AI is evolving at an unprecedented rate [1], there is a possibility that new records in the academic and grey literatures will have been published after the conclusion of our search, and prior to publication. Some recent examples of related articles have very much been in line with our findings, drawing light to many of the pertinent ethical issues of AI in healthcare discussed in the

literature reviewed [18, 127–132]. Few, however, appear to have discussed the ethical application of AI in LMICs [117, 133] or public health [117, 130], so despite any new literature that may have arisen, there is still further work to be done in these areas. Furthermore, given our search strategy was limited to the English language, we may have missed valuable insights from publications written in other languages. The potential impact on our results is that we underrepresented the authorship from LMICs, and underreported the amount of literature on the ethics of AI within the context of LMICs. Furthermore, by not engaging with literature in other languages, we risk contradicting recommendations for an inclusive approach to the ethics discourse. Indeed, we may be missing important perspectives from a number of country and cultural contexts that could improve the ethical development and application of AI in health globally. To address this limitation, future researchers could collaborate with global partner organizations, such as WHO regional offices, in order to gain access to literatures which would otherwise be inaccessible to research teams. An additional limitation lies in our grey literature search. As part of a systematic search strategy, we pursued targeted website searches in order to identify any literature that did not emerge from our grey literature database and customized Google searches. These websites were chosen based on the expert knowledge of the research team, as well as stakeholders operating within the AI space, however there is a chance that additional relevant websites, and thus reports, proceedings, and other documents, exist beyond what was included in this review. Nevertheless, this scoping review offers a comprehensive overview of the current literature on the ethics of AI in health, from a global health perspective, and provides a valuable direction for further research at this intersection.

Conclusions

The ethical issues surrounding the introduction of AI into health and health systems are both vast and complex. Issues of privacy and security, trust, bias, and accountability and responsibility have dominated the ethical discourse to date with regard to AI and health, and as this technology is increasingly taken to scale, there will undoubtedly be more that arise. This holds particularly true with the introduction of AI in public health, and within LMICs, given that these areas of study have been largely omitted from the ethics literature. AI is being developed and implemented worldwide, and without considering what it means for populations at large, and particularly those who are hardest to reach, we risk leaving behind those who are already the most underserved. Thus, the dearth of literature on the ethics of AI within public health and LMICs points to a critical need to devote further research in these areas. Indeed, a greater concentration of ethics research into AI and health is required for all of its many applications. AI has the potential to help actualize universal health coverage, reduce health, social, and economic inequities, and improve health outcomes on a global scale. However, the bourgeoning field of AI is outpacing our ability to adequately understand its implications, much less to regulate its responsible design, development, and use for health. Given the relatively uncharted territory of AI in health, we must be diligent to both consider and respond to the ethical implications of its implementation, and whether if in every case it is indeed ethical at all. Amidst the tremendous potential that AI carries, it is important to approach its introduction with a degree of cautious optimism, informed by an extensive body of ethics research, to ensure its development and implementation is ethical for everyone, everywhere.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12910-021-00577-8.

Additional file 1. Search Strategy for the Academic Literature.

Additional file 2. Search Strategy and Results of Grey Literature Search.

Additional file 3. Data Charting Form Template.

Additional file 4. Bibliography of the 103 records included in analysis.

Abbreviations

AI: Artificial intelligence; CIHR: Canadian Institutes of Health Research; SDG(s): Sustainable Development Goal(s); LMIC(s): Low- and middle-income country/ies; HIC(s): High-income country/ies; HCP(s): Health care professionals; R&D: Research and development; CRISPR: Clustered regularly interspaced short palindromic repeats.

Acknowledgements

We would like to thank participants at the Ethics and AI for Good Health Symposium, whose thoughtful comments informed our thinking throughout the analysis of the literature. We would also like to acknowledge the kind contributions of Mikaela Grey at the Gerstein Science Information Centre, for peer reviewing our search strategy.

Authors' contributions

KM assisted in developing the search strategies, managing the data, as well as screening, charting, and analyzing the data. She was the primary contributor in writing the manuscript. NM assisted with the article screening, charting, and analysis, and was a significant contributor to the writing of the manuscript. JC and NM assisted with the article screening, data charting and analysis, and developed the graphical representations of the findings. They also supported the writing of the manuscript. VL helped to develop the search strategy to appropriately address our research question, and to write said search strategy in the methods section of the manuscript. JG, ED, RU, and DW, conceived the idea for this study, and devised its approach. They provided oversight and strategic direction, and participated in the article screening process, oversaw the data analysis, and contributed valuable feedback throughout each step of the manuscript-writing process. All authors read and approved the final manuscript.

Funding
This study was supported by funding from the Joint Centre for Bioethics (JCB) Jus Innovation Fund. The JCB Jus Innovation Fund provided salary support for trainees (Murphy, Cai, Malhotra, Malhotra) working on the project.

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author upon reasonable request.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Joint Centre for Bioethics, Dalla Lana School of Public Health, University of Toronto, 155 College Street, Suite 754, Toronto, ON M5T 1P8, Canada. [2] Office of Global Health Education and Training, Dalla Lana School of Public Health, University of Toronto, 155 College Street, Room 408, Toronto, ON M5T 3M7, Canada. [3] Division of Clinical Public Health, Dalla Lana School of Public Health, 155 College Street, Toronto, ON M5T 3M7, Canada. [4] Bridgepoint Collaboratory for Research and Innovation, Lunenfeld Tanenbaum Research Institute, Sinai Health System, 1 Bridgepoint Drive, Toronto, ON M4M 2B5, Canada. [5] Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public, Health Sciences Building, Health University of Toronto, 155 College Street, Suite 425, Toronto, ON M5T 3M6, Canada. [6] Gerstein Science Information Centre, University of Toronto, 9 King's College Circle, Toronto, ON M7A 1A5, Canada.

Received: 30 April 2020 Accepted: 20 January 2021
Published online: 15 February 2021

References

1. Schwab K. The Fourth Industrial Revolution: what it means and how to respond. World Economic Forum. 2016 [cited 2020 Sep 23]. https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/.
2. AI in the UK: ready, willing and able? United Kingdom: authority of the house of lords; 2018. (Intelligence SCoA, editor). https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf.
3. Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, et al. Deep learning for health informatics. IEEE J Biomed Health Inform. 2017;21(1):4–21.
4. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44.
5. Future of Life Institute. National and international AI strategies. n.d. https://futureoflife.org/national-international-ai-strategies/.
6. The Canadian Institute for Advanced Research 4. CIFAR Pan-Canadian Artificial Intelligence Strategy. 2020. https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy.
7. Price Waterhouse Cooper. Sizing the prize: What's the real value of AI for your business and how can you capitalise? 2017. https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html.
8. Bossmann J. Top 9 ethical issues in artificial intelligence. World Economic Forum. https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/.
9. Gibney E. The battle for ethical AI at the world's biggest machine-learning conference. Nature. 2020;577(7792):609–609.
10. Ouchchy L, Coin A, Dubljević V. AI in the headlines: the portrayal of the ethical issues of artificial intelligence in the media. AI Soc. 2020. https://doi.org/10.1007/s00146-020-00965-5.
11. Floridi L. Soft ethics: its application to the general data protection regulation and its dual advantage. Philos Technol. 2018;31(2):163–7. https://doi.org/10.1007/s13347-018-0315-5.
12. Partnership on AI. Partnership on AI. 2020. https://www.partnershiponai.org/.
13. OpenAI. OpenAI. https://openai.com/.
14. Responsible Robotics. Responsible robotics: accountable innovation for the humans behind the robots. https://responsiblerobotics.org/.
15. AI Ethics Initiative. The ethics and governance of artificial intelligence initiative. n.d. https://aiethicsinitiative.org.
16. Université de Montréal. Montréal declaration for a responsible development of artificial intelligence. p. 4–12. https://5dcfa4bd-f73a-4de5-94d8-c010ee777609.filesusr.com/ugd/ebc3a3_506ea08298cd4f8196635545a16b071d.pdf.
17. Fairness, Accountability, and Transparency in Machine Learning 12. Principles for accountable algorithms and a social impact statement for algorithms. n.d. https://www.fatml.org/resources/principles-for-accountable-algorithms.
18. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell. 2019;1(9):389–99.
19. Social Sciences and Humanities Research Council 14. Canada-UK Artificial Intelligence Initiative. 2020. https://www.sshrc-crsh.gc.ca/funding-financement/programs-programmes/canada-uk_ai/index-eng.aspx.
20. Canadian Institute for Advanced Research 15. Canada, France, UK launch research workshops exploring societal implications of Artificial intelligence. 2019. https://www.cifar.ca/cifarnews/2019/04/15/canada-france-uk-launch-research-workshops-exploring-societal-implications-of-artificial-intelligence.
21. Canadian Institute for Advanced Research 17. AI & Society. CIFAR. https://www.cifar.ca/ai/ai-society.
22. Wellcome Trust. The ethical, social and political challenges of using artificial intelligence in healthcare|Wellcome. https://wellcome.org/grant-funding/ethical-social-and-political-challenges-using-artificial-intelligence-healthcare.
23. Bhagyashree SIR, Nagaraj K, Prince M, Fall CHD, Krishna M. Diagnosis of dementia by machine learning methods in epidemiological studies: a pilot exploratory study from south India. Soc Psychiatry Psychiatr Epidemiol. 2018;53(1):77–86. https://doi.org/10.1007/s00127-017-1410-0.
24. Zhang X, Pérez-Stable EJ, Bourne PE, Peprah E, Duru OK, Breen N, et al. Big data science: opportunities and challenges to address minority health and health disparities in the 21st century. Ethn Dis. 2017;27(2):95.
25. Johnson W, Pauwels E. How to optimize human biology:27. https://www.wilsoncenter.org/sites/default/files/media/documents/publication/how_to_optimize_human_biology.pdf.
26. Ghebreyesus T. Artificial intelligence for good global summit. World Health Organization. http://www.who.int/dg/speeches/2018/artificial-intelligence-summit/en/.
27. World Health Organization. WHO guideline recommendations on digital interventions for health systems strengthening. 2019. 123 p. http://www.ncbi.nlm.nih.gov/books/NBK541902/.
28. Chan M. Opening remarks at the artificial intelligence for good global summit. World Health Organization. 2017. https://www.who.int/dg/speeches/2017/artificial-intelligence-summit/en/.
29. Koplan JP, Bond TC, Merson MH, Reddy KS, Rodriguez MH, Sewankambo NK, et al. Towards a common definition of global health. The Lancet. 2009;373(9679):1993–5.
30. Ghosh S, Mitra I, Nandy P, Bhattacharya U, Dutta U, Kakar D. Artificial intelligence in India—hype or reality: impact of artificial intelligence across industries and user groups.. Price Waterhouse Cooper India; 2018. p. 1–32. https://www.pwc.in/assets/pdfs/consulting/technology/data-and-analytics/artificial-intelligence-in-india-hype-or-reality/artificial-intelligence-in-india-hype-or-reality.pdf.
31. A blueprint for the future of AI: 2018–2019. Brookings Institute. 2020. https://www.brookings.edu/series/a-blueprint-for-the-future-of-ai/.
32. United Nations. Sustainable Development Goal 3: ensure healthy lives and promote well-being for all at all ages. n.d. https://sdgs.un.org/goals/goal3.
33. United Nations. Sustainable development goal 10: reduce inequality within and among countries. n.d. https://sdgs.un.org/goals/goal10.
34. United Nations Committee for Development Policy. Leaving no on behind. 2018. p. 1–4. https://sustainabledevelopment.un.org/content/documents/2754713_July_PM_2._Leaving_no_one_behind_Summary_from_UN_Committee_for_Development_Policy.pdf.

35. Vinuesa R, Azizpour H, Leite I, Balaam M, Dignum V, Domisch S, et al. The role of artificial intelligence in achieving the sustainable development goals. Nat Commun. 2020;11(1):233.
36. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? BMJ Glob Health. 2018;3(4):e000798. https://doi.org/10.1136/bmjgh-2018-000798.
37. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci. 2010;5(1):69. https://doi.org/10.1186/1748-5908-5-69.
38. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med. 2018;169(7):467. https://doi.org/10.7326/M18-0850.
39. Godin K, Stapleton J, Kirkpatrick SI, Hanning RM, Leatherdale ST. Applying systematic review search methods to the grey literature: a case study examining guidelines for school-based breakfast programs in Canada. Syst Rev. 2015;4(1):138.
40. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol. 2008;8(1):45.
41. Sharkey A, Sharkey N. Granny and the robots: ethical issues in robot care for the elderly. Ethics Inf Technol. 2012;14(1):27–40. https://doi.org/10.1007/s10676-010-9234-6.
42. West DM. What happens if robots take the jobs? The impact of emerging technologies on employment and public policy. Center for Technology Innovation at Brookings. 2015. https://www.brookings.edu/wp-content/uploads/2016/06/robotwork.pdf.
43. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, et al. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface. 2018;15(141):20170387. https://doi.org/10.1098/rsif.2017.0387.
44. Infosys Limited. AI for healthcare: balancing efficiency and ethics. 2018;14. https://www.infosys.com/smart-automation/docpdf/ai-healthcare.pdf.
45. Paul Y, Hickok E, Sinha A, Tiwari U, Bidare PM. Artificial intelligence in the healthcare industry in India. 2018;45. https://cis-india.org/internet-governance/files/ai-and-healtchare-report.
46. Luxton DD. Recommendations for the ethical use and design of artificial intelligent care providers. Artif Intell Med. 2014;62(1):1–10.
47. Suominen H, Lehtikunnas T, Back B, Karsten H, Salakoski T, Salanterä S. Applying language technology to nursing documents: pros and cons with a focus on ethics. Int J Med Inf. 2007;76:S293-301.
48. Crawford K, Whittaker M. The AI now report: the social and economic implications of artificial intelligence technologies in the near-term. White House and New York University's Information Law Institute; (AI Now Public Symposium 2016).
49. Denton S, Pauwels E, He Y, Johnson W. There's nowhere to hide: Artificial intelligence and privacy in the fourth industrial revolution. Wilson Center, Synenergene, and the Institute for Philosophy & Public Policy. 2018. https://iapp.org/media/pdf/resource_center/ai_and_privacy.pdf.
50. Markowetz A, Błaszkiewicz K, Montag C, Switala C, Schlaepfer TE. Psycho-informatics: big data shaping modern psychometrics. Med Hypotheses. 2014;82(4):405–11.
51. AI Now 2016 Symposium: The social implications of artificial intelligence technologies in the near-term. New York. 2016. (AI Now 2016 Primers.). https://ainowinstitute.org/AI_Now_2016_Primers.pdf.
52. Bollier D. The promise and challenge of integrating AI into cars. Healthcare and journalism: a report on the inaugural aspen institute roundtable on artificial intelligence. United States: The Aspen Institute. 2017. https://assets.aspeninstitute.org/content/uploads/2017/01/2017-Artificial-Intelligence-REP-FINAL.pdf.
53. Kantarjian H, Yu PP. Artificial intelligence, big data, and cancer. JAMA Oncol. 2015;1(5):573. https://doi.org/10.1001/jamaoncol.2015.1203.
54. Bowser A, Sloan M, Michelucci P, Pauwels E. Artificial intelligence: a policy-oriented introduction. Wilson Center. 2017. https://www.wilsoncenter.org/sites/default/files/media/documents/publication/wilson_center_policy_brief_artificial_intelligence.pdf.
55. UK Government. Guidance: initial code of conduct for data-driven health and care technology. United Kingdom: Department of Health and Social Care, editor. 2018. https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology.
56. Hengstler M, Enkel E, Duelli S. Applied artificial intelligence and trust—the case of autonomous vehicles and medical assistance devices. Technol Forecast Soc Change. 2016;105:105–20.
57. Mohandas S, Ranganathan R. AI and healthcare in India: looking forward roundtable report. India: The Centre for Internet and Society, India. 2017. https://cis-india.org/internet-governance/files/ai-and-healthcare-report.
58. Powles J, Hodson H. Google DeepMind and healthcare in an age of algorithms. Health Technol. 2017;7(4):351–67. https://doi.org/10.1007/s12553-017-0179-1.
59. Fenech M, Strukelj N, Buston O. Ethical, social, and political challenges of artificial intelligence in health. Future advocacy & wellcome trust. https://wellcome.ac.uk/sites/default/files/ai-in-health-ethical-social-political-challenges.pdf.
60. Kenneth K, Eggleton A. Challenge ahead: Integrating robotics, artificial intelligence and 3D printing technologies into Canada's healthcare systems. 2017. p. 1–44. (The Standing Senate Committee on Social Affairs SaT, editor). https://sencanada.ca/content/sen/committee/421/SOCI/reports/RoboticsAI3DFinal_Web_e.pdf.
61. Nuffield Council on Bioethics. 50. Artificial intelligence (AI) in healthcare and research. 2018. https://www.nuffieldbioethics.org/publications/ai-in-healthcare-and-research.
62. Boissoneault J, Sevel L, Letzen J, Robinson M, Staud R. Biomarkers for musculoskeletal pain conditions: use of brain imaging and machine learning. Curr Rheumatol Rep. 2017;19(1):5. https://doi.org/10.1007/s11926-017-0629-9.
63. Verghese A, Shah NH, Harrington RA. What this computer needs is a physician: humanism and artificial intelligence. JAMA. 2018;319(1):19. https://doi.org/10.1001/jama.2017.19198.
64. Stone P, Brooks R, Brynjolfsson E, Calo R, Etzioni O, Hager O, et al. Artificial Intelligence and Life in 2030. One hundred year study on artificial intelligence: report of the 2015–2016 study panel. https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai_100_report_0831fnl.pdf.
65. McBee MP, Awan OA, Colucci AT, Ghobadi CW, Kadom N, Kansagra AP, et al. Deep learning in radiology. Acad Radiol. 2018;25(11):1472–80.
66. Mesko B. The role of artificial intelligence in precision medicine. Expert Rev Precis Med Drug Dev. 2017;2(5):239–41. https://doi.org/10.1080/23808993.2017.1380516.
67. Siqueira-Batista R, Souza CR, Maia PM, Siqueira SL. Robotic surgery: bioethical aspects. ABCD Arq Bras Cir Dig São Paulo. 2016;29(4):287–90.
68. Monteith S, Glenn T. Automated decision-making and big data: concerns for people with mental illness. Curr Psychiatry Rep. 2016;18(12):112. https://doi.org/10.1007/s11920-016-0746-6.
69. Hope M. Computer-aided medicine: present and future issues of liability. Comput Law J. 1989;9(2):177–203.
70. Balthazar P, Harri P, Prater A, Safdar NM. Protecting your patients' interests in the era of big data, artificial intelligence, and predictive analytics. J Am Coll Radiol. 2018;15(3):580–6.
71. Yuste R, Goering S, Arcas BAY, Bi G, Carmena JM, Carter A, et al. Four ethical priorities for neurotechnologies and AI. Nature. 2017;551(7679):159–63.
72. O'Brolcháin F. Robots and people with dementia: Unintended consequences and moral hazard. Nurs Ethics. 2019;26(4):962–72. https://doi.org/10.1177/0969733017742960.
73. Decker M. Caregiving robots and ethical reflection: the perspective of interdisciplinary technology assessment. AI Soc. 2008;22(3):315–30. https://doi.org/10.1007/s00146-007-0151-0.
74. Russell S. Ethics of artificial intelligence. 521:415–8. https://www.nature.com/articles/521415a.pdf?origin=ppub.
75. Coeckelbergh M, Pop C, Simut R, Peca A, Pintea S, David D, et al. A survey of expectations about the role of robots in robot-assisted therapy for children with ASD: ethical acceptability, trust, sociability, appearance, and attachment. Sci Eng Ethics. 2016;22(1):47–65. https://doi.org/10.1007/s11948-015-9649-x.
76. Corbett J, d'Angelo C, Gangitano L, Freeman J. Future of health: findings from a survey of stakeholders on the future of health and healthcare in England. RAND Corporation. 2017. p. 1–90. https://www.rand.org/pubs/research_reports/RR2147.html.

77. Campolo A, Sanfilippo M, Whittaker M, Crawford K. AI now 2017 report. New York University. 2017. (AI Now 2017 Symposium and Workshop). https://ainowinstitute.org/AI_Now_2017_Report.pdf.
78. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. N Engl J Med. 2018;378(11):981–3. https://doi.org/10.1056/NEJMp1714229.
79. Howard A, Borenstein J. The ugly truth about ourselves and our robot creations: the problem of bias and social inequity. Sci Eng Ethics. 2018;24(5):1521–36. https://doi.org/10.1007/s11948-017-9975-2.
80. International Telecommunications Union, XPrize. AI for good global summit report. AI for good global summit. Geneva, Switzerland; 2017. https://www.itu.int/en/ITU-T/AI/Documents/Report/AI_for_Good_Global_Summit_Report_2017.pdf.
81. Williams AM, Liu Y, Regner KR, Jotterand F, Liu P, Liang M. Artificial intelligence, physiological genomics, and precision medicine. Phys Genomics. 2018;50(4):237–43. https://doi.org/10.1152/physiolgenomics.00119.2017.
82. Mooney SJ, Pejaver V. Big data in public health: terminology, machine learning, and privacy. Annu Rev Public Health. 2018;39(1):95–112. https://doi.org/10.1146/annurev-publhealth-040617-014208.
83. Senders JT, Zaki MM, Karhade AV, Chang B, Gormley WB, Broekman ML, et al. An introduction and overview of machine learning in neurosurgical care. Acta Neurochir (Wien). 2018;160(1):29–38. https://doi.org/10.1007/s00701-017-3385-8.
84. Lee CH, Yoon H-J. Medical big data: promise and challenges. Kidney Res Clin Pract. 2017;36(1):3–11. https://doi.org/10.23876/j.krcp.2017.36.1.3.
85. Borenstein J, Pearson Y. Robot caregivers: harbingers of expanded freedom for all? Ethics Inf Technol. 2010;12(3):277–88. https://doi.org/10.1007/s10676-010-9236-4.
86. Sharkey A. Robots and human dignity: a consideration of the effects of robot care on the dignity of older people. Ethics Inf Technol. 2014;16(1):63–75. https://doi.org/10.1007/s10676-014-9338-5.
87. Laitinen A, Niemelä M, Pirhonen J. Social robotics, elderly care, and human dignity: a recognition-theoretical approach. In: What social robots can and should do. IOS Press. 2016. https://doi.org/10.3233/978-1-61499-708-5-155
88. Vandemeulebroucke T, Dierckx de casterlé B, Gastmans C. The use of care robots in aged care: a systematic review of argument-based ethics literature. Arch Gerontol Geriatr. 2018;74:15–25.
89. Coeckelbergh M. Artificial agents, good care, and modernity. Theor Med Bioeth. 2015;36(4):265–77. https://doi.org/10.1007/s11017-015-9331-y.
90. Sharkey N, Sharkey A. The eldercare factory. Gerontology. 2012;58(3):282–8.
91. Sorell T, Draper H. Robot carers, ethics, and older people. Ethics Inf Technol. 2014;16(3):183–95. https://doi.org/10.1007/s10676-014-9344-7.
92. Sparrow R, Sparrow L. In the hands of machines? The future of aged care. Minds Mach. 2006;16(2):141–61. https://doi.org/10.1007/s11023-006-9030-6.
93. Coeckelbergh M. Health care, capabilities, and AI assistive technologies. Ethical Theory Moral Pract. 2010;13(2):181–90. https://doi.org/10.1007/s10677-009-9186-2.
94. Wachsmuth I. Robots like me: challenges and ethical issues in aged care. Front Psychol. 2018;9:432. https://doi.org/10.3389/fpsyg.2018.00432/full.
95. Coeckelbergh M. Care robots and the future of ICT-mediated elderly care: a response to doom scenarios. AI Soc. 2016;31(4):455–62. https://doi.org/10.1007/s00146-015-0626-3.
96. Gallagher A, Nåden D, Karterud D. Robots in elder care: some ethical questions. Nurs Ethics. 2016;23(4):369–71. https://doi.org/10.1177/0969733016647297.
97. Vandemeulebroucke T, de Casterlé BD, Gastmans C. How do older adults experience and perceive socially assistive robots in aged care: a systematic review of qualitative evidence. Aging Ment Health. 2018;22(2):149–67. https://doi.org/10.1080/13607863.2017.1286455.
98. Dahl T, Boulos M. Robots in health and social care: a complementary technology to home care and telehealthcare? Robotics. 2013;3(1):1–21.
99. Kohli M, Prevedello LM, Filice RW, Geis JR. Implementing machine learning in radiology practice and research. Am J Roentgenol. 2017;208(4):754–60. https://doi.org/10.2214/AJR.16.17224.
100. Kruskal JB, Berkowitz S, Geis JR, Kim W, Nagy P, Dreyer K. Big data and machine learning—strategies for driving this bus: a summary of the 2016 intersociety summer conference. J Am Coll Radiol. 2017;14(6):811–7.
101. Vogel L. What, "learning" machines will mean for medicine. Can Med Assoc J. 2017;189(16):E615–6. https://doi.org/10.1503/cmaj.1095413.
102. Ethically aligned design: a vision for prioritizing human well-being with autonomous and intelligent systems. The IEEE global initiative on ethics of autonomous and intelligent systems. n.d. https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf.
103. Mentis A-FA, Pantelidi K, Dardiotis E, Hadjigeorgiou GM, Petinaki E. Precision medicine and global health: the good, the bad, and the ugly. Front Med. 2018;5:67. https://doi.org/10.3389/fmed.2018.00067/full.
104. Albrecht S, Bouchard B, Brownstein JS, Buckeridge DL, Caragea C, Carter KM, et al. Reports of the 2016 AAAI workshop program. AI Mag. 2016;37(3):99.
105. Conway M, O'Connor D. Social media, big data, and mental health: current advances and ethical implications. Curr Opin Psychol. 2016;9:77–82.
106. Flahault A, Geissbuhler A, Guessous I, Guerin PJ, Bolon I, Marcel S, et al. Precision global health in the digital age. Swiss Med Wkly. 2017;147(1314).
107. Huschilt J, Clune L. The use of socially assistive robots for dementia care. J Gerontol Nurs. 2012;38(10):15–9. https://doi.org/10.3928/00989134-20120911-02.
108. Sharts-Hopko NC. The coming revolution in personal care robotics: what does it mean for nurses? Nurs Adm Q. 2014;38(1):5–12.
109. van Wynsberghe A. Designing robots for care: care centered value-sensitive design. Sci Eng Ethics. 2013;19(2):407–33. https://doi.org/10.1007/s11948-011-9343-6.
110. Morley J, Machado CCV, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: a mapping review. Soc Sci Med. 2020;260:113172.
111. Stahl BC, Timmermans J, Mittelstadt BD. The ethics of computing: a survey of the computing-oriented literature. ACM Comput Surv. 2016;48(4):1–38. https://doi.org/10.1145/2871196.
112. Asaro PM. AI ethics in predictive policing: from models of threat to an ethics of care. IEEE Technol Soc Mag. 2019;38(2):40–53.
113. Insights Team. Forbes insights: 4 industries that feel the urgency of AI ethics. Forbes. 2019 [cited 2020 Sep 23]. https://www.forbes.com/sites/insights-intelai/2019/03/27/4-industries-that-feel-the-urgency-of-ai-ethics/#7ec15d7372be.
114. Blasimme A, Vayena E. The ethics of AI in biomedical research, patient care and public health. SSRN Electron J. 2019. https://www.ssrn.com/abstract=3368756.
115. Panch T, Pearson-Stuttard J, Greaves F, Atun R. Artificial intelligence: opportunities and risks for public health. Lanc Dig Health. 2019;1(1):e13–4.
116. Canadian Public Health Association. Public health: a conceptual framework. 2017. https://www.cpha.ca/sites/default/files/uploads/policy/ph-framework/phcf_e.pdf.
117. US Agency for International Development 104. Artificial intelligence in global health: defining a collective path forward. 2019. p. 1–42. https://www.usaid.gov/sites/default/files/documents/1864/AI-in-Global-Health_webFinal_508.pdf.
118. Bedaf S, Marti P, De Witte L. What are the preferred characteristics of a service robot for the elderly? A multi-country focus group study with older adults and caregivers. Assist Technol. 2019 May 27;31(3):147–57. https://doi.org/10.1080/10400435.2017.1402390.
119. Draper H, Sorell T. Ethical values and social care robots for older people: an international qualitative study. Ethics Inf Technol. 2017;19(1):49–68. https://doi.org/10.1007/s10676-016-9413-1.
120. Brundage M. Artificial intelligence and responsible innovation. https://www.milesbrundage.com/uploads/2/1/6/8/21681226/ai_ri_slides.pdf.
121. Bonderman D. Artificial intelligence in cardiology. Wien Klin Wochenschr. 2017;129(23–24):866–8. https://doi.org/10.1007/s00508-017-1275-y.
122. Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. PeerJ. 2019;7:e7702.

123. Smith C, McGuire B, Huang T, Yang G. The history of artificial intelligence. https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf.
124. Robitzski D. You have no idea what artificial intelligence really does: the world of AI is full of hype and deception. Futurism. https://futurism.com/artificial-intelligence-hype.
125. Dutton T. An overview of national AI strategies. Medium. 2018. https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd.
126. Dutton T, Barron B, Boskovic G. Building an AI world: report on national and regional AI strategies. Canadian Institute for Advanced Research. 2018. https://www.cifar.ca/docs/default-source/ai-society/buildinganaiworld_eng.pdf?sfvrsn=fb18d129_4.
127. Sun TQ, Medaglia R. Mapping the challenges of artificial intelligence in the public sector: evidence from public healthcare. Gov Inf Q. 2019;36(2):368–83.
128. Guan J. Artificial intelligence in healthcare and medicine: promises, ethical challenges, and governance. Chin Med Sci J. 2019;99.
129. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat Med. 2019;25(9):1337–40.
130. Matheny M, Israni S, Ahmed M, Whicher D. Artificial Intelligence in health care: the hope, the hype, the promise, the peril. Washington, DC: National Academy of Medicine. 2019 p. 1–245. (NAM Special Publication.). https://nam.edu/wp-content/uploads/2019/12/AI-in-Health-Care-PREPUB-FINAL.pdf.
131. Powell J. Trust Me, I'm a chatbot: how artificial intelligence in health care fails the turing test. J Med Internet Res. 2019;21(10):16222.
132. Morley J, Machado CCV, Burr C, Cowls J, Joshi I, Mariarosaria T, et al. The debate on the ethics of AI in health care: a reconstruction and critical review. 2019. https://doi.org/10.13140/RG.2.2.27135.76960
133. Davies SE. Artificial intelligence in global health. Ethics Int Aff. 2019;33(02):181–92.

Publisher's Note