Health Information and Quality Authority

An tÚdarás Um Fhaisnéis agus Cáilíocht Sláinte

Evidence review to inform the development of National Guidance for the Responsible and Safe use of Artificial Intelligence in Health and Social Care Services

January 2026

Supporting National Standards

About the Health Information and Quality Authority (HIQA)

The Health Information and Quality Authority (HIQA) is an independent statutory body established to promote safety and quality in the provision of health and social care services for the benefit of the health and welfare of the public.

Reporting to the Minister for Health and engaging with relevant government Ministers and departments, HIQA has responsibility for the following:

- **Setting standards for health and social care services** — Developing person-centred standards and guidance, based on evidence and international best practice, for health and social care services in Ireland.

- **Regulating social care services** — The Chief Inspector of Social Services within HIQA is responsible for registering and inspecting residential services for older people and people with a disability, and children's special care units.

- **Regulating health services** — Regulating medical exposure to ionising radiation.

- **Monitoring services** — Monitoring the safety and quality of permanent international protection accommodation service centres, health services and children's social services against the national standards. Where necessary, HIQA investigates serious concerns about the health and welfare of people who use health services and children's social services.

- **Health technology assessment** — Evaluating the clinical and cost effectiveness of health programmes, policies, medicines, medical equipment, diagnostic and surgical techniques, health promotion and protection activities, and providing advice to enable the best use of resources and the best outcomes for people who use our health service.

- **Health information** — Advising on the efficient and secure collection and sharing of health information, setting standards, evaluating information resources and publishing information on the delivery and performance of Ireland's health and social care services.

- **National Care Experience Programme** — Carrying out national service-user experience surveys across a range of health and social care services, with the Department of Health and the HSE.

Visit www.hiqa.ie for more information.

Table of contents

Tables and Figures 4

Acknowledgements 4

Executive Summary 5

1. Introduction 9
   1.1. What is Artificial Intelligence? 9
   1.2. Artificial Intelligence in health and social care 9
   1.3. Previous research, strategies, and policy documents 12
   1.4. The Irish context 13
   1.5. Purpose of this review 14
2. Methods 14
   2.1. Eligibility criteria 14
   2.2. Search strategy 15
   2.3. Evidence screening and selection 16
   2.4. Data extraction 17
   2.5. Data analysis 18
3. Results 19
   3.1. Overview of included sources 19
   3.2. Concepts and guidance 21
   3.3. Themes 24
4. Discussion 48
   4.1. Overview of findings 48
   4.2. Mapping to evidence-based principles for person-centred care and support 49
   4.3. Limitations and Strengths 56
   4.4. Gaps and future research 56
   4.5. Conclusion and next steps 57
5. References 59
6. Appendices 70

Tables and Figures

Table 1. Summary of 12 themes generated from the evidence reviewed 6

Table 2. Concept grid outlining the review question and the key concepts 15

Figure 1. Flow chart demonstrating the number of articles excluded at each stage of screening 17

Figure 2. Diagram summarising method used to synthesise data 19

Figure 3: Geographic distribution of sources 20

Figure 4. Word cloud representing the number of sources each concept was identified in 22

Figure 5. Top thirty concepts by frequency of occurrence (number of sources) 23

Table 3. Matrix showing themes cross-cutting with evidence-based HIQA principles for person-centred care and support 50

Acknowledgements

This review was led by staff in HIQA's Health Information and Standards Directorate.

Name	Role
Caoimhe O'Reilly	Project Lead
Louise Foley	Research Officer
Daráine Murphy	Senior Business Analyst
Deirdre Hyland	Programme Manager

The following individuals contributed to the management, technical writing or dissemination of this report: Rachel Flynn, Linda Weir, Fiona Geaney and Elena Vaughan.

The following individuals contributed to the initial research which informed the background to this report: Deirdre Connolly, Ailbhe O'Rourke, Marie Carrigan.

Conflicts of Interest

None reported.

Executive Summary

Introduction

Artificial Intelligence (AI) can be described as a machine-based system capable of operating autonomously and producing outputs like predictions, recommendations, or decisions based on input data.[1] Potential applications of AI in health and social care services are growing, such as in medical imaging and diagnostics, predictive analytics and early disease prevention, administration, drug discovery, and personalising health and social care.[2] Some examples of AI tools that could be used in health and social care services include:

- AI scribes that can prepare clinical notes in real-time during consultations
- AI tools that can enhance the quality of images captured during scans
- AI tools that can risk-stratify populations using large datasets
- AI tools that can monitor information about health and wellbeing and alert staff if a person is at risk of a health problem.

As the potential utility of AI tools in health and social care services increases, there needs to be measures in place, including strong governance and accountability, to ensure it is used in an ethical, safe and transparent way.[3,4] If AI is to contribute positively to the health and social care system, policies and guidelines must be in place to ensure it is used responsibly and safely.

The AI landscape is evolving rapidly, and an increasing amount of legislation, frameworks, position papers, and articles are being produced internationally to guide the use of AI.[5-10] In Ireland, the use of AI tools in health and social care services will be underpinned by the EU AI Act.[11] The AI Act addresses potential risks of AI to citizens' health, safety, and fundamental rights; it provides AI developers and those working with AI with clear requirements and obligations regarding AI tools, including AI-enabled medical devices.

Health and social care services will require further sector-specific guidance on the use of AI, and people using services will need to know what to expect if AI is used in their care. HIQA is currently developing national guidance that will promote awareness and build good practice among services and staff about the responsible and safe use of AI. To ensure the guidance is relevant, timely, and effective, a review of recent international evidence was conducted.

Aim and method

The aim of this evidence review was to identify and synthesise currently available concepts and guidance that facilitate the responsible and safe use of AI in health and social care services, in order to inform the development of *National Guidance for the Responsible and Safe Use of AI in Health and Social Care Services* in Ireland.

A systematic search of academic and grey literature was conducted in March 2025. Sources were screened against pre-defined eligibility criteria and relevant data were extracted from included sources. Concepts identified across the sources were synthesised to generate themes relating to the responsible and safe use of AI in health and social care.

Overview of key findings

A total of 55 evidence sources were identified across the academic and grey literature, representing perspectives from diverse stakeholders including professional associations, national professional or regulatory bodies, patient organisations, academic researchers, industry, and intergovernmental organisations. Over one-third of sources had a global focus, while the remainder were from across 11 geographical regions. A total of 12 themes were generated which are summarised below (table 1).

Table 1. Summary of 12 themes generated from the evidence reviewed

- **Transparency:**
  Considers the properties of transparent AI tools such as explainability and traceability, as well as the extent to which people are made aware that AI is being used and how data is processed. Guidance and recommendations include auditable AI systems and provision of information.
- **Inclusivity and non-discrimination:**
  Describes the need for AI tools to be accessible and to be used fairly and equitably and considers the various types of bias and risk of bias associated with AI in healthcare. Guidance and recommendations include user-friendly AI systems, use of representative data, and critical analysis of AI outputs.
- **Privacy:**
  Discusses the importance of upholding peoples' rights for their health and social care information to remain confidential and measures to ensure privacy. Guidance and recommendations include compliance with privacy laws and secure data storage.
- **Human agency and oversight:**
  Emphasises the need for human involvement, oversight, monitoring, and governance throughout the development and deployment of AI systems. Guidance and recommendations include regular auditing and human validation.
- **Responsibility:**
  Captures the importance of responsibility and accountability for AI systems in health and social care, clarity on roles and responsibilities throughout the development and use of AI systems, and compliance with legal, ethical and

regulatory requirements. Guidance and recommendations include clarifying individual and organisational responsibilities, measures to ensure adequate and accessible redress mechanisms, and human oversight and assurance when using AI systems in health and social care.

- **Upholding people's rights:**
  Examines the need for AI systems used to respect and protect internationally recognised human rights, including autonomy, and to reflect a person's personal preferences when using AI in the delivery of care. Guidance and recommendations include considering individual needs and personal preferences.
- **Safe care:**
  Considers the duty to ensure that AI is used safely and to the benefit of people using health and social care services, and to mitigate any mental or physical harm. Types of risks are also explored, including novel considerations for the safe delivery of care when using AI. Guidance and recommendations include validation of AI systems and adherence to regulatory standards.
- **Integration into care:**
  Discusses the incorporation of AI into existing health and social care systems in a way that adds value and improves services, and considers the importance of interoperability, clinical relevance and environmental sustainability. Guidance and recommendations include interoperable systems and integrating tools that lead to better outcomes for people who use services.
- **Education, training, development and information provision:**
  Explores the need for education and information for both staff and the public, including people using services, in order to facilitate informed decisions and to support confident and responsible use of AI in health and social care. Guidance and recommendations include continuous education and training for professionals and promoting AI literacy among people using services and providers.
- **Data quality:**
  Examines the need for consistent, standardised, accurate, complete and reliable data for AI systems used in health and social care services, and for data to align with privacy and cybersecurity requirements. Guidance and recommendations include monitoring of AI outputs and infrastructure to support quality data.
- **Technical robustness and security:**
  Considers the importance of AI systems being secure against unauthorised access, breaches, attacks, and malicious interference, as well as the need to

ensure AI systems used in health and social care can perform under varying conditions. Guidance and recommendations include secure data storage and contingency planning.

- **Human connection:**
  Explores the importance of maintaining person-centred, human, and caring interactions between the people who use services and health and social care providers when AI is used. Guidance and recommendations include integrating person-centred values and involving people who use services in decisions.

