Health Policy and Technology 13 (2024) 100868

Contents lists available at ScienceDirect

Health Policy and Technology

journal homepage: www.elsevier.com/locate/hlpt

Review Article

Broadening the HTA of medical AI: A review of the literature to inform a tailored approach

Bart-Jan Boverhof [a,*], W. Ken Redekop [b], Jacob J. Visser [c], Carin A. Uyl-de Groot [b], Maureen P.M.H. Rutten-van Mölken [b]

[a] Erasmus School of Health Policy and Management, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3000 DR Rotterdam, the Netherlands
[b] Erasmus School of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, the Netherlands
[c] Radiology & Nuclear Medicine, Erasmus Medical Centre, Rotterdam, the Netherlands

ARTICLE INFO

Keywords:
Health-technology assessment methodology
Artificial intelligence

ABSTRACT

Objectives: As current health technology assessment (HTA) frameworks do not provide specific guidance on the assessment of medical artificial intelligence (AI), this study aimed to propose a conceptual framework for a broad HTA of medical AI.

Methods: A systematic literature review and a targeted search of policy documents was conducted to distill the relevant medical AI assessment elements. Three exemplary cases were selected to illustrate various elements: (1) An application supporting radiologists in stroke-care (2) A natural language processing application for clinical data abstraction (3) An ICU-discharge decision-making application.

Results: A total of 31 policy documents and 9 academic publications were selected, from which a list of 29 issues was distilled. The issues were grouped by four focus areas: (1) Technology & Performance, (2) Human & Organizational, (3) Legal & Ethical and (4) Transparency & Usability. Each assessment element was extensively discussed in the test, and the elements clinical effectiveness, clinical workflow, workforce, interoperability, fairness and explainability were further highlighted through the exemplary cases.

Conclusion: The current methodology of HTA requires extension to make it suitable for a broad evaluation of medical AI technologies. The 29-item assessment list that we propose needs a tailored approach for distinct types of medical AI, since the conceptualisation of the issues differs across applications.

Introduction

The number of publications on medical AI has skyrocketed in recent years [1], with the majority concerning clinical diagnostics [2]. AI is for instance leveraged to detect diabetic retinopathy in retinal images and [3] to detect large vessel occlusions on computed tomography angiography (CTA) images [4]. AI has also been used in treatment support for mental health issues [5,6], error prevention in medicine prescription [7], AI-based risk prediction and triage [8], patient flow prediction & management [9], telemonitoring & alerting [10], etc.

Consistent with the hype surrounding medical AI, increasing attention has been given to its regulation and evaluation. In Europe, guidelines were introduced through the Medical Device Regulation (MDR) in 2021. This regulation presents a novel classification of medical devices and overall requires more comprehensive evidence of efficacy and safety [11] The MDR targets medical devices including medical AI, but lacks full-fledged guidance on the regulation of AI and its unique issues. This regulatory gap is also evident in the USA [12]. Nevertheless, proceedings in AI regulation are ongoing in Europe in the form of the AI Act [11], as well as through various whitepapers in the USA [12].

Whilst the regulations are lagging, health technology assessment (HTA) could play a vital role in valuing health technologies. The HTA Core Model constitutes a good starting point, as it includes a broad selection of topics to include in the evaluation, including legal, ethical, social, and organizational aspects [13]. Despite the wide scope, however, certain AI-specific assessment elements remain unaddressed [14,15]. These include challenges regarding how to assess continuous learning AI technologies, explainability (i.e., insight into how the AI decision was made) [15], and interoperability (i.e., the degree to which an application interacts smoothly with other hospital systems) [14].

Against this background, we address this methodological gap by reviewing the policy and academic literature to identify the assessment

* Corresponding author.
E-mail address: boverhof@eshpm.eur.nl (B.-J. Boverhof).

https://doi.org/10.1016/j.hlpt.2024.100868

Available online 4 May 2024

elements that should be included in the evaluation of medical AI. By drawing from this body of literature, we propose a list of elements relevant to a comprehensive HTA of medical AI. We furthermore illustrate some of the more AI-specific assessment elements by three exemplary cases, one taken from stroke care, one from clinical data abstraction, and one from intensive care unit (ICU) decision-making.

Methodology

Systematic literature review of the academic literature

We conducted a systematic literature review (SLR) of the academic literature, targeting publications that discuss the methodological gaps in HTA of medical AI. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram is depicted in Fig. 1. The search was conducted in multiple databases: Medline, Embase, Web of Science, Cochrane, CINAHL Plus, and Scopus. We utilized a combination of Medical Subject Heading (MeSH) terms related to HTA and AI. The search strategy (provided in supplementary Table 1) was created in collaboration with an information specialist from the Erasmus University Medical Centre's academic library and carried out on November 10, 2023. The search yielded 548 eligible records after de-duplication. These were screened independently by WKR and BJB, after which BJB conducted a full-text assessment of 44 articles. Papers were included if they provided a direct description of barriers and challenges in the HTA of medical AI. The criteria for exclusion are shown in Fig. 1. This resulted in 9 included publications that provided input for the comprehensive list of assessment elements that need to be addressed in a comprehensive HTA of medical AI [15–23] (Supplementary Table 3).

Targeted review of the policy literature

We also conducted a targeted literature search of English policy publications addressing challenges and barriers surrounding medical AI, including publications of international organizations (e.g., World Health Organization), national institutions (e.g., the UK National Health Service), supranational organizations (e.g., the European Parliament), organizations representing industry (e.g. European Confederation of Pharmaceutical Entrepreneurs (EUCOPE), and consultancies (e.g., PricewaterhouseCoopers). Relevant publications were identified in three ways: 1) searching the websites of regulatory bodies, HTA agencies, governmental and non-governmental healthcare organizations, industry organizations, consultancies, and other relevant organizations considering the regulation and/or barriers of medical AI; 2) a Google search in August 2022 with the prompt "Report on issues of medical Artificial Intelligence"; and 3) through a snowballing in the already found publications until saturation was reached. This resulted in a list of 31 included documents. A full overview of all publications is provided in the supplementary Table 3.

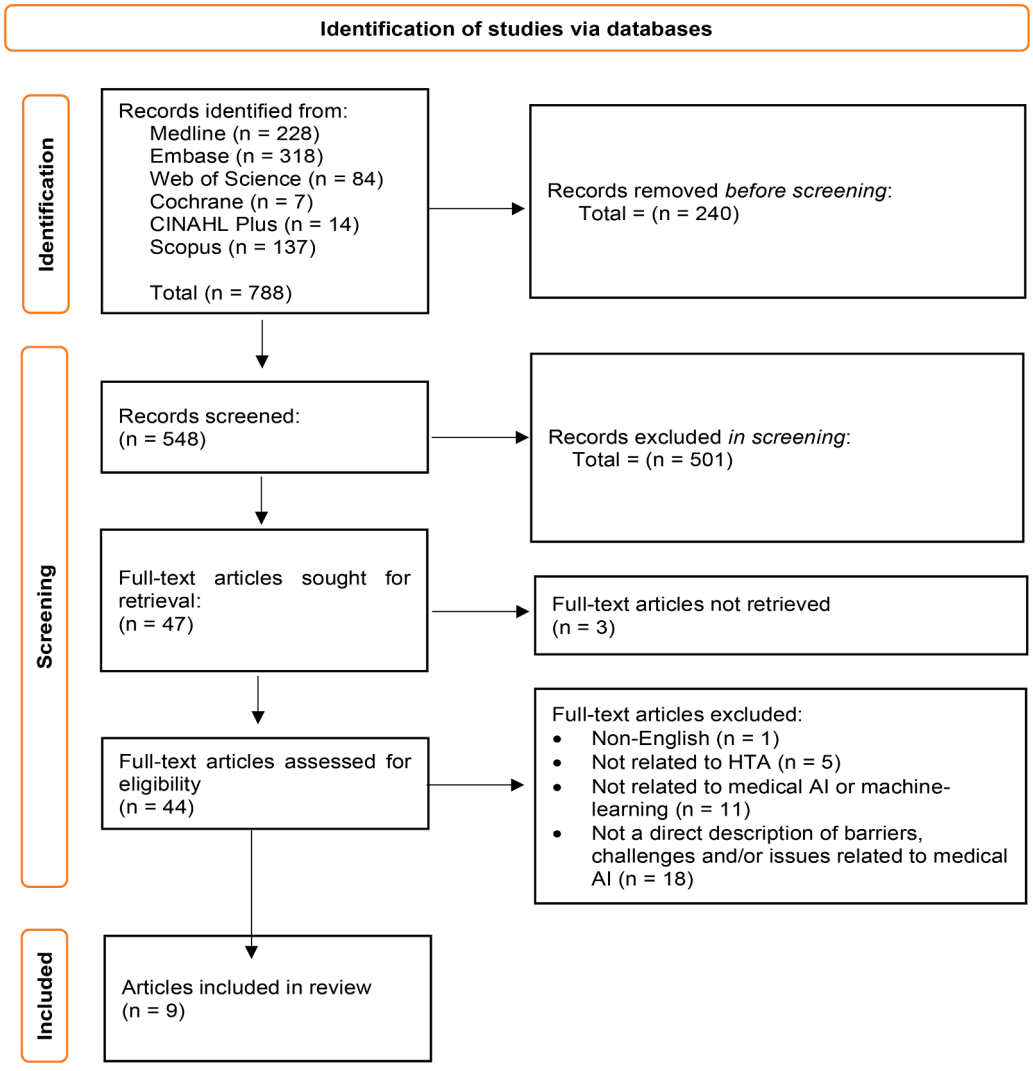

Fig. 1. PRISMA flow diagram depicting the study selection process and exclusion criteria.

Elements to address in HTA of medical AI

A synthesis of the academic publications and policy documents resulted in a comprehensive list of assessment elements for the HTA of medical AI. The EUnetHTA HTA core model served as the foundation in which we infused the additional AI-specific assessment elements [13]. All new elements were recorded and integrated into the list, regardless of how extensive they were presented in the publications. They were recorded by three iterations of full-text reading by BJB to ensure that nothing was missed. The HTA Core Model served as a foundation to ensure that our synthesized framework is sufficiently comprehensive. The elements were furthermore clustered in an iterative process between the authors, presented in a table, and explained in the text. Exemplary cases have been described in Boxes (Box 1,2,3) further clarifying some of the relevant assessment elements involved in each case. We focus on those elements that benefit from further explanation, particularly those specific to AI. The first case is an AI technology designed to assist in stroke diagnosis and decision-making. The second involves an AI technology dedicated to drawing insights from unstructured data in electronic health records. The third case describes an AI technology that supports discharge decisions for ICU patients.

Results

A total of 31 policy documents and 9 academic publications were identified for in-depth reading (see supplementary materials for reference to all documents). The academic literature included various HTA frameworks, which were mostly compiled by expert stakeholders and focused on the more AI-specific assessment elements. Our efforts in combining these existing HTA frameworks with more pragmatic insights from policy documents, as well as the HTA Core Model, resulted in a well-rounded list that can be the starting point of a comprehensive HTA of medical AI.

Our framework compromises 29 assessment elements across four main domains: Technology & Performance, Human & Organizational, Legal & Ethical, and Transparency & Usability. Table 1 shows the elements, together with brief elaborations. Fig. 1 depicts the frequency of appearance of the elements across publications. Elements such as patient privacy, conformity to regulation, training data, and fairness are addressed relatively often, whereas elements such as budget impact, environmental impact, reporting quality, and local governance are only mentioned occasionally. The following sections contain details about the different elements.

Technology and performance

The *Health Problem & PICO (Patient, Intervention, Comparator, Outcome)* element emphasizes the importance of understanding and communicating the context in which the AI is to operate. The targeted health problem must be clearly described: particularly the disease mechanism or treatment pathway that the AI technology addresses [13, 24]. Doing so should demonstrate the AI's theoretical basis and its relevance to successfully address the problem being targeted. Identifying the Population involves specifying the patients' demographics and clinical characteristics that the AI is designed to target. The Intervention must be carefully detailed in terms of the AI's functionality (including the algorithm design, training data, data processing methods, etc.). The Comparator should be defined as the existing treatment or diagnostic tool that the AI is meant to improve upon or replace. The Outcomes are the health benefits or risks the AI is expected to produce, which could range from diagnostic accuracy to patient survival rates. Formally defining these components can contribute to an understanding of the context in which the AI is to operate.

Training data revolves around the "garbage in garbage out" principle, prescribing that high-quality representative datasets are required to train an AI technology to perform well [25,26]. Such datasets can be

Table 1
Broad health technology assessment framework for a comprehensive assessment of medical artificial intelligence technologies.

Elements	Elaboration
Technology & Performance	
Health problem & PICO	To what extent is the targeted health problem, as well as the target population, (AI) intervention, comparator, and outcome (PICO) defined and communicated?
Training data	To what extent is the AI technology trained on high quality and representative data that are adequately annotated?
Patient safety	To what extent is patient safety ensured?
Clinical validation	To what extent is the AI technology validated in a clinical setting?
Clinical effectiveness	What is the clinical effectiveness of the AI application?
Cost-effectiveness	What is the cost-effectiveness of the AI technology compared to the standard of care?
Budget impact	What is the impact of adopting the AI technology on the budget?
Generalizability	To what extent does the AI technology's performance generalize across institutions / countries etc.?
Continuous learning	To what extent is the continuous learning component of the AI technology monitored?
Environmental impact	To what extent is the environmental impact of the AI technology clear?
Human & Organizational	
Clinical workflow	To what extent are changes in clinical workflow necessary to facilitate the AI technology's implementation?
Workforce	To what extent are workforce alterations and re-education necessary to facilitate the AI technology's implementation?
Stakeholder involvement	To what extent are relevant stakeholders (patients, clinician, etc.) involved in the AI technology's development process and to what extent is their satisfaction and willingness to adopt assessed?
User bias	To what extent does the human-AI collaboration pose risks concerning automation bias, aversion bias, or alert fatigue?
Interoperability	To what extent is the AI technology interoperable with other systems and data in place?
Equipment	To what extent are investments in equipment required?
Local governance	To what extent do institutions have a clear AI governance system in place?
Legal & Ethical	
Regulation	To what extent does the AI technology comply with legislation and regulation?
Intellectual property	To what extent is it clear who holds ownership over the data, algorithm, and platform?
Fairness	To what extent might the AI technology treat different subgroups unfairly?
Patient privacy	To what extent is patient privacy sufficiently addressed?
Informed consent	To what extent is patient consent with the use of AI technology ensured?
Accountability	To what extent is it clear who is held accountable in case of failure?
Cybersecurity	To what extent is cybersecurity ensured?
Environmental impact	To what extent is the environmental impact of the AI technology clear?
Transparency & Usability	
Explainability	To what extent are the AI technology's decisions explainable and should it be?
Access & availability	To what extent are data available and accessible to clinicians and researchers?
Communication	To what extent is it clearly explained what the AI technology can (and cannot) do?
Usability	To what extent is it clearly explained how the AI technology should be used?
Reporting quality	To what extent is the reporting conforming to relevant guidelines (e.g., EQUATOR guidelines, such as CONSORT, PRIMSA, CHEERS, etc.)

AI: Artificial Intelligence, CHEERS: Consolidated Health Economic Evaluation Reporting Standards, CONSORT: Consolidated Standards of Reporting Trials, EQUATOR: Enhancing the QUAlity and Transparency of health Research, PICO:

Patient, Intervention Comparator and Outcomes, PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses.

created by linking separate datasets into large multifaceted datasets [26]. High-quality annotations are an essential component of high-quality data and good AI performance [27,28], however, this is often very resource-intensive due to the need for labor-intensive expert annotation [27]. Poor annotation may also result in fairness concerns (see *fairness* element). For instance, using historical healthcare costs as an annotation proxy in predicting the healthcare needs of US citizens led to the disadvantageous treatment of black respondents (because previous costs were strongly correlated to affordability) [29]. High-quality annotated training data is thus an essential factor in AI quality and should be addressed in the assessment.

Patient safety refers to any unwanted or harmful effects that may be caused by the AI technology [24]. Clinical safety is one aspect of patient safety, constituting the AI's potentially negative impact on the well-being of patients and clinicians, as well as any incident of adverse events. Technical safety is another vital aspect, for instance requiring that the IT systems that facilitate the AI are robust towards technical failures that could comprise patient safety [20], also related to the *cybersecurity* element.

Clinical validation is important because the clinical reality is often messier than the context in which the AI was trained, potentially jeopardizing generalizability [26,30,31]. Real-world validation is therefore vital. Furthermore, many clinicians insist on seeing the tool operate within the clinical setting before fully trusting it [26], adding to the importance of clinical validation. Clinical validation is not only relevant in HTA but is also required in regulatory approval [26]. It lastly provides essential insight into how the AI fits into the actual workflow [31], further discussed in the *clinical workflow* element.

Clinical effectiveness largely depends on how the AI technology changes the existing workflows and how humans adapt to these changes and utilize the AI in daily practice. Some medical AI technologies are even more complex interventions because they change dynamically due to continuous learning. Conducting randomized controlled trials (RCTs) may consequently be challenging. This is endorsed by a 2022 systematic review, reporting on a scarcity of RCTs of medical AI applications [32]. Alternative approaches to assessing clinical effectiveness are therefore necessary for complex AI applications, such as quasi-experimental designs (Box 1 & Box 3).

Cost-effectiveness can be addressed by various approaches, such as cost-effectiveness analysis (CEA) or cost-utility analysis (CUA) if health benefits are gained (Box 3), and cost-minimization analysis (CMA) when the AI intends to only reduce costs (Box 2) [33]. A 2021 systematic review of these assessments of medical AI found various expenses to be widely overlooked (e.g., any costs accrued due to *equipment investments, cybersecurity, workforce* re-education, etc.) [34]. All relevant costs accrued in the different elements we propose should be incorporated, as neglecting them paints an unrealistic assessment of cost-effectiveness.

Budget impact analysis (BIA) assesses the affordability of the AI technology in terms of the local budget and the consequences for the budget holder. A medical AI technology may be affordable on a large scale but not in a smaller hospital ("economy of scale" principle). Institutions of different sizes may have varying affordability thresholds for medical AI applications, and BIA helps to assess affordability on the institutional level. An assessment on the local level is essential if it intends to inform decision-making locally [35].

Generalizability refers to the ability of AI to work adequately on data from a new, independent cohort of patients. Real-world data can be messier than curated data used during training and development, harming the AI's generalizability [30,31] (see also *training data*). Furthermore, if an AI algorithm is trained on data representing only a certain subgroup (e.g., a certain ethnicity), it may not generalize well to a more diverse population (see also *fairness*). Additionally, single-source training data (e.g., from a single institution or type of equipment) may not generalize well to other institutions or equipment [30,36] (Box 2). Generalizability can mostly be safeguarded by ensuring high diversity in training data [37].

Continuous learning refers to the ability of AI to continuously learn from processed data throughout its life cycle. This is a potential strength of AI, but ongoing monitoring is crucial in ensuring safety. Ongoing monitoring is furthermore essential since clinical AI is sensitive to model drift, which is defined as diminishing performance over time that occurs post-implementation due to changing circumstances [38]. In 2022, the UK National Institute for Health and Care Excellence (NICE) updated its evidence standards framework (ESF), introducing an outline for a pre-specified plan for reporting evidence of continuously learning AI technologies [39]. The FDA advocated a similar approach [12], though in both cases this is currently not implemented legally.

Environmental impact is underrepresented in medical AI, due to which the net carbon footprint remains unclear. Training and utilizing AI requires extensive computational resources, negatively impacting the carbon footprint [40]. Certain applications could, however, also reduce emissions, for instance through a reduction in commuting because of telemonitoring. Regardless, the net impact of AI in terms of carbon emission is often expected to be negative. Embedding environmental impact into HTA could very well become more commonplace, which is especially legitimate in AI assessment.

Box 1
Stroke care decision-making.

Application. Stroke care requires fast diagnosis, triage, and treatment, which is captured by the idiom "time is brain"[71]. Rapid diagnosis and treatment minimize the patient's risks of mortality and long-term morbidity: the longer the brain is deprived of oxygen and nutrients, the higher the risk of severe long-term damage, due to which it is vital to restore blood flow as quickly as possible[72]. AI can speed up various aspects of the stroke care workflow, for instance in processing and interpreting computed tomography angiography (CTA) images and generating a PDF report highlighting potential occlusions, which are a common and treatable cause of ischemic stroke[73].

Relevant HTA elements. The introduction of AI throughout the stroke pathway may pose implications for the *clinical workflow*. The traditional stroke workflow is highly optimized, as it requires coordination between the emergency, radiology, and neurology departments. Introducing AI may alter the workflow of involved clinical stakeholders. For instance, there may be a shift from detecting and classifying stroke type to overseeing the characterization of the AI. Changes in the workflow of one party may also have consequences for another, as stroke care requires careful alignment between the stakeholders. It is vital to include this in the HTA process, as it would constitute an important consequence of adopting the technology.

In the context of stroke care an optimized workflow is vital. Introduction of the AI technology may require changes in *workforce*, as all involved actors must be well-adapted to the novel workflow. Education on the fundamental principles, pitfalls, and situations in which the AI assessment requires careful monitoring is essential to facilitate safe and optimal usage of the technology.

The automated and fast dispersion of information to all involved actors has implications for *interoperability,* prescribing health technologies to be able to operate jointly with technologies already in place. For instance, an imaging AI technology for stroke should integrate with the utilized Picture Archiving and Communication System (PACS), electronic health record, etc.

Stroke AI technologies are relatively complex interventions that leverage AI to automate and optimize multiple aspects of the pathway. As mentioned, this may result in actual health improvements. The assessment of *clinical effectiveness* for such a multifaceted technology may not be straightforward, as argued in the text. Assessment of clinical value may therefore require quasi-experimental approaches since the workflow affects the entire organization, and patient-level randomization is not possible. Positioning the technology to be able to result in health outcomes also has implications for *cost-effectiveness,* for which a CUA or CEA may be best suited.

Box 2
Clinical data abstraction.

Application. The electronic health record (EHR) contains structured and unstructured data. Unstructured data is often not annotated and contains many different documents such as clinical notes, radiology reports, genomic test-results, etc. Extracting insights from unstructured data (i.e. clinical data abstraction) is highly useful, but also challenging and labor-intensive[74]. AI and Natural Language Processing (NLP) offer solutions to unlock a rich body of information from unstructured data[74,75]. AI comes in to construct a structured patient database from (unstructured) EHR data, which can be utilized to unlock a rich composite of real-world evidence. For instance, detailed patient cohorts with non-small cell lung cancer (NSCLC) can be extracted, resulting in a rich source of data with which scientific inference can be conducted[75]. Another application of clinical data abstraction with AI is the ability to detect risk profiles of patients with a rare disease, such as Transthyretin Amyloid Cardiomyopathy (ATTR-CM)[74]. Leveraging AI in clinical data abstraction may alleviate labor-intensive human work, but also has the potential to improve health outcomes by identifying patients with (rare) diseases.

Relevant HTA elements. Standardization of the EHR format has been attempted, but different institutions and departments still use varying structures and standards[76], making *interoperability* essential. The application must be adaptable to different EHR structures, which developers can for instance achieve by offering different application programming interfaces (APIs).

Differences may furthermore exist in EHR content (e.g., regarding what information is generally retained) across departments, disciplines, and institutions. Different countries furthermore retain information in different languages. AI for clinical abstraction should be *generalizable* to these different contexts, mostly ensured by diversity in the training dataset (e.g., by including different institutions and different disciplines).

Fairness can be jeopardized when the records of certain patients tend to be less populated (e.g., subgroups that have limited healthcare access). The AI may perform subpar within these groups, resulting in unfair performance. Addressing this is mostly a matter of ensuring adequate representation of such patients in the training data, which can be achieved through actively ensuring diverse representation, or post-hoc corrections (e.g., weighting or dataset linking).

Human and organizational

Clinical workflow. Early research often compared AI with human performance [41], facilitating a direct comparison between the two. Recent studies, however, have focused more on the human-AI collaboration [27,30,31,42], with the potential for significant time savings [25]. The adoption of medical AI may often require adaptations in the workflow of clinical personnel, shifts in task delegation, and (re-)distribution of "invisible" work (e.g., reassuring and reminding patients) [43]. Financial implications and willingness to adopt these changes should be addressed in the assessment (Box 1).

Workforce. Changes in the *clinical workflow* may call for workforce considerations such as training sessions on safe and responsible use, common pitfalls, etc. This could also include training in digital literacy and data science [26], which is important as AI illiteracy may encourage irrational fear (e.g., concerning job displacement). The extent to which workforce re-composition is necessary should be considered, as well as its acceptance and financial implications (Box 1 & Box 3).

Stakeholder involvement is essential in successful medical AI development. Relevant stakeholders oftentimes include clinicians, as they are in many cases the intended users of the product. Their involvement throughout the AI developmental cycle is vital, as it ensures alignment of the product with clinical needs. Stakeholder involvement furthermore contributes to the readiness of organizations to adopt AI [16]. Ensuring this involves clear communication with stakeholders (see also *communication*), setting up a training strategy for end-users (see also *workforce*), and management strategies that resonate with stakeholders. Involving stakeholders is thus vital in ensuring clinical utility and organizational readiness, which are both conducive to the successful adoption of medical AI.

User bias could hinder successful collaboration between the AI and the human user [31]. One type is automation bias, which reflects the tendency to accept AI decisions uncritically, to which inexperienced clinicians are more prone [30]. Experienced clinicians are more at risk of aversion bias, reflecting the tendency to disregard an AI technology's decision to rely on personal experience. Alert fatigue bias could occur if alerts provided by an AI system are largely ignored due to a history of too many false positives [31]. Assessments should be conducted to ascertain these biases, and proper education to prevent them.

Interoperability is the ability of digital systems to work together, often ensured by using standardized formats. The importance of achieving interoperability is widely recognized but remains challenging in many circumstances [44]. Interoperability should be included in the assessment, for instance, with the UK National Health Service (NHS) digital technology assessment criteria that guide interoperability in program interface, data format, electronic health record communication, and communication with other devices [45] (Box 1 & Box 2).

Equipment requirements in the form of hardware investments should be considered since utilizing AI may demand substantial computational resources [30]. Commercial technologies, however, typically run in the cloud, due to which extra on-site hardware investments are not always relevant. Furthermore, equipment is closely linked to the *cybersecurity* element as ensuring this may require higher investments in equipment. The assessment should incorporate any (financial) implications including acquisition, maintenance, and operation.

Local governance structures are vital in the management of hospital-wide medical AI. Most publications advocate for a multidisciplinary team to facilitate and oversee the adoption of AI, which could include the identification of institutional processes that can be automated with AI [46], making an inventory of potential solutions, and upholding ethical principles and legal oversight [47]. An institution-wide governance structure for AI thus plays an important role in upholding

Box 3
Intensive care unit discharge decision-making.

Application. Optimizing the moment of discharge on the intensive care unit (ICU) is crucial, as too late ICU discharge causes mental distress in patients and their loved ones and too early discharge increases the risk of re-entry and mortality, posing vital concerns regarding *patient safety*[9]. Optimizing discharge timing is also crucial in optimal resource management, ensuring that beds are available for those patients in need[9]. AI can be used to optimize the discharge decision-making process by leveraging patient demographic data (e.g., age and sex), clinical observations (e.g., nursing scores and Glasgow Coma Scale score), and automated physiological measurements from equipment and devices[77]. This may result in positive patient health gains, which has been shown by an early *cost-effectiveness* analysis[9].

Relevant HTA elements. When using a wide variety of sensitive patient data, biases may be amplified, and *fairness* concerns arise. Physiological health markers used by ICU discharge technologies[77] (e.g., as arterial blood pressure) depend upon lifestyle and may differ across socio-demographic subpopulations. If such patients were underrepresented in the training data, the AI is prone to reach subpar performance and exacerbate health inequalities. The fact that AI influences the allocation of scarce resources (ICU beds) gives further rise to the importance of *fairness* across subgroups[67]. An increased *fairness* risk gives rise to an increasing importance of *explainability*, since ensuring explainable AI (especially when sensitive data is used) will help to uncover any violations of fairness.

An elevated *fairness* risk furthermore raises concerns surrounding *accountability*, since clear legal assertion of who should be held accountable in case of unfair treatment may foster more awareness. This may contribute to reducing the risk of unfair treatment, which is especially important in a situation where the stakes are high, such as ICU decision-making.

Lastly, it is vital for such a high-stake AI application that adequate *workforce* re-education is conducted. Intensivists should be educated on the potential pitfalls of the AI application, to ensure that any potential shortcomings are recognized and handled.

oversight and could play a role in assessment.

Legal and ethical

Regulation for medical AI is vital, but guidelines often lag technological advancement. This is endorsed by a survey of various public authorities across the European Union (EU), which found that clear legislation was lacking in 75 % of the cases [48]. Regulation on broader digital health technologies has been overhauled in the EU MDR, but dedicated AI regulation is coming in the form of the EU AI act [11]. Although exact regulatory demands are often unclear, developers of AI technologies should conform to regulations to the extent possible.

Intellectual property (IP) legislation is critical for data ownership. Developers of medical AI should be aware of local IP laws for patient data, especially if data is monetized [49]. From a developer's perspective, challenges may arise in the IP protection of their algorithms, as they can be a black-box and difficult to protect [50]. It is thus vital to consider IP rights as part of the assessment.

Fairness issues are mostly related to unrepresentative training datasets. Heterogeneous data is especially for AI processing sociodemographic and clinical characteristics (such as genotypes and phenotypes). If certain subgroups are underrepresented in the training data, AI performance will be subpar for these groups in the clinical reality [37,49,50], resulting in discriminatory outcomes [37]. Incorporating fairness into the assessment is therefore vital (Box 2 & Box 3).

Patient privacy is highly relevant in medical AI due to the need for large, multifaceted, and rich training datasets. Stricter privacy regulations (as occurred through the EU's General Data Protection Regulation) could limit the widespread availability of such data and affect performance. Pseudonymization and anonymization of data can increase patient privacy [49], but privacy violations remain possible (e.g., trough linkage) [51]. A state-of-the-art approach is differential privacy, which adds the exact amount of random noise such that any individual dataset remains completely private [51]. Privacy is furthermore the most frequently appearing element among the included publications (Fig. 2) and should naturally be incorporated into the assessment (Box 1).

Informed consent prescribes that patients should actively consent to the use of a particular intervention [36]. This is often not the case when medical AI is used, especially if it is part of routine procedures [52]. A reason for this is a lack of clear regulation, although this is under development [53]. It might not always be necessary to ask patients for informed consent, for instance when the medical AI is primarily used to improve efficiency with little impact on clinical decision making. However, clear guidance on when to ask or not to ask for approval is needed.

Accountability concerns responsibility in case of AI misprediction or failure. In reality, clinicians are typically held accountable, but there is often a lack of clarity surrounding this [50]. This ambiguity is especially problematic with opaque algorithms that lack *explainability,* as this leaves clinicians unable to reassure the logic of the AI's choice. A lack of clarity on accountability is harmful and impedes the clinician's willingness to adopt [36]. Clear legislation on accountability is therefore essential, and any accountability issues should incorperated (Box 3).

Cybersecurity measures are necessary as medical AI often processes sensitive medical data. It could be ensured through multi-factor authentication or anomaly detection of unusual activity [50]. Ensuring the security of the algorithm itself is also crucial, calling for security measures throughout the entire IT sphere (e.g., protecting data storage, defending against DDoS attacks, etc.) [50]. Cybersecurity must be considered, which can be done with the NHS's extensive checklist for data and algorithm security [45] (Box 1).

Transparency and usability

Explainability refers to AI technology's ability to present an outcome in understandable terms to the human user [54]. Various publications argue that explainability is a precursor to user trust [47,55–57], *fairness* [55], and clarification of *accountability* [36]. Some argue, however, that

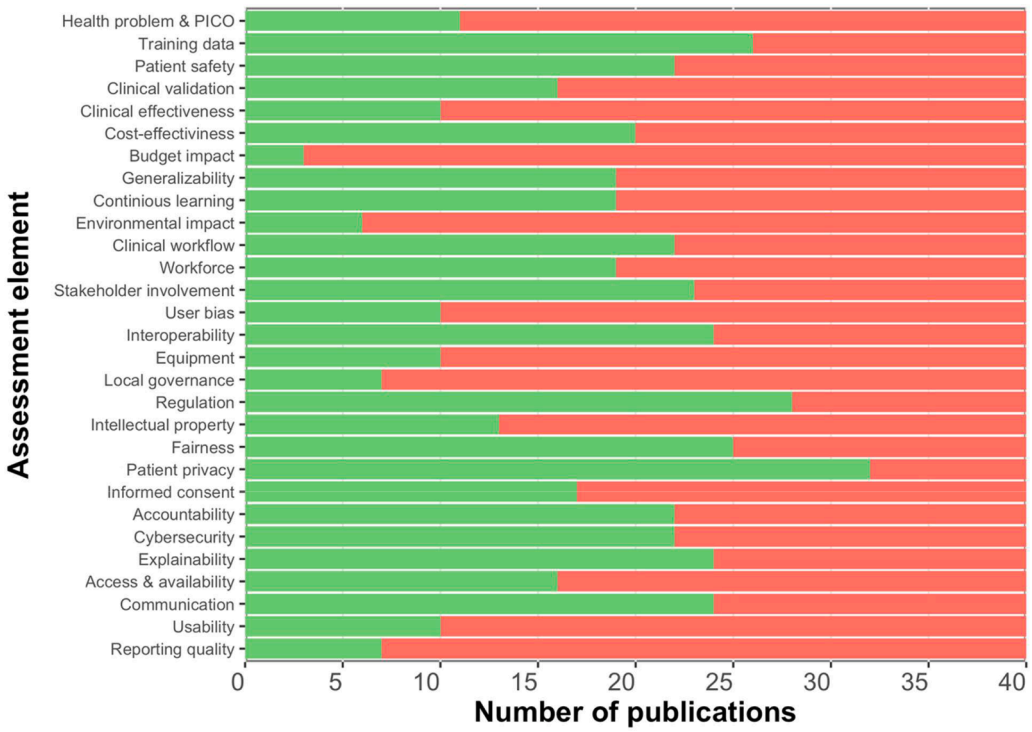

Fig. 2. Appearance of relevant elements across the included publications (grey and white literature). The green bars reflect the number of publications in which the respective elements have appeared. Elements were marked to appear if mentioned as relevant anywhere throughout the publication, regardless of how extensive the elaboration was.
PICO: Patient, Intervention Comparator and Outcomes.

fully explainable AI is often impossible to reach due to the complexity of most models [26,55]. In applications where explainability is imperative, they argue, we should rely on naturally explainable techniques such as regression methods. This may be feasible in some circumstances, but simpler models perform substantially worse with highly complex data, such as medical images, genomic data, etc. Explainability is perhaps not required if safety and robustness are guaranteed, but it becomes increasingly important when *fairness* is jeopardized [55], for instance, if the AI contributes to the allocation of scarce resources between patients (Box 3). Explainability should thus become integral to the assessment of various subtypes of medical AI, especially where high-stake decisions are involved.

Access & availability of the data concerns availability to clinicians, researchers and (AI) developers. Large datasets are important (see also *training data*), but potentially endanger *patient privacy*. Creating large datasets from scratch is very expensive, due to which more widespread availability of existing data is highly beneficial. This would, however, raise concerns about *patient privacy* and *cybersecurity*. This is where synthetic data could come in, which is a newly generated dataset based on the original source, but respecting patient privacy [58].

Communication should be clear and transparent about AI capabilities and limitations to establish (user) trust [27,58–60]. Over-hype in the 1970s resulted in the AI winters, where funding diminished due to unrealistic expectations [27,60]. A clear and realistic communication strategy helps to manage realistic expectations, fosters user trust, and should therefore be assessed.

Usability is considered a major driving force behind adoption [61], as users are often clinicians without expertise in AI. Creating an easy-to-use product with clear guidelines is therefore vital. The NHS provides guidance to ensure usability [45], and including this in HTA is vital.

Reporting quality in HTA studies that evaluate medical AI technologies leaves room for improvement. This has been underscored by various systematic reviews, showing that HTA studies failed to report on key elements [19,34]. Standardized reporting guidelines could help to improve reporting quality, such as those offered by the EQUATOR Network (specifically the CHEERS guidelines for economic evaluations) [62]. The list of elements outlined in this paper can contribute to ensuring adequate reporting on HTA for medical AI.

Discussion

We have performed a systematic review of the academic literature and a targeted review of policy documents on barriers, challenges, and HTA assessment elements to inform an extension of the methodology for HTA of medical AI. This resulted in a comprehensive list of 29 elements, divided over four domains. We elaborated on all elements in the text, and some of the more AI-specific ones were further elaborated on with reference to the exemplary cases. As for the assessment of many health technologies, we think that the EUnetHTA Core Model is an adequate start for HTA of medical AI, despite that it does not address various elements presented in Table 1. This research could firstly provide an amendment of existing HTA frameworks to AI-specific assessment elements, and secondly, help reposition original elements for medical AI. Consequently, our list could benefit (HTA) researchers, clinicians considering the adoption of medical AI in their work, healthcare institutions tasked with the acquisition of medical AI, as well as developers creating it.

We found a sizable difference in how often elements appeared across publications, with patient privacy, training data, regulation, and fairness appearing most frequently. This could be because they are most important, most discussed in the public discourse, or a combination. Various publications specifically targeted a single element such as fairness [37] or explainability [56], and many publications focused on a specific subset of legal and ethical elements [36,42,47,59,63–65]. Some publications focused on delineating a subset of AI-specific HTA elements [16,17,20–22] but did not intend to provide a comprehensive list. By drawing from all these publications, our list comprehensively covers HTA of medical AI broadly.

Most publications presented the elements one-dimensionally, without acknowledging that urgency varies between different types of medical AI. Contemplation of elements, such as informed consent and accountability, in relation to the risk factor of a particular medical AI technology helps to illustrate this. The risk of a medical AI device is legally determined by (1) the state of the patient's condition (from non-serious to critical) and (2) the AI's impact on clinical decision-making (from low to high influence) [66]. If the AI application has little influence on the decision-making, the clinician's decision is mostly independent, making concerns surrounding informed consent and accountability less pressing. As medical AI influences diagnostic or treatment decisions more strongly, concerns surrounding consent and accountability become increasingly pressing.

Differences in AI applications and their most pressing elements were likewise seen across the exemplary cases. For instance, how to approach cost-effectiveness depends largely on the purpose of the AI technology and its clinical objective (i.e., health outcomes, workflow efficiency, or both). When the objective is to enhance efficiency a relatively simple cost-minimization analysis can suffice, but a technology setting to increase patient health might require cost-effectiveness or cost-utility analysis.

A similar consideration surrounds the elements of explainability, characterized by opposing views in the literature. If an AI technology involves the allocation of (scarce) resources and/or leverages sensitive information, explainability becomes more pressing in monitoring whether patients with different socio-demographic profiles are treated fairly [67]. In scenarios where the AI technology performs diagnostic tasks with a proficiency comparable to that of clinicians, explainability becomes less urgent as clinicians can use their own expertise to validate or override the AI's decisions, for instance in radiology. We do not argue that explainability should be neglected, as there is still a benefit in it [68], but rather it illustrates the point that assessment elements have varying urgency across different applications.

This upholds similarly for patient privacy and cybersecurity which become more urgent if sensitive data is processed. Accountability and workforce become more pressing when the AI technology has a strong influence on decision-making and/or the stakes are high. Workforce and clinical workflow become more urgent if the integration of the application is highly disruptive to the current standard of care.

It may be clear that elements carry a different urgency across applications. It is therefore vital that the HTA approach is tailored to the specific application under assessment. One could do this by weighting the various elements according to their importance for the case, as is done with multi-criteria decision analysis [69].

The assessment furthermore strongly relies on the local context in which the AI technology might be embedded. This translation is vital, as the general assessment of an AI technology does not necessarily translate towards the local context, calling for a different perspective [35]. In follow-up research we endeavor to explore and employ the elements in an HTA of a medical AI technology, taking the next step towards adequate HTA of AI.

Our approach has various shortcomings. In our targeted literature review of policy documents, we may have missed relevant publications from institutions that were not targeted or did not appear in the Google searches. In our systematic review of the academic literature, we might have missed publications that were not available in one of the six targeted databases. Moreover, we restricted both searches to publications written in English. In the targeted search of policy documents, we focused on publications from institutions in the UK, US, European Union, and globally operating. This is expected to be no issue, however, as institutions from those countries are the most active explorers of medical AI strategies. Both the UK and the US are furthermore classified within the top four "AI-ready countries" [70], due to which it is unlikely that broadening to other national institutions would have revealed novel

elements.

Conclusion

This review of the literature on barriers, challenges, and assessment elements has identified and explored the vital elements in conducting a health technology assessment of medical AI. We have advocated for an extension of the current methodology that incorporates elements such as interoperability, explainability, accountability, and cybersecurity. A total of 29 elements were identified and organized over the focus areas: (1) Technology & Performance, (2) Human & Organizational, (3) Legal & Ethical and (4) Transparency & Usability. With reference to our list of elements, HTA researchers ensure a comprehensive overview of the value of a particular medical AI application for different stakeholders.

Author contribution

Concept and design: Boverhof, Redekop, Rutten-van Mölken
Acquisition of data: Boverhof
Analysis and interpretation of data: Boverhof, Redekop, Rutten-van Mölken
Drafting of the manuscript: Boverhof
Critical revision of the paper for important intellectual content: Boverhof, Redekop, Rutten-van Mölken, Visser, Uyl-de Groot
Supervision: Redekop, Rutten-van Mölken

Funding

None.

Competing interests

Jacob J. Visser is medical advisor for Contextflow, Quibim, Noaber Ventures, NLC Ventures.

Ethical approval

Not required.

Acknowledgements

None.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.hlpt.2024.100868.

References

[1] Secinaro S, Calandra D, Secinaro A, Muthurangu V, Biancone P. The role of artificial intelligence in healthcare: a structured literature review. BMC Med Inform Decis Mak 2021;21(1):1–23.
[2] Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017;2(4).
[3] Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. Ophthalmology 2017;124(7):962–9.
[4] van Leeuwen KG, Meijer FJA, Schalekamp S, Rutten M, van Dijk EJ, van Ginneken B, et al. Cost-effectiveness of artificial intelligence aided vessel occlusion detection in acute stroke: an early health technology assessment. Insights imaging 2021;12(1):133. Sep 25.
[5] Woebot Health. Woebot Health. 2022 [cited 2022 Sep 10]. What powers Woebot. Available from: https://woebothealth.com/what-powers-woebot/.
[6] Darcy A, Daniels J, Salinger D, Wicks P, Robinson A, others. Evidence of human-level bonds established with a digital conversational agent: cross-sectional, retrospective observational study. JMIR Form Res 2021;5(5):e27868.
[7] Rozenblum R, Rodriguez-Monguio R, Volk LA, Forsythe KJ, Myers S, McGurrin M, et al. Using a machine learning system to identify and prevent medication prescribing errors: a clinical and cost analysis evaluation. The Joint Commission Journal on Quality and Patient Safety 2020;46(1):3–10.
[8] Babylon Health [Internet]. [cited 2022 Sep 12]. Our AI-powered Symptom Checker can analyse your symptoms, potential causes and possible next steps. Available from: https://www.babylonhealth.com/en-gb/product/ask-babylon.
[9] de Vos J, Visser LA, de Beer AA, Fornasa M, Thoral PJ, Elbers PW, et al. The Potential Cost-Effectiveness of a Machine Learning Tool That Can Prevent Untimely Intensive Care Unit Discharge. Value in Health 2022;25(3):359–67.
[10] Halcox JP, Wareham K, Cardew A, Gilmore M, Barry JP, Phillips C, et al. Assessment of remote heart rhythm sampling using the AliveCor heart monitor to screen for atrial fibrillation: the REHEARSE-AF study. Circulation 2017;136(19):1784–94.
[11] Det Norske Veritas. AI Medical Device Software under the MDR. Det Norske Veritas 2022.
[12] U.S. Food and Drug Administratio. Artificial intelligence and machine learning (AI/ML) software as a medical device action plan [Internet]. U.S. Food and Drug Administration; 2021. Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device.
[13] Kristensen FB, Lampe K, Wild C, Cerbo M, Goettsch W, Becla L. The HTA Core Model®—10 Years of Developing an International Framework to Share Multidimensional Value Assessment. Value in Health 2017;20(2):244–50.
[14] Kolasa K, Kozinski G. How to Value Digital Health Interventions? A Systematic Literature Review. International Journal of Environmental Research and Public Health [Internet] 2020;17(6). Available from: https://www.mdpi.com/1660-4601/17/6/2119.
[15] Alami H, Lehoux P, Auclair Y, de Guise M, Gagnon MP, Shaw J, et al. Artificial intelligence and health technology assessment: anticipating a new level of complexity. J Med Internet Res 2020;22(7):e17707.
[16] Alami H, Lehoux P, Denis JL, Motulsky A, Petitgand C, Savoldelli M, et al. Organizational readiness for artificial intelligence in health care: insights for decision-making and practice. J Health Organ Manag 2021;35(1):106–14.
[17] Belisle-Pipon JC, Couture V, Roy MC, Ganache I, Goetghebeur M, Cohen IG. What Makes Artificial Intelligence Exceptional in Health Technology Assessment? Front artif intell 2021;4:736697.
[18] Carter SM, Rogers W, Win KT, Frazer H, Richards B, Houssami N. The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. BREAST 2020;49:25–32. Feb.
[19] Farah L., Davaze-Schneider J., Martin T., Nguyen P., Borget I., Martelli N. Are current clinical studies on artificial intelligence-based medical devices comprehensive enough to support a full health technology assessment? A systematic review. Artif Intell Med. 6AD;140:102547.
[20] Fasterholdt I, Kjølhede T, Naghavi-Behzad M, Schmidt T, Rautalammi QTS, Hildebrandt MG, et al. Model for ASsessing the value of Artificial Intelligence in medical imaging (MAS-AI). Int J Technol Assess Health Care 2022;38(1):e74. Jan.
[21] Fasterholdt I., Naghavi-Behzad M., Rasmussen B.S.B., Kjolhede T., Skjoth M.M., Hildebrandt M.G., et al. Value assessment of artificial intelligence in medical imaging: a scoping review. BMC med imaging. 10 31;22(1):187.
[22] Hendrix N, Veenstra DL, Cheng M, Anderson NC, Verguet S. Assessing the Economic Value of Clinical Artificial Intelligence: Challenges and Opportunities. Value in Health 2022;25(3):331–9.
[23] Willem T, Krammer S, Böhm AS, French LE, Hartmann D, Lasser T, et al. Risks and benefits of dermatological machine learning health care applications—An overview and ethical analysis. J Eur Acad Dermatol Venereol 2022;36(9):1660–8.
[24] EUnetHTA. HTA core model ® version 3.0 (Pdf) [Internet]. EUnetHTA. EUnetHTA; 2016. Available from: www.htacoremodel.info/BrowseModel.aspx.
[25] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25(1):44–56.
[26] EIT Health, McKinsey & Company. Transforming healthcare with ai. EIT Health; 2021.
[27] Medical Futurist. A guide to artificial intelligence in healthcare. Medical Futurist; 2021.
[28] Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. Ann Surg 2018;268(1):70.
[29] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science (1979) 2019;366(6464):447–53.
[30] The Academy of Medical Sciences. Artificial intelligence and health. The Academy of Medical Sciences; 2019.
[31] National health service (NHS) ai lab, health education england. understanding healthcare workers' confidence in ai. National Health Service (NHS); 2022.
[32] Lam TYT, Cheung MFK, Munro YL, Lim KM, Shung D, Sung JJY. Randomized Controlled Trials of Artificial Intelligence in Clinical Practice: Systematic Review. J Med Internet Res 2022;24(8):e37188. Aug 25.
[33] Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford university press; 2015.
[34] Voets MM, Veltman J, Slump CH, Siesling S, Koffijberg H. Systematic Review of Health Economic Evaluations Focused on Artificial Intelligence in Healthcare: The Tortoise and the Cheetah. Value in Health 2021.
[35] Boverhof B.J., Redekop K., Bos D., Starmans M., Birch J., Rockall A., et al. Radiology AI Deployment and Assessment Rubric (RADAR) for value-based AI in Radiology. Insights into Imaging. In press.
[36] European Parliamentary Research Service. Artificial intelligence in healthcare: applications, risks and ethical and societal implications. European Parliament; 2022.
[37] Imperial college London. Addressing racial and ethnic inequities in datadriven health technologies. Imperial college London; 2022.

[38] Feng J, Phillips RV, Malenica I, Bishara A, Hubbard AE, Celi LA, et al. Clinical artificial intelligence quality improvement: towards continual monitoring and updating of AI algorithms in healthcare. npj Digit Med 2022;5(1):1–9. May 31.
[39] National Institute for Health and Care Excellence (NICE). Evidence standards framework for digital health technologies. National Institute for Health and Care Excellence (NICE); 2018.
[40] Organisation for Economic Co-operation and Development. Laying the foundations for artificial intelligence in health. Organisation for Economic Co-operation and Development; 2021.
[41] Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med 2022:1–8.
[42] High-Level Expert Group on Artificial Intelligence. Assessment list for trustworthy artificial intelligence (ALTAI). European Commission; 2020.
[43] Carboni C, Wehrens R, van der Veen R, de Bont A. Conceptualizing the digitalization of healthcare work: A metaphor-based Critical Interpretive Synthesis. Soc Sci Med (1967) 2022;292:114572. Jan 1.
[44] European Commission. Interoperability of electronic health records in the eu. European Commission; 2021.
[45] National Health Service (NHS). The digital technology assessment criteria for health and social care (DTAC). National Health Service (NHS); 2021.
[46] Deloitte. The future of artificial intelligence in health care. Deloitte; 2019.
[47] High-Level Expert Group on Artificial Intelligence. Ethics guidelines for trustworthy ai. European Commission; 2019.
[48] PricewaterhouseCoopers. Sherlock in health. how artificial intelligence may improve quality and efficiency, whilst reducing healthcare costs in europe. PricewaterhouseCoopers; 2021.
[49] Academy of Medical Royal Colleges. Artificial intelligence in healthcare. Academy of Medical Royal Colleges; 2021.
[50] Cloud Security Alliance. Artificial intelligence in healthcare. Cloud Security Alliance; 2022.
[51] Zhu T, Li G, Zhou W, Yu PS. Differentially Private Data Publishing and Analysis: A Survey. IEEE Trans Knowl Data Eng 2017;29(8):1619–38. Aug.
[52] Cohen I.G. Informed Consent and Medical Artificial Intelligence: What to Tell the Patient? [Internet]. Rochester, NY; 2020 [cited 2023 Feb 21]. Available from: https://papers.ssrn.com/abstract=3529576.
[53] Astromskė K, Peičius E, Astromskis P. Ethical and legal challenges of informed consent applying artificial intelligence in medical diagnostic consultations. AI & Soc 2021;36(2):509–20. Jun 1.
[54] Doshi-Velez F., Kim B. Towards a rigorous science of interpretable machine learning. arXiv preprint 2017.
[55] EIT Health. Contribution to the discussion on the european commission's data strategy and ai white paper. EIT Health; 2020.
[56] Future Advocacy. Explinability in data-driven health and care technology. Future Advocacy; 2021.
[57] U.S. Government Accountability Office. Artificial intelligence in health care benefits and challenges of technologies to augment patient care. U.S. Government Accountability Office; 2020.
[58] The Academic Health Science Networks Network. Accelerating artificial intelligence in health and care: results from a state of the nation survey. Department of Health & Social Case and NHS; 2018.
[59] Deloitte. The socio-economic impact of ai in healthcare. Deloitte; 2020.
[60] European Confederation of Pharmaceutical Entrepreneurs (EUCOPE). Artificial intelligence in health: a european opportunity. European Confederation of Pharmaceutical Entrepreneurs (EUCOPE); 2021.
[61] Organisation for Economic Co-operation and Development. Empowering the health workforce: strategies to make the most of the digital revolution. Organisation for Economic Co-operation and Development; 2021.
[62] Kirtley, Shona. Reporting guidelines. 2020; Available from: https://www.rcslt.org/wp-content/uploads/2021/01/Feb-24-25_RandD-Colu-1.pdf.
[63] Future Advocacy. Ethical, social and political challenges of artificial intelligence. Future Advocacy; 2018.
[64] EIT Health. AI and ethics in the health innovation community. EIT Health; 2019.
[65] World Health Organization. Big data and artificial intelligence for achieving universal health coverage: an international consultation on ethics. World Health Organization; 2017.
[66] Beckers R, Kwade Z, Zanca F. The EU medical device regulation: Implications for artificial intelligence-based medical device software in medical physics. Physica Medica 2021;83:1–8. Mar 1.
[67] Babic B, Gerke S, Evgeniou T, Cohen IG. Beware explanations from AI in health care. Science (1979) 2021;373(6552):284–6. Jul 16.
[68] Groen AM, Kraan R, Amirkhan SF, Daams JG, Maas M. A systematic review on the use of explainability in deep learning systems for computer aided diagnosis in radiology: Limited use of explainable AI? Eur J Radiol 2022;157:110592.
[69] Thokala P, Duenas A. Multiple criteria decision analysis for health technology assessment. Value in health 2012;15(8):1172–81.
[70] Oxford Insights, International Development Research Centre. Artificial intelligence readiness index 2019. Oxford Insights; 2019.
[71] Patil S, Rossi R, Jabrah D, Doyle K. Detection, Diagnosis and Treatment of Acute Ischemic Stroke: Current and Future Perspectives. Front Med Technol 2022;4.
[72] Musuka TD, Wilton SB, Traboulsi M, Hill MD. Diagnosis and management of acute ischemic stroke: speed is critical. Cmaj 2015;187(12):887–93.
[73] Fasen BA, Berendsen R, Kwee RM. Artificial intelligence software for diagnosing intracranial arterial occlusion in patients with acute ischemic stroke. Neuroradiology 2022:1–5.
[74] Estevez M, Benedum CM, Jiang C, Cohen AB, Phadke S, Sarkar S, et al. Considerations for the Use of Machine Learning Extracted Real-World Data to Support Evidence Generation: A Research-Centric Evaluation Framework. Cancers (Basel) 2022;14(13):3063.
[75] Rich A.S., Leybovich B., Estevez M., Irvine J., Singh N., Cohen A.B., et al. Extracting non-small cell lung cancer (NSCLC) diagnosis and diagnosis dates from electronic health record (EHR) text using a deep learning algorithm. 2021.
[76] Sreenivasan M, Chacko AM. Interoperability issues in EHR systems: Research directions. Data analytics in biomedical engineering and healthcare 2021:13–28.
[77] Thoral PJ, Fornasa M, de Bruin DP, Tonutti M, Hovenkamp H, Driessen RH, et al. Explainable machine learning on AmsterdamUMCdb for ICU discharge decision support: uniting intensivists and data scientists. Crit Care Explor 2021;3(9).