Conclusion and next steps

This evidence review identified and synthesised concepts and guidance relating to the responsible and safe use of AI in health and social care. Overall, the findings highlighted the range of distinct but interconnected factors that need to be considered to guide the responsible and safe use of AI in services. Twelve themes were generated, pointing to key considerations for actors across the AI lifecycle. These themes were mapped to established, evidence-based principles for person-centred care and support. These four principles underpin national standards and guidance developed by HIQA, and comprise accountability, a human rights-based approach, safety and wellbeing and responsiveness. The 12 themes generated were mapped to the HIQA principles in a manner consistent with previous approaches to standards and guidance development and informed principles-based guidance for the responsible and safe use of AI in health and social care services.

The review forms one part of an evidence-based and collaborative process to develop *National Guidance for the Responsible and Safe Use of AI in Health and Social Care* in Ireland. As part of this process, HIQA is also engaging with relevant stakeholders through a steering group, a co-production working group, a public scoping consultation, focus groups and interviews, and a public consultation on draft guidance. The findings from this evidence review, together with insights gathered from stakeholders in the Irish health and social care context, will inform the development of the National Guidance.

1. Introduction

1.1. What is Artificial Intelligence?

Artificial Intelligence (AI), in very simple terms, is a machine-based solution that learns from data. This learning is then used to create content, predict, or emulate human behaviour. AI is described by the Organisation for Economic Co-operation and Development (OECD) as a range of digital technologies that are capable of mimicking human intelligence and performing tasks at a scale that exceeds what humans are capable of.[9] AI is not a new phenomenon – the term Artificial Intelligence has been in existence since 1955 and AI systems have been in use for many years.[11] However, recent advances in AI technology, especially its accessibility to the general public, means that AI has become much more widely known and used in the last number of years. It is difficult to define AI as it is constantly evolving. The current legal definition from the EU AI Act describes an AI system as "a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments".[11] In Ireland, the Department of Public Expenditure, Infrastructure, Public Service Reform and Digitalisation defines AI as "a machine-based system capable of operating autonomously and producing outputs like predictions, recommendations, or decisions based on input data".[1]

1.2. Artificial Intelligence in health and social care

1.2.1 AI use cases

Health and social care services are facing increasing challenges including increased demand due to an ageing population, increased prevalence of chronic and complex conditions and a shortage of health and social care staff. AI is increasingly being utilised in health and social care services to address some of the challenges faced, for example, AI is being used for medical imaging and diagnostics; in predictive analytics and early disease prevention; for administrative duties through the use of chatbots, virtual assistants, and robotic process automation; to accelerate drug discovery; and to personalise health and social care.[2] Previous research has found that AI can enhance clinical decision-making, improve efficiency and optimise resource allocation;[2] has the potential for macro-level impact through its ability to examine large scale trends in data and forecast disease outbreaks or pandemics;[12] can enhance diagnostic accuracy and efficiency in relation to the analysis of medical images;[2] and has the potential to maximise personalised care and individualised solutions for people using services.[2] Furthermore, research suggests that AI has the potential to maximise patient-clinician time by reducing administrative burden on

specialised staff or by automating certain tasks – for example, remote monitoring systems that monitor vital signs.[2] Research also suggests that AI may expedite the drug development process potentially leading to new discoveries and vital cures.[13]

While it is evident that there are benefits relating to the use of AI in health and social care services, there needs to be measures in place, including strong governance and accountability, to ensure it is used in an ethical, safe and transparent way.[3,4] Although AI systems are advancing rapidly, the development and incorporation of supporting policies and regulations are still evolving, for example, data privacy and security policies tailored specifically to AI are emerging, but are not yet fully established. Therefore, additional safeguards need to be in place to ensure that the rights of people using services are protected.[3] AI systems, as with any technological device, can produce false or inaccurate outputs. Therefore, it is important that human oversight is maintained, particularly for high-risk AI systems.[3] AI systems that are trained on AI algorithms based on non-representative or biased datasets need to be identified and the risk minimised to prevent discrimination to already vulnerable patient groups.[3]

1.2.2 AI output types

There are eight main output types from AI systems – generative, labels, prediction, recommendation, optimisation, translation, interaction, and autonomy. A form of AI that most people will be familiar with is generative AI. Large language models such as ChatGPT, Claude, and Deep Seek are examples of generative AI. This form of AI can be used to answer questions, provide information, and generate images. For health and social care, generative AI has shown promise in producing patient summaries, discharge summaries, procedural coding, and referral letters, creating accessible plain language explanations for people who use services, synthesising population health data, or directly interacting with people who use services (through the use of chatbots).[14,15] Generative AI, for example, has been successfully used to help patients with glycaemia and weight loss.[16]

AI can also be used for labelling. For example, using data from X-ray images to label a patient as having, or not having, a given disease. Such AI systems can also be used in medical imaging, for example, to interpret chest radiographs, which has been found to enhance diagnostic accuracy and reduce radiologist work load.[17] AI has also shown promise for detecting malignant breast lesions in X-rays and identifying diabetic retinopathy in fundus imaging and lung nodules in computed tomography scans.[15]

AI can be used to predict medical outputs. For example, clinical data (that is, blood tests, demographics, vital signs) can be input to machine learning models to identify disease risk or presence and forecast disease progression. Predictive AI systems can be used to support clinicians to predict the risk of mental health crisis among people

receiving psychiatric care,[18] and to predict postpartum depression, and can be developed to transform pre-eclampsia diagnosis and care.[19] Predictive models were used during the COVID-19 pandemic to understand and predict patterns of virus spreading, allowing for preventive measures to be put in place to protect macro-level health.[20]

AI can be used for recommendation. In health and social care services, AI systems can be used to provide recommendations for personalised treatment based on data such as a patient's medical history, symptoms, lab results, and evidence-based guidelines.[21]

AI can be used for optimisation. In health and social care, AI can be used to optimise emergency service routes, to target vaccination outreach to high-risk areas with low uptake, and to optimise patient admissions, transfers and discharges. A prognostic and treatment algorithm for breast cancer has been developed which optimises treatment options for female breast cancer patients.[22]

AI can be used for translation. For example, such AI systems can be used to support interpretation to allow people using services to communicate with health and social care workers in a language they do not speak.[23]

AI can be used for interaction. AI systems can respond to voice commands in an emergency to contact services, for example, if an older person has a fall. AI systems can use facial analysis to support the assessment of pain among people who cannot verbally communicate their pain, for example research has been conducted with people living with dementia in this context.[24,25]

AI can be used for control – also known as autonomous or real-time AI. Examples of autonomous AI that could be used in health and social care services include AI-controlled surgical robot assistants,[26] AI-controlled robots to assist those with additional physical needs,[27] and wearable AI systems to detect fall risk.[28]

The potential of AI to improve the health and social care system has been recognised at a health policy and service delivery level in Ireland.[29,30] However, there needs to be proactive measures in place to ensure that AI is used in a responsible and safe way. On one hand, research has shown that AI has the potential to enhance health and social care. On the other hand, concerns related to the responsible and safe use of AI need to be considered.[3,4] If AI is to improve the health and social care system, policies and guidance must be in place to ensure the responsible and safe use of AI. To ensure safer, better care for people using health and social care services, it is vital to have national guidance to facilitate the responsible and safe use of AI in health and social care services in Ireland.

1.3. Previous research, strategies, and policy documents

Internationally, many frameworks, legislations, position papers, and articles have been published to guide the responsible and safe use of AI.[5-10] For example, a plethora of reviews across various domains have been conducted internationally to assess AI governance, ethical principles, barriers to adopting ethical AI principles, and best practices.[31-34] In healthcare, systematic and scoping reviews have explored applications of AI in healthcare and how these impact safety, transparency, and ethics;[35] healthcare professionals' experiences of using AI tools to inform clinical decision making;[36] the processes and challenges, and barriers to incorporating trustworthy AI in healthcare;[37,38] what responsible AI in digital health is;[39] the benefits and risks of AI in healthcare;[40] ethics relating to AI and healthcare;[41,42] and the unique biases introduced through AI use in healthcare.[43-45]

While previous reviews, frameworks, guidance, and concepts provide valuable insights into responsible AI implementation, they were not conducted specifically to inform the development of a national guidance for the responsible and safe use of AI specific to Irish health and social care services. Every health and social care system operates within a unique legal, ethical, and organisational context. Considering the Irish context, Digital for Care's health framework for Ireland[30] highlights AI's potential to reduce pressure on acute and community services, facilitating the timely and efficient delivery of care. Ireland's national AI strategy,[46] the *Progress Report on the National AI Strategy*,[47] and the *Guidelines for the Responsible use of AI in the Public Service*[1] emphasise the need for ethical and trustworthy AI adoption. However, these strategies primarily draw from high-level international frameworks such as the EU AI high-level expert group *Ethical Guidelines for Trustworthy AI* and the OECD *Recommendation of the Council on AI*. While these frameworks provide useful guiding principles, they are not sector-specific and were published before the enactment of the EU AI Act and OECD AI guiding principles.[9,11] Furthermore, the scope of this review was more expansive than those previously conducted due to the inclusion of additional grey literature sources from regulatory, professional and international bodies such as WHO (World Health Organisation), OECD and those specifically relevant to the Irish context.

The EU AI Act[11] is a regulatory and legal framework for AI in the EU and addresses potential risks of AI to citizens' health, safety, and fundamental rights. It provides developers and those who are working with AI tools with clear requirements and obligations regarding specific uses of AI and outlines four risk levels of AI use ranging from low to unacceptable risk. Many applications of AI systems in health and social care will fall under the "high risk" category of AI and thus will require rigorous oversight and governance.[11,48] The EU AI Act outlines seven principles to ensure the reliable and correct use of AI systems including human agency and oversight; technical robustness and safety; privacy and data governance; transparency;

diversity, non-discrimination and fairness; societal and environmental wellbeing; and accountability. The OECD AI principles document[9] is the first intergovernmental standard on AI to promote innovative, trustworthy AI that respects human rights and democratic values. The OECD standard is composed of five values-based principles – sustainable development; human-centred values and fairness; transparency; safety; and accountability. In the Irish context, AI guidance will be underpinned by the EU AI Act and the OECD AI principles.

To date, no evidence-based review has systematically examined responsible AI use to inform Ireland's health and social care sector, nor is there a sector-specific national guidance for the use of AI in Ireland's health and social care services. Given the evolving AI landscape, there is a need to synthesise recent literature, concepts, and guidance relevant to responsible and safe AI usage in Irish health and social care services to inform the development of a national guidance that aligns with the needs and challenges of the Irish health and social care system.

1.4. The Irish context

In Ireland, the Department of Health, the Health Service Executive (HSE) and the Health Information and Quality Authority (HIQA), are working together to ensure that the health and social care sector is ready for forthcoming policy and legislative requirements for AI. A system-wide approach to supporting the responsible and safe use of AI in health and social care is needed and the Department of Health, HSE and HIQA are developing separate but interlinked programmes of work in line with their respective organisational remits. The Department of Health oversees Ireland's health and social care system and is responsible for policy development and oversight, funding and resource allocation, developing legislation and regulation, leading responses to public health, and collaborating with global bodies on health initiatives. The HSE delivers Ireland's public health and social care service in hospitals and communities nationally, both directly and through partnerships. The HSE is responsible for providing safe, high-quality, and accessible healthcare.

The Department of Health and the HSE are developing an AI in Health Strategy to promote and support innovation and digital transformation in health, as part of a commitment in the *Programme for Government 2025*. This strategy will promote the use of AI in healthcare and the responsible and safe use in a number of areas such as clinical patient care, operations and administration, research and innovation, patient engagement and experience and public health. The HSE is also working on a corresponding strategic roadmap and implementation framework to ensure that AI systems are used responsibly, legally, safely and effectively. The implementation framework will set parameters for AI use ensuring organisational accountability for decisions and risk while adhering to ethical, legal and regulatory frameworks. The implementation framework will also address the technical quality, data management

and cybersecurity aspects of AI systems as well as risk management, prioritisation processes and partnership approaches.

HIQA is an independent statutory body established to promote safety and quality in the provision of health and social care services for the benefit of the health and welfare of the public. The Department of Health commissioned HIQA to develop national guidance to promote and drive the responsible and safe use of AI in health and social care. HIQA is working collaboratively with policy-makers, people using and delivering health and social care services and members of the public to develop an evidence-based national guidance to promote the responsible and safe use of AI in health and social care services in Ireland.

1.5. Purpose of this review

The aim of this evidence review is to identify and synthesise currently available guidance and concepts for facilitating the responsible and safe use of AI in health and social care services. The evidence gathered will be used to inform the national guidance for the responsible and safe use of AI in health and social care services in Ireland. For the purpose of this review, guidance can be defined as "advice or information aimed at resolving a problem or difficulty, especially as given by someone in authority."[49] A concept is an idea or principle,[50] operationalised for this review as "a statement of a duty or a responsibility in the context of the development, deployment and continuing assessment of AI technologies for health."[51]

The evidence review summarises and synthesises relevant concepts and guidance and will be used to inform the development of national guidance for the responsible and safe use of AI in health and social care services in the Irish context, as requested by the Department of Health. The main purpose of this guidance is to promote awareness and build good practice among services and staff about the responsible and safe use of AI in their services. The guidance will also be of use to people using services by educating and empowering them on what their expectations should be in respect of how AI can be used safely and responsibly while engaging with health and social care services. The evidence review aims to answer the following research question: What concepts and guidance are available to facilitate the responsible and safe use of AI in health and social care services?

2. Methods

2.1. Eligibility criteria

Table 2 outlines the review question and the key concepts using a concept grid. Guided by the review question, evidence sources were considered relevant for inclusion in the review if their content addressed concepts 1 to 4 outlined in table 2. Concept 1 specifies principles, concepts, guidance, or frameworks. A principle was

defined as "a statement of a duty or a responsibility in the context of the development, deployment and continuing assessment of AI technologies for health,"[51] guidance as "advice or information aimed at resolving a problem or difficulty, especially as given by someone in authority,"[49] and framework as "a structured approach or system that provides guidance, objectives, and methodologies to implement policies, programmes, or projects effectively".[52] A framework involves a structured set of ideas, concepts, principles, or rules used to organise and guide actions or decisions related to a specific area of health or public health. For a source to be included in the review, the criterion listed in concept 1 had to be related to the responsible, safe, trustworthy, and or ethical use (concept 2) of AI (concept 3) in contemporary health and or social care settings (concept 4).

Grey literature sources published during or after 2019 and peer-reviewed literature published during or after 2022 were considered for inclusion, to align with publication of the *EU Ethics Guidelines for Trustworthy AI* in 2019.[53] The full list of inclusion and exclusion criteria is outlined in appendix 1.

Table 2. Concept grid outlining the review question and the key concepts

Review question	Concept 1: Guiding concepts	Concept 2: AI	Concept 3: Approach to AI	Concept 4: Setting
What guidance and concepts are available to facilitate the responsible and safe use of AI in health and social care services?	- Principles - Concepts - Guidance - Frameworks	- Artificial intelligence	- Responsible - Safe - Trustworthy - Ethical	- Health care services - Social care services

2.2. Search strategy

The search strategy was developed by a junior librarian and peer reviewed by a senior librarian in HIQA using the Peer Review of Electronic Search Strategies (PRESS) checklist.[54] The search strategy was piloted by one reviewer. Firstly, two

key articles from the peer-reviewed literature and three key grey literature sources identified as relevant from a preliminary literature review were searched for within the results, to ensure the database results included those key sources. Secondly, a rapid screening of the first 100 citations from the academic database search was conducted to test the precision of the search strategy (as recommended by Pawliuk et al.[55]). The full search strategy is outlined in appendix 2.

2.2.1 Academic database search

The academic database search was conducted in March 2025 on MEDLINE yielding 1,593 full texts. The urgent requirement for this evidence review to inform the national guidance for responsible and safe use of AI in health and social care in Ireland, alongside the time and resources available to conduct the review constrained the search to a single academic database, and necessitated the use of the cut-off date applied. In particular, MEDLINE was chosen given its extensive coverage of literature relevant to the health and social care context. All articles identified via the search strategy were exported to EndNote (Clarivate, Philadelphia, PA, USA, London, UK) and duplicates were removed (n=10). The remaining sources were then imported to Covidence for screening.

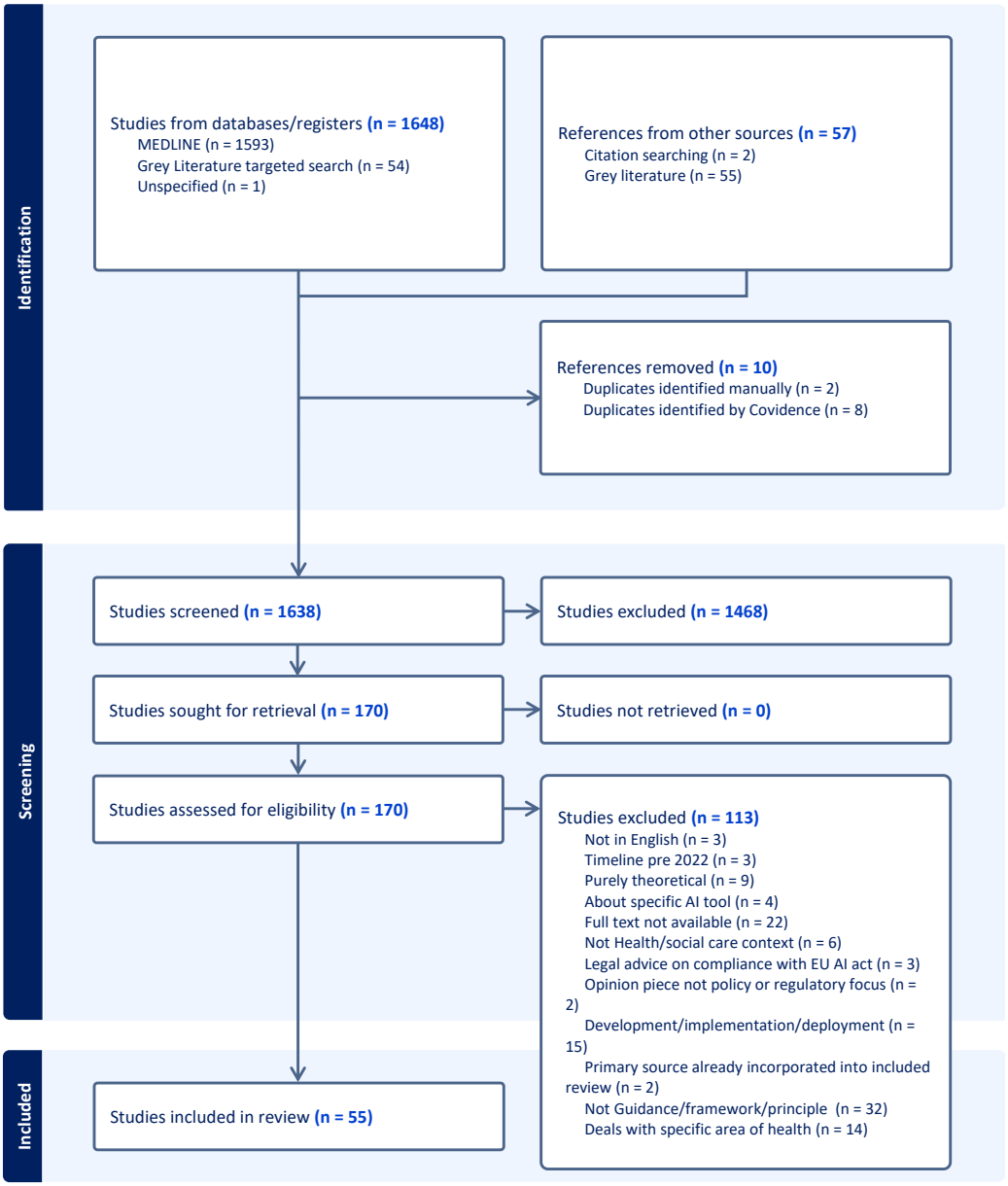
2.2.2 Grey literature search

The grey literature systematic search was conducted in March 2025 using websites of organisations and bodies previously deemed as relevant based on a preliminary review of the literature conducted in January 2025 (see appendix 3 for list of websites). Two reviewers searched half of the websites each (split based on an alphabetically ordered list). A standardised approach to searching was followed: the search box on each website was located and used to search a list of keywords (appendix 2). The process and results for each website were logged in a search log template. This yielded 55 sources. The grey literature sources identified were then imported to Covidence for screening.

2.3. Evidence screening and selection

The identified sources were screened using the eligibility criteria outlined in appendix 1. Title and abstract screening were conducted blindly by two reviewers using Covidence. Discrepancies were discussed and resolved by the two primary reviewers. Sources considered potentially relevant for inclusion progressed to full-text screening (n=170). Studies identified through citation chaining progressed directly to inclusion. Two reviewers each screened 50% of the full texts. To ensure consensus, approximately 30% were cross-checked by both reviewers. A total of 113 articles were excluded with reasons recorded. The study selection process is documented via the flow chart (figure 1).[56]

Figure 1. Flow chart demonstrating the number of articles excluded at each stage of screening



2.4. Data extraction

Data was extracted in two stages using Covidence. Two reviewers each conducted data extraction for approximately 50% of the included sources. The data extraction process was piloted using 25 randomly selected sources to assess consistency between reviewers. During final data extraction, each source was screened by one reviewer and approximately 15% of sources were cross checked by a third reviewer to ensure accuracy and consistency. The first step of data extraction involved

extracting data on source characteristics including, where available, the source title, country or location, year, aim, design or methods, stakeholder perspective, target audience of the source, official outputs from the evidence source (such as guidance, framework, principles, recommendations), type of organisation, and underpinning ethical, policy, or conceptual frameworks.

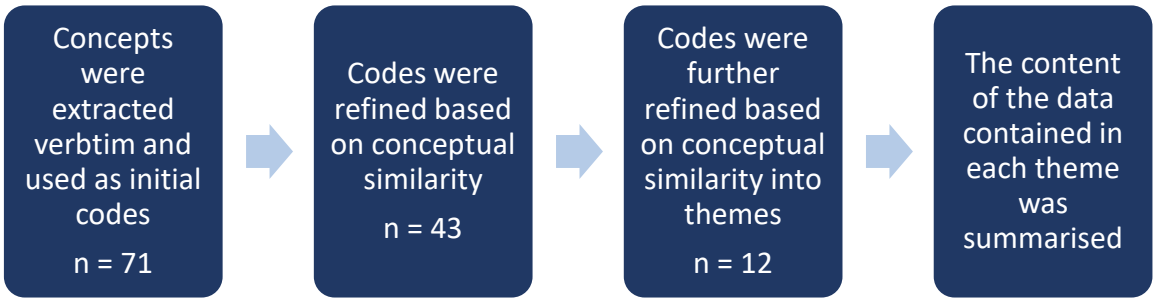
The second step involved using an iterative process to extract data that captured concepts for the responsible and safe use of AI.[56] Data were extracted if they aligned with the definition of a principle or guidance outlined in section 2.1. If a source did not explicitly label such segments as a principle, guideline, or recommendation, an element of interpretation by the reviewer was introduced; that is, if the reviewer deemed a segment of text as aligning with the aforementioned definition, the segment was extracted as a concept. For example, Alelyani referred to "key factors that emerged as critical for assessing the trustworthiness of autonomous systems in healthcare"[57] and those key factors were extracted as concepts.

Extraction of the concepts involved iteratively adding rows to the data extraction table in Covidence each time a new concept was identified. A row was added with the name of the concept, and verbatim text from the source defining the concept was extracted. For each new concept, a second row was added, and verbatim text from the source outlining how to uphold the concept in practice was extracted. After piloting the data extraction procedure, no major discrepancies were found between the two reviewers. A total of 71 concepts were extracted. The extracted data was then imported to NVivo and the 71 concepts were used as initial codes for further qualitative analysis.

2.5. Data analysis

A summary of the method used to extract, synthesise, and analyse the data from included sources can be found in figure 2. Two team members reviewed the data associated with each of the 71 initial concepts and used these as preliminary codes. The team members discussed potential areas of overlap and commonality between the preliminary codes. The 71 initial codes were then independently grouped and refined based on their content by the two coders. The refinement was discussed in two workshops with four team members and consensus was reached to refine the 71 codes into 43 refined codes. Refinement was repeated by two independent coders, discussion workshops were held, and consensus was reached to refine the 43 codes to 12 refined themes. At this point, the team agreed that saturation had been reached and no further refinement occurred.

Figure 2. Diagram summarising method used to synthesise data



3. Results

The academic and grey literature searches together yielded 1,648 results. Of these, 10 duplicates were removed. During title and abstract screening, 1,468 sources were excluded. Full-text screening was conducted for 170 sources and 115 were excluded (see figure 1 for exclusion rationale). In total, 55 sources were deemed to meet the inclusion criteria and were included in the review (28 academic literature sources and 27 grey literature sources).

3.1. Overview of included sources

In line with the inclusion criteria, sources from the academic literature search were published between January 2022 and March 2025 and sources from the grey literature search were published between January 2019 and March 2025 (see appendix 4 for characteristics of sources included in the review). The majority of sources were published in 2024. Eleven specified geographical regions were identified across the 55 sources, and 19 sources specified having a global focus (figure 3). The majority of evidence reviewed was targeted at multiple stakeholder groups including people who work in health and social care services, AI developers, researchers and people who use health and social care services.

Figure 3: Geographic distribution of sources

Note: Larger circle represents higher quantity of sources

3.1.1 Overview of academic sources

Among the 28 academic literature sources, 15 were described as a review.[21,58-72] Of these, three were further described as a scoping review,[66,69,71] six as a narrative or literature review,[60,62,63,67,68,70] and one reported a discourse analysis of published documents.[58] While the majority of reviews did not involve consultation with experts or stakeholders, two incorporated consensus processes with groups representing various backgrounds and expertise.[70,71] For example, Reddy et al.[70] involved an international panel with expertise in medicine, data science, healthcare policy, biomedical research and healthcare commissioning. Three sources conducted primary qualitative research using interviews, including with professionals working in healthcare,[73] medical doctors,[74] and experts in AI system technologies in healthcare[57] (range of sample = 15-25 participants). Two further sources used consensus methods involving multiple stakeholders,[7,15] for example Lekadir et al.[7] engaged a consortium of 117 participants from clinical, technical, AI, ethics, social science, legal, industry, regulatory, and patient advocate perspectives. Other sources included a document content analysis,[75] two frameworks,[17,76] a summary and grouping paper,[77] a case example approach,[78] and two discussion papers.[79,80]

3.1.2 Overview of grey literature sources

The 27 grey literature sources included nine position papers,[8,81-88] four participatory reports,[6,89-91] one white paper,[92] one consultation submission,[93] and one webpage.[94] The remainder were general reports.[9,51,95-103] Six were produced by professional associations,[81,86-88,93,95] five by national professional or regulatory bodies,[82,83,94,97,101] four by patient organisations,[6,89-91] two by an industry trade association,[8,84] and four by intergovernmental organisations.[9,51,98,99] A further six were produced by other types of organisations.[85,92,96,100,102,103] Several grey literature sources reported the approach to gathering information and or involving experts and stakeholders. Three reported conducting a review of evidence.[93,96,100] Other approaches included a citizens' jury with 24 members of the public,[91] clinical simulations combined with interviews and surveys,[92] qualitative interviews with people who use services, their representatives, technologists, researchers, and AI policy experts,[90] and regulatory sandboxing involving technology suppliers and their clinical partners.[97] Two sources combined a survey with consultation among diverse stakeholders;[6,95] for example the Australian Alliance for Artificial Intelligence in Healthcare surveyed 152 stakeholders and consulted with a working group and industry advisory group.[95] Five other sources reported involvement of stakeholders via consultation,[86,89,99] stakeholder discussion,[85] and expert/working groups.[51,89,99]

3.2. Concepts and guidance

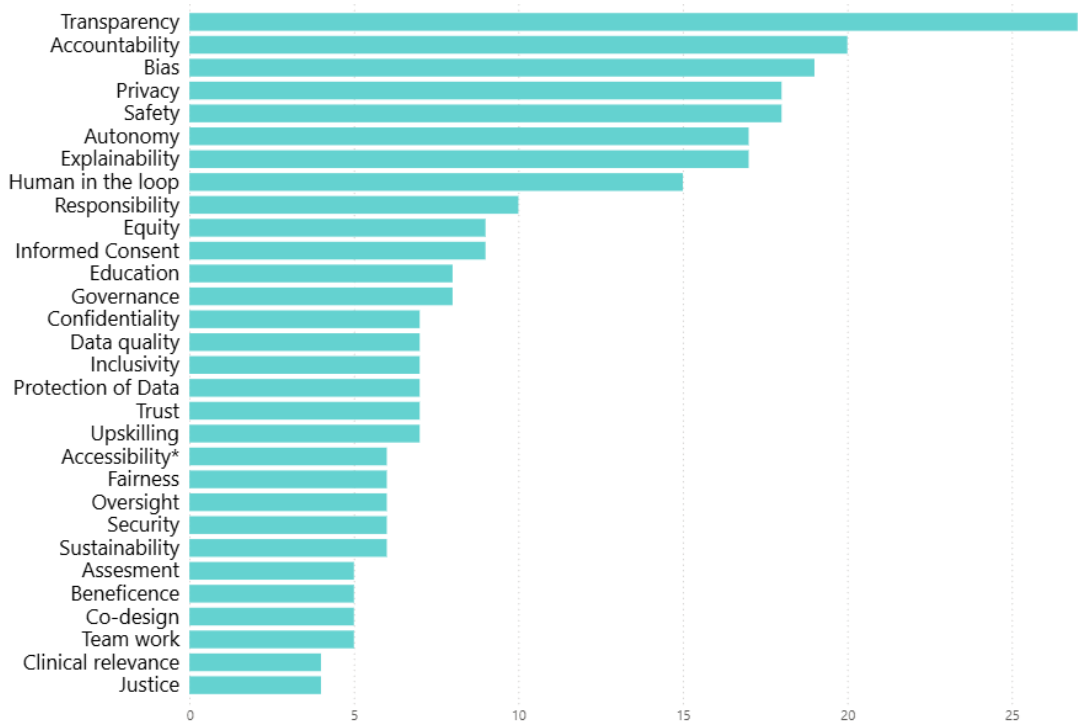
In total, 71 concepts were identified across the 55 papers. Figure 4 visually represents the frequency of occurrence of each concept in the form of a word cloud. The concept mentioned most often was Transparency (27 articles) followed by Accountability (20 articles), Bias (19 articles), Safety (18 articles), Privacy (18 articles), Autonomy (17 articles), Explainability (17 articles), and Human in the Loop (15 articles). All other concepts were mentioned in 10 or less sources as outlined in figure 5. The 71 concepts were synthesised to generate 12 refined themes (see appendix 5 for details on code refinement leading to generation of 12 themes):

- Transparency, which includes the subtheme Trust
- Inclusivity and non-discrimination
- Privacy
- Human agency and oversight
- Responsibility
- Upholding people's rights
- Safe care
- Integration into care, which includes the subtheme Eco-responsibility and sustainability
- Education, training, development and information provision which includes the subtheme Public education and information provision
- Data quality
- Technical robustness and security
- Human connection.

Information about the types of sources and evidence that refer to each theme can be found in appendix 6. The conceptualisation of each theme, as well as related guidance and recommendations, are summarised in section 3.3.

Figure 4. Word cloud representing the number of sources each concept was identified in

Figure 5. Top thirty concepts by frequency of occurrence (number of sources)



3.3. Themes

3.3.1 Theme one: Transparency

Thirty-seven sources contained concepts captured by the theme "transparency".[6-9,15,17,21,51,57,61-68,70,71,73,74,76,77,79-82,86,89,91,94,96-98,100-102] Twenty-one were academic sources and 16 were grey literature sources. (See appendix 6 for detail on the types of evidence supporting each theme).

Conceptualisation of transparency

The theme transparency encapsulates the properties of transparent AI tools such as explainability and traceability, as well as the extent to which people are made aware that AI is being used and how data is processed. This theme also captures some of the advantages of transparency, including trust, accountability and oversight. Trust is discussed below as a sub-theme of transparency.

Explainability was discussed by several sources,[7,9,15,17,51,57,61,62,64,65,68,70,73,74,80,81,89,100-102] including its role in understanding how AI systems produce decisions,[57] the "logic" behind AI outputs,[7] and the "reasoning processes" of AI technologies.[100] Jha and colleagues cited explainability as a key ethical issue in the development of medical AI systems,[17] while the Irish College of General Practitioners stated that both providers and users of services require some level of understanding of how AI produces recommendations.[101] Eight sources further discussed the "black-box" nature of certain AI systems,[7,17,64,67,70,73,80,102] with algorithmic explainability proposed by the National Health Service (NHS) as a measure to mitigate the "black-box" issue, promoting understanding of how AI models function.[102] In their qualitative study with healthcare professionals, Elgin and Elgin reported that the need for explainability may vary according to the nature of contexts and decisions, with strong emphasis on explainability for decisions directly related to care.[73] The importance of balancing accuracy and explainability was highlighted in two sources,[64,68] with Mennella et al. describing this as a key consideration.[68] Lekadir et al. discussed traceability, referring to the need for "detailed and continuous" information about AI tools across their lifecycle,[7] while Sousa-Pinto et al. highlighted the requirement for credibility of AI systems.[71]

As well as transparency in AI algorithms, 10 sources referred to transparency regarding whether an AI system is being used.[6,8,17,66,71,81,82,94,96,98] In their scoping review, Maccaro et al. noted the importance of users' awareness of when they are interacting with AI.[66] The Australian Commission on Safety and Quality in Health Care (ACSQHC) cited the need for transparency about using AI as part of care,[96] and the European Patients' Forum referred to the need to inform users if interactions are with a "non-human agent".[6] Referencing the clinical context, the Australian Medical Association stated the importance of transparency if an AI system is used to

determine a diagnosis or treatment recommendation,[81] while the College of Physicians and Surgeons of British Columbia emphasised the need for transparency regarding the degree to which a provider uses an AI system in clinical decision-making.[82] Transparency around the processing of data was mentioned in three sources.[63,79,96] Seroussi and Zablit noted the distinction between primary and secondary uses of data and underlined the need for transparency about processing and access,[79] while Harishbhai Tilala et al. emphasised the importance of transparency regarding data use, risks, and privacy.[63] Other areas requiring transparency mentioned within the reviewed sources included governance,[96] potential risks,[6] limitations,[6,51,82,96,102] decision-making,[51] evaluation,[51,96] training data, bias and performance.[96]

The advantages of promoting transparency were cited in multiple sources. Regarding model explainability, both Lekadir et al. and Alelyani noted that transparency supports interpretation and understanding of AI tools and outputs.[7,57] Seven sources discussed transparency from an accountability perspective,[6,7,17,61-63,96] and as many referred to the importance of transparency for promoting trust.[6,17,57,62,63,80,89] Five sources discussed the relevance of interpretability and transparency to oversight and monitoring,[7,17,51,57,98] while the OECD pointed to the importance of understanding AI systems to ensure they can be "effectively challenged".[9] Moreover, the WHO noted that transparency requires adequate information to be available prior to the design and roll-out of AI systems to facilitate consultation and debate.[51] From a quality and safety perspective, the WHO pointed to the value of transparency for enhancing quality, ensuring patient safety and promoting public health safety;[51] the Australian Commission on Safety and Quality in Health Care highlighted that transparency facilitates the regulation of safety;[96] Upadhyay et al. stated that transparency in AI models is necessary for fairness and prevention of biases;[80] and Jeyaraman and colleagues underlined that low explainability may result in adverse effects in the context of clinical decision support systems.[64] Four sources discussed the relevance of transparency when communicating with people using services and or enabling informed decisions.[21,63,73,96] Other advantages of transparency included understanding the abilities and limitations of a tool,[7] acceptance of AI tools in practice,[63] and issues related to certification, approval and liability.[64]

Guidance and recommendations

Guidance and recommendations to promote transparency were discussed in a number of sources. The evidence suggested that transparency can be promoted through measures including "interpretability and explainability, communication, auditability, traceability, provision of information, record-keeping, data governance, and documentation".[67] Other sources referred to audit,[61,74] including developing AI systems in a way that is auditable and in a format that can be understood by

clinicians.[61] The role of training and communication for end users to promote their understanding of AI system outputs and to support their communication with those impacted by AI outputs was also noted.[8] Regarding the provision of information, sources mentioned the need to provide adequate information,[68] for explanations to be suitably adapted for the recipient,[51,66,68] and for information to be as accessible and comprehendible as possible.[63,71] On the topic of governance, recommendations included building on established expertise and governance processes to oversee transparency in the use of AI.[96] Other recommendations to facilitate transparency included making legal and best-practice guidelines available,[76] and documenting all AI tools, data sources and methodologies that have been used.[71]

Conceptualisation of sub-theme: Trust

The sub-theme trust considers the nature of trust in the context of AI use, the relationship between trust and reliance, and the role of humans in providing care.

Three sources discussed the nature of trust generally, and or in the context of AI use specifically.[21,66,80] Savulescu et al. characterised trust as a relationship between humans that involves relying on another person perceived to have requisite knowledge, skills and moral attributes.[21] The meaning of trust in the context of AI systems was further explored, questioning the extent to which humans can plausibly trust tools, including AI systems; the view that such a relationship may more suitably be framed from the perspective of reliability than trust was noted.[21] Similarly, Upadhyay et al. asserted that the extent to which decisions based on AI can be trusted depends on the apparent reliability and validity of the AI system.[80] The idea that trust in AI metaphorically reflects trust in the designers and or users of an AI tool who are responsible and accountable was also presented.[21] In their discussion of health equity, Tierney et al. highlighted that use of AI technologies in healthcare can potentially exacerbate low trust among people who are marginalised if it is not appropriately introduced.[77] Indeed, the trustworthiness of AI systems is described by Upadhyay et al. as critical for them to be accepted and used effectively.[80]

The relationship between trust and (over)reliance in the context of AI systems was explored in two sources.[100,101] The Irish College of General Practitioners pointed to a tension between the potential negative impacts of overly depending on AI – citing errors, skills erosion and issues with appraising AI systems – and the constraints on potential improvements in practice if trust is low.[101] Highlighting further the tension between trust and reliability, a report from the European Commission Joint Research Centre underscored the need to examine how contradictory conclusions reached by an AI system and a human can be managed.[100] Two further sources discussed the impacts of AI on the role of humans in healthcare,[21,67] with Savulescu and colleagues reflecting on concerns that AI could result in obsolescence of humans, or deskilling due to overreliance.[21] Similarly, Marques et al. referred to potential

impacts on employment.[67] However, returning to the nature of trust, Savulescu et al. emphasised that wherever trust and accountability are required in the provision of care, humans will also be required, suggesting that these features of care cannot be replaced by AI systems.[21]

Guidance and recommendations

Guidance and recommendations to uphold trust were identifiable in a number of sources. Regarding features of trustworthy AI use, cited mechanisms included informing individuals when an AI system is being used and providing explainable outcomes where possible.[79] Regarding the features of a caring approach, recommendations included conserving the emotional element of care, with the need for empathetic and compassionate care emphasised.[91] Other noted characteristics of a caring approach included addressing emotional needs, informing service users of how machine learning is incorporated in their care, explaining diagnostic results, describing how data is processed, and attending to concerns.[97]

3.3.2 Theme two: Inclusivity and non-discrimination

Thirty-five sources contained concepts captured by the theme "inclusivity and non-discrimination". [6,7,9,17,21,51,57,59-68,70,74,75,77,79,81,82,86,89-92,96,98-102] Nineteen were academic sources and 16 were grey literature sources (See appendix 6 for detail on the types of evidence supporting each theme).

Conceptualisation of inclusivity and non-discrimination

The theme inclusivity and non-discrimination encapsulates the need for AI tools to be accessible and to be used fairly and equitably and also considers the various types of bias and risk of bias associated with AI.

Twenty-four sources addressed the topic of bias in AI systems.[6,7,17,51,57,60-64,66-68,70,75,82,86,89-91,96,98,100,101] The WHO noted that bias is "a threat to inclusiveness and equity, as it can result in a departure, often arbitrary, from equal treatment".[51] A number of sources discussed the types of bias relevant to AI.[7,17,60-62] As outlined by Lekadir et al., bias in healthcare AI can be due to differences in the attributes of individuals (such as sex, gender, age, ethnicity, socioeconomic status, medical conditions) or the data (such as acquisition site, machines, operators, annotators).[7] Corfmat et al. and Elendu et al. discussed two types of bias caused by AI.[60,62] The first replicates societal and historical biases present in machine learning data, which can disproportionately impact particular groups. The second relates to incomplete or under-representative data, especially that which over- or under-represents a subgroup such as a minority group, a vulnerable group, or disease subtype. Drabiak further identified a third type of bias in their review, contextual bias, where an AI system developed in one medical setting does not apply to another, for example, an

AI system developed in a high-resource centre being applied in a low-resource setting.[61]

Five sources discussed how under-representative data, inappropriately representative data, or flawed algorithm development can lead to inaccurate AI outcomes, which in turn can lead to potentially harmful decisions thereby perpetuating healthcare disparities.[6,57,63,64,70] Jha et al. noted the potential impacts of such bias, from providing suboptimal care to marginalised communities, to eroding trust in healthcare institutions, potentially deterring these populations from seeking necessary medical help.[17] However, the Council of Europe suggested that detecting bias in AI systems may not be straightforward as biased decision-making rules can be hidden in black-box models.[98]

Fourteen sources highlighted the importance of AI systems being fair and equitable.[7,21,51,57,59,63,66-68,74,81,91,96,99] Alelyani discussed the need to ensure that AI systems' decisions, actions and outcomes are unbiased, equitable, and do not disproportionately favour or discriminate against individuals or groups, for example based on race or socioeconomic status.[57] However, Lekadir et al. noted that while AI systems should perform the same across all individuals, perfect fairness might not be possible to achieve in practice.[7] Six sources advocated that the benefits and costs of AI systems be fairly distributed,[51,59,63,66,68,81] while four others noted that AI should be used to resolve issues concerning equitable access to healthcare and should not be limited to private services.[67,74,91,99] Marques et al. provided an example of how AI technologies could broaden healthcare access in rural areas through remote diagnoses, treatment recommendations, and health monitoring, mitigating challenges such as staff shortages. They also highlighted how AI technologies could worsen health inequalities, for example, if people with lower incomes are unable to afford advanced AI treatments.[67] Finally, Mennella et al. and the WHO noted that AI technologies should not solely align with the needs and usage patterns of high-income settings, but must be adaptable across contexts.[51,68]

Ten sources explored the issue of accessibility of AI systems, ensuring they are inclusive and that they are developed based on appropriate consultation with people who use services.[6,9,17,51,57,62,75,77,79,81] Alelyani noted that AI systems must be user-friendly and accepted by people providing and using services, as user experience can influence trust in the system and likelihood of adoption.[57] The European Patients' Forum noted that accessibility and inclusive design which accounts for diverse needs, including people with physical and mental disabilities, must be fundamental principles for the development and use of AI.[6]

Guidance and recommendations

Guidance and recommendations to promote inclusivity and non-discrimination were noted in several sources.[6,7,17,51,57,67,68,75,79,82,86,91,96] Cited measures included

developing AI technologies that accommodate varying levels of access to digital and technical infrastructure[68] and providing guidelines for monitoring and auditing AI outcomes.[96] From an accessibility perspective, recommendations encompassed user-friendly systems, including those meeting the needs of people with disabilities, providing guidance for users, and maintaining access to human support as needed.[79] Recommendations also included engagement with diverse stakeholders.[6,75] In relation to bias, noted measures included development of transparent and inclusive data development processes before training machine learning algorithms,[17,67] use of representative data,[57,67] and regular auditing and monitoring.[17,67] Other guidance included transparency over the process, data and outcomes used during development, use, and monitoring,[86] regular updates to capture new data,[91] diverse hiring practices,[51] and risk assessment frameworks to address risk of bias, discrimination and unfairness.[96] Guidance for clinicians included awareness of bias and critically analysing AI outputs through an equity, diversity, and inclusion lens.[82]

3.3.3 Theme three: Privacy

Twenty-nine sources contained concepts captured by the theme "privacy".[6,17,21,51,57,60,62-68,70,73,76,81,82,86,91,94-96,98-103] Fourteen were academic sources and 15 were grey literature sources (See appendix 6 for detail on the types of evidence supporting each theme).

Conceptualisation of privacy

The theme of privacy in the literature refers to ensuring that patients' right for their health and social care information to remain confidential is upheld.

Eleven sources suggested that AI systems must protect patient data, safeguard personal information, maintain confidentiality, and inform patients that their data is being collected.[6,57,62,63,73,81,82,86,91,99,102] with three sources stressing that privacy is particularly important in health and social care given the sensitivity of health data.[57,60,62] In their academic papers, Jha et al. and Harishbhai Tilala et al. highlighted that this is a fundamental requirement in healthcare systems.[17,63] The Council of Europe highlighted that the right to privacy and entitlement to know any information collected about one's health are outlined in article 10 of the Oviedo Convention – a legally binding international treaty by the Council of Europe.[99]

Five sources highlighted the tension between maintaining data privacy and gathering data comprehensive enough to be informative.[60,73,95,96,99] Corfmat et al. argued that the more data is anonymised, the higher the risk that its ability to provide important insights is reduced.[60] This is echoed by participants in Elgin and Elgin's study who described difficulties balancing gathering comprehensive data and maintaining data privacy.[73] In a similar vein, the Australian Alliance for Artificial Intelligence in Healthcare highlighted the importance of data collection and sharing for research

whilst acknowledging that healthcare is a high-risk sector in need of strict data security and privacy requirements.[95] The ACSQHC also recognised the importance of data collection to maximise the benefits of AI but highlighted that this must be done in a way that ensures sensitive health data is protected.[96] The Council of Europe also highlighted this tension, explaining that, while data confidentiality is paramount, on the other hand, openness may also help to identify potential biases in data in turn helping to mitigate discrimination of certain demographics.[99]

A number of sources highlighted the difficulty of maintaining data privacy when it comes to AI systems. For example, Harishbhai Tilala et al. asserted that privacy concerns related to AI systems transcend concerns related to traditional systems [63] while Jeyaraman et al. noted that big data use creates unique privacy issues such as loss of data control and the unauthorised use of personal data in predictive analysis.[64] Corfmat et al., and Jeyaraman et al., asserted that the variety of means of data collection made possible through AI use are more portable and diverse (such as through traditional healthcare systems as well as self-tracking using digital technologies; data shared on social networks and wellness applications) making protection, security, and confidentiality increasingly difficult, risk of ransomware attacks higher,[60,64] and in general opens the healthcare system to the security and privacy vulnerabilities of new medical devices in healthcare.[64] Marques and colleagues underscored the difficulties of ensuring that the security of data and maintenance of confidentiality is upheld at every level.[67] A report from the European Commission Joint Research Centre highlighted complexities regarding what happens when a person dies, for example, whether their data remains available to AI systems and whether their data can be inherited.[100]

A number of sources highlighted the risks to privacy related to AI systems. Two sources asserted that re-identification is possible, for example through reverse engineer AI algorithms,[21,95] highlighting the risks associated with assurances of de-identification. Several sources highlighted that protection of data from cybersecurity threats in particular is challenging but critical.[51,95] For example, the WHO highlight that inadequate data privacy and security can leave patients vulnerable to cyber-theft, accidental disclosure, and privacy concerns. They noted that such concerns are augmented for stigmatised and vulnerable populations. They also noted that "it may be illegal for third parties to use "new" health data", such as recommendations produced by an AI system using a person's health data.[51] The European Commission Joint Research Centre highlighted that data alterations can have severe consequences such as blackmail and discrediting of individuals and groups.[100] The College of Physicians and Surgeons of British Columbia highlighted that many openly available large language models at present do not comply with current privacy regulations,[82] a risk also highlighted by the Royal Australian and New Zealand College of Radiologists (RANZCR).[88] The Irish College of General Practitioners

(ICGP) also cautioned against sharing sensitive information with currently available AI tools given the lack of data security currently available.[101]

Five sources specified that informed consent for data collection and sharing is necessary to authorise data use,[17,21,70,81,86] while others implicitly referenced the need for informed consent through their reference to existing privacy laws.[62,96,101] The AMA specified that informed consent applies to both identified and de-identified patient data.[81] In contrast to practices that currently exist for non-AI tools, the Irish Platform for Patients' Organisations, Science and Industry (IPPOSI), in their 2024 report, outlined recommendations generated by a citizens' jury on AI who recommended that service users should have the right to opt out of AI-enabled care, that there should be an option to opt out of auto-enrolment of health data for training AI, and that patient choice should be central to any AI use in health and social care.[91]

Guidance and recommendations

Guidance and recommendations to promote privacy were noted in several sources.[6,17,21,60,62,64,66,67,70,76,81,82,86,91,94-96,99,101] For example, several sources noted that it is the duty of those entrusted with data, such as healthcare professionals, to maintain the privacy of patients' personal health information and to ensure that their patients' right for their health and social care information to remain confidential is upheld.[66,81,82,99]

Several sources provided guidance related to data; any data collected should be minimised to strictly necessary data;[63,67] datasets should be anonymised before public release;[17] data should be approved by patients before sharing; data should be securely stored, authorities should be created to manage data and protect confidentiality, and breaches should be reported;[21,81,95,96] and AI systems must comply with existing privacy and security laws.[62,86,96,101] One source suggested that a legally binding data processing contract between developers and users could all help to ensure privacy and confidentiality are upheld.[96]

Guidance was also provided in relation to informed consent. Several sources suggested that informed consent for data collection and sharing is necessary to authorise data use.[17,21,60,62,70,81,86,96,101] One source, the Australian Health Practitioner Regulation Agency, referenced specific AI tools and use cases highlighting that AI scribing tools that use generative AI to collect personal data legally require informed consent prior to recording of consultations.[94]

In contrast to practices that currently exist for non-AI tools, recommendations from the citizens' jury on AI conducted by IPPOSI asserted that patient choice should be central to any AI use in health and social care.[91] The Council of Europe also asserted that patients should be able to accept or refuse the use of AI systems in their care.[99]

3.3.4 Theme four: Human agency and oversight

Twenty-seven sources contained concepts captured by the theme "human agency and oversight".[6,8,9,51,57,60,61,68,71,74,76,77,82-84,86,89-92,95-98,100,102,103] Eight were academic sources and 19 were grey literature sources. (See appendix 6 for detail on the types of evidence supporting each theme).

Conceptualisation of human agency and oversight

The theme human agency and oversight emphasises the need for human involvement, oversight, monitoring, and governance throughout the development and use of AI systems.

Three sources discussed the relevance of oversight and evaluation prior to the use of AI systems.[60,96,97] In their literature review on AI development in healthcare, Corfmat et al. noted the importance of assessing potential negative outcomes associated with the use of an AI system, as well as benefits, risks and adherence to established ethical principles.[60] Similarly, the Australian Commission on Safety and Quality in Health Care called for "high-quality, local, practice-relevant" evidence to be available prior to AI use.[96] Eight sources further discussed the continuous monitoring and oversight of AI tools throughout usage.[57,60,71,77,82,96,97,103] Alelyani highlighted the importance of validation to ensure accurate and reliable outcomes, thereby promoting generalisability.[57] Corfmat et al. asserted it should be possible for AI systems to be assessed "continuously, systematically, and transparently".[60] The importance of monitoring is also noted by several other sources.[71,82,96,97,103] Regarding impacts of regular monitoring, the UK's Care Quality Commission noted that auditing can enable the identification and management of issues, thereby promoting safety and quality;[97] the ACSQHC similarly asserted that monitoring can highlight issues quickly, facilitating intervention.[96] Meanwhile, Tierney at al. noted the limits of currently available methods for assessing the quality of outputs from large language models.[77]

In their report identifying principles for the responsible and safe implementation of AI in healthcare, the Australian Commission on Safety and Quality in Health Care discussed "evaluation, monitoring and maintenance as an issue for governance".[96] Eight sources reflected on governance at various levels.[9,51,84,86,91,95-97] Of these, three discussed the importance of governance at the provider level.[86,91,97] Referring to diagnostic services using machine learning, the UK's Care Quality Commission attested that governance of the "clinical, information, technical and human aspects of the application" is required,[97] while the Royal Australian and New Zealand College of Radiologists noted the need for governance that is transparent and accountable to ensure oversight of AI use and monitoring, as well as compliance.[86] Other sources referred to governance at a national level, with the Australian Alliance for Artificial Intelligence in Healthcare discussing the potential benefits of a "whole-

of-government" approach,[95] while in their report the Australian Commission on Safety and Quality in Health Care noted that some countries centralise coordination of governance within a healthcare system.[96] The value of a strong governance framework to maximise the potential of AI with regulation based on risk and ethics was highlighted by MedTech Europe.[84]

The need for human involvement and oversight when using AI systems was highlighted by a range of sources.[57,61,68,74,77,83] The importance of healthcare providers being able to override AI decisions was noted, for example, by Kahraman et al. following their interviews with medical doctors[74] and Mennella et al. in their narrative review.[68] Based on reviewed evidence, Drabiak noted cautions against fully automating certain decisions or over-relying on AI systems.[61] Tierney and colleagues highlighted that AI systems are fallible and can provide erroneous recommendations based on errors in the data used to train them, and that human oversight can be a protective barrier.[77] Several sources highlighted the importance of preserving clinical expertise by involving humans in decision-making.[6,89,91,92] The Council of Europe caution against automation bias, which refers to over-reliance or over-trust by clinicians,[98] while a report from the MPS Foundation suggested that AI systems should provide information, not recommendations.[92]

Guidance and recommendations

Guidance and recommendations to uphold human agency and oversight were mentioned by a number of sources. In terms of monitoring and oversight, these included assessing AI tools prior to usage,[97] validating AI models,[57] revalidation during use,[103] using established safety and quality systems,[96] oversight involving updating and maintaining AI tools,[82] surveillance and evaluation,[103] regular auditing,[97] and continuous evaluation against defined principles.[71] Two sources also noted the need to increase capacity to facilitate evaluation and monitoring.[96,103] Regarding governance, the need for a policy context that enables responsible AI was discussed in the OECD's report on AI in health.[9] Other recommendations included requiring major healthcare settings to create an AI policy,[91] developing a framework for governing AI implementation, and creating practice standards and a risk-based framework to ensure safety.[96] The need for an AI in healthcare strategy that provides a regulatory framework to guide the development and use of AI was noted,[103] alongside calls for an independent commissioner, domestic legislation, and a regulatory body to be established.[91] Regarding human involvement, sources recommended that AI should not replace physician evaluation, interpretation, or existing validation procedures by qualified humans,[83] that decisions made by AI systems should be validated by adequately trained professionals,[98] and that it should be made clear in new healthcare AI policy guidance and guidance from healthcare organisations how clinicians should manage conflicts of opinion with AI systems.[92]

3.3.5 Theme five: Responsibility

Twenty-seven sources contained concepts captured by the theme "responsibility".[6,8,9,17,21,51,57,60-62,64,66-68,71,73,75,76,81,83,85,86,91,94,96,99,101] Fourteen were academic sources and 13 were grey literature sources (See appendix 6 for detail on the types of evidence supporting each theme).

Conceptualisation of responsibility

The theme responsibility captures the importance of responsibility and accountability. It includes challenges arising in the context of AI, clarity on roles and responsibilities throughout development and use, including shared responsibilities, and compliance with legal, ethical and regulatory requirements.

The majority of sources highlighted the importance of responsibility and accountability in the use of AI, with Elendu et al. emphasising that clarity over responsibilities is "essential for ethical use".[62] The European Patients' Forum noted the significance of responsibility for promoting ethics and fairness in the use of AI[6] and MedTech Europe highlighted the importance of accountability for quality, compliance, safety, approval, and readiness for inspection.[8] However, several sources further reflected on challenges raised in the context of AI.[21,62,64,67,68,73,96] For example, Elgin and Elgin described uncertainties among healthcare professionals about the assignment of responsibility if decisions informed by AI lead to a negative outcome;[73] Marques et al. noted ethical and legal questions regarding the responsibilities of healthcare professionals and developers of AI tools;[67] and Savulescu et al. cited the attribution of responsibility among the "great conundrums" for clinicians using AI.[21]

The involvement of humans was discussed by nine sources in relation to responsibility and accountability.[17,51,61,64,68,71,73,90,94] For example, in their review article, Drabiak noted the idea that moral accountability promotes the interests of people using services, acknowledging the challenge arising for decisions based on AI systems.[61] Similarly, Jeyaraman et al. cited discussions of moral agency as a human quality that is not present in AI, noting calls to analyse the "causal chain of human agency" to support attribution of accountability.[64]

The need for clarity on roles and responsibilities was discussed in eight sources, with some also offering views on the assignment of responsibility.[6,17,21,76,81,83,85,96] Jha et al. stated that roles and responsibilities of care providers, AI developers and vendors must be defined,[17] while the Australian Medical Association simultaneously called for clear lines of accountability for the use of AI in healthcare.[81] When referring to responsibility for harm, Savulescu et al. indicated that practitioners would bear responsibility if they do not evaluate AI performance, communicate risks, benefits, alternatives and confidence in the tool, and use AI appropriately, while noting that AI designers respectively need to ensure the safety, reliability and effectiveness of

the AI system, and to clarify the values driving it and its limits for specific groups.[21] In their position paper, the College of Physicians and Surgeons of British Columbia outlined that medical directors are responsible for "acceptance testing and quality assurance", and physicians hold responsibility for the "clinical interpretation and management" of results from AI tools.[83] Finally, the UK's Careworkers' Charity expressed their expectation that employers bear responsibility if a harm occurs from the use of AI, if policies and procedures were appropriately followed.[85]

Five sources referred to diffusion of responsibility and or collective responsibilities.[51,64,68,86,96] with Jeyaraman et al. outlining that diffusion of responsibility can arise when there are multiple options and a number of parties involved.[64] The WHO proposed "collective responsibility" as a mechanism through which all involved in developing and using AI are positioned as accountable, thereby countering the diffusion of responsibility, promoting integrity and limiting harm.[51,68,96] The Royal Australian and New Zealand College of Radiologists further specified the need for shared responsibility between healthcare professionals, service management, and AI developers.[86]

Four sources referred to compliance with existing legislation and policies when implementing AI.[57,71,96,99] This included specific mention of medical device regulations and privacy laws,[57] data privacy, consumer law, cybersecurity policy, and national ethics frameworks,[96] as well as general reference to legal and regulatory frameworks.[71] The Council of Europe further noted key policy and legal documents such as the Declaration of Geneva and the Oviedo Convention and highlighted that professional obligations, for example codes of conduct or legal obligations, ensure standards and quality of care.[99] They also asserted that guidelines and regulations are needed for AI development and use, noting that such guidelines should ensure developers are accountable for AI system errors.[99] They further suggested that regulatory compliance is governed by regulatory bodies and asserted that healthcare professionals and healthcare providers are obligated to use AI in line with ethical and legal guidelines.[99] The duty and obligations of care providers were further discussed by others, including the importance of fidelity and professional faithfulness,[68] and the responsibility to deliver care in accordance with professional obligations.[94]

Guidance and recommendations

Guidance and recommendations to promote responsibility were noted in a number of sources. Cited mechanisms included auditability of AI systems,[66] transparency and traceability,[64] clinicians' evaluation of evidence and appropriate use of AI,[21] and informing service users of benefits, risks and confidence in a tool.[21] "Human warranty" was additionally noted as an approach to promote responsibility during the development and use of AI systems,[17,51] with others citing human assurance,[68] human oversight,[6,71,94] and measures to ensure professional control.[73]

Recommendations also included clarifying individual and organisational responsibilities for decisions informed by AI,[96] and seeking advice regarding the shifting of responsibility from AI developers to users.[101] Finally, redress was discussed in five sources,[6,17,51,66,68] with recommendations that redress mechanisms are established, adequate, and accessible.[6,17,51,66]

3.3.6 Theme six: Upholding people's rights

Twenty-five sources contained concepts captured by the theme "upholding people's rights".[6,9,15,17,21,51,59,60,62,63,66,68,71,74-76,79,81,86,87,89,91,94,96,98] Fourteen were academic sources and 11 were grey literature sources (See appendix 6 for detail on the types of evidence supporting each theme).

Conceptualisation of upholding people's rights

The theme of upholding people's rights in the literature refers to AI systems respecting and protecting internationally recognised human rights.

Eight sources outlined how AI use should respect the fundamental principles of healthcare including adhering to and protecting human rights.[6,71,79,86,87,89,91,94] For example, the OECD report discussed how AI should be designed in a way that "respects the rule of law, human rights, democratic values, and diversity" and that there should be safeguards in place to ensure that AI is used in a fair way.[9]

Eight articles focused specifically on ensuring that a patient's autonomy is respected when AI is used in their care.[17,21,59,60,62,63,66,68] Armitage, Maccaro et al. and Elendu et al. noted that a patient should have a right to make their own choices about the use of AI in their care without undue pressure, solicitation or coercion.[59,62,66] Mennella et al. advocated that individuals have a right to make an informed decision about the use of AI in their care.[68] Jha et al. suggested that the integration of AI into healthcare must be balanced with a patient's autonomy. They noted that a fundamental principle of healthcare is that competent adults have the right to make an informed decision about their medical care, even if AI systems increasingly inform clinical decision-making.[17] Harishbhai Tilala et al. similarly noted that central to the concept of autonomy is the recognition that individuals are rational agents capable of self-determination and personal choice. They advocated for shared decision-making between a professional and a patient because they noted that "respecting patients' autonomy not only fosters trust and collaboration but also upholds their inherent dignity and autonomy as moral agents."[63] Corfmat et al. also noted how AI can erode a patient' autonomy, and advocated for patients' co-participation in their care and their ability to refuse care or request additional medical advice if required.[60] Savulescu et al., suggested that AI could enhance patients' autonomy, enhancing decision-making and patient empowerment.[21]

Four articles referenced the role of informed consent in ensuring that a patient's autonomy is respected.[51,59,60,75] The WHO noted that, in relation to healthcare data, the collection of data without the informed consent of an individual "undermines the agency, dignity and human rights of those individuals".[51]

Four articles discussed the role of professional autonomy in ensuring that people who use services' rights are respected.[51,76,81,98] The Council of Europe noted that protecting professionals' autonomy and decision-making power over an AI tool is critical.[98] The WHO also noted that any extension of AI autonomy should not undermine human autonomy.[51]

Guidance and recommendations

Guidance and recommendations to uphold people's rights were noted in several sources.[6,51,60,79,81,86,87,89,94] Several sources advocated for the role of the service and the professional in ensuring that an individual's human rights are respected.[6,86,87,89,94] Examples provided included that professionals must always protect patients' fundamental rights and ensure that individuals receive care in line with their preferences and values.[86,87] Professionals also have an obligation to respect people's rights such as data confidentiality and privacy and to protect people from diverse backgrounds from the risk of bias.[94]

Several sources noted that AI tools should not operate autonomously if they cannot function in line with established medical and ethical principles.[81,86,87] For example, the evidence asserts that a medical professional must always maintain clinical independence and have oversight over any decision from an AI tool.[81,86,87] If an occasion arises when a medical professional determines that the treatment or management of a patient is different from an AI tool, healthcare organisation protocols must ensure that clinical independence is not undermined by AI.[81]

Several sources noted the importance of information. For example, services must ensure that staff have sufficient information to use AI in a safe and effective way and that they can use this information to support people who use services to understand the role of AI in their care.[51] Patients should be provided with objective, accurate and easily understandable information prior to giving informed consent as it can be difficult for patients to challenge a decision if the healthcare professional cannot explain clearly to the patient how or why they propose a certain treatment or procedure.[60]

3.3.7 Theme seven: Safe care

Twenty-three sources contained concepts captured by the theme "safe care".[6,8,9,15,17,21,51,57,59,63,67,68,70,74,75,81,86,89,91,95-98] Eleven were academic sources and 12 were grey literature sources (See appendix 6 for detail on the types of evidence supporting each theme).

Conceptualisation of safe care

The theme safe care considers the duty to ensure that AI is used safely and to the benefit of people using services, mitigating mental or physical harm, as well the types of risks, including novel considerations for the safe delivery of care when using AI.

Nine sources reflected on the duty to ensure safe care and mitigate harm.[51,57,59,63,68,70,74,75,91] In general terms, four sources outlined the importance of acting for the benefit of people who use services while preventing and avoiding harm,[59,63,68,74] with seven elaborating further on this in the context of using AI.[51,57,59,68,70,75,91] In a qualitative study involving experts in autonomous systems technologies in healthcare, Alelyani noted that safety involves the assurance of no harm or risk arising from use of an AI system.[57] Meanwhile, a review from Armitage noted that where a tool has plausible benefits for outcomes, there may be a duty to introduce the tool. Similarly, they posited that where there is a notable likelihood of harm or worse outcomes, there may be a duty to not introduce the tool.[59] Further discussions across the evidence sources referred to the need for proportionality in appraising risks,[75] oversight to ensure AI models function as intended,[68] identifying and avoiding "foreseeable and unintentional harms" arising from AI,[70] prioritising "the common good",[91] and mitigating physical or mental harm.[51] In addition, the WHO outlined that prevention of harm also includes managing AI-derived diagnoses or warnings that cannot be addressed, for example due to inaccessible or unaffordable care, in a careful and balanced way.[51]

Three sources noted that AI introduces novel challenges and new types of considerations for ensuring safe care.[17,81,86] For example, the Royal Australian and New Zealand College of Radiologists highlighted that alongside the potential to enhance care, the introduction of AI tools brings "new risks".[86] Several other sources discussed types of risks and potential harms, with Reddy et al. describing the need to identify and avoid "physical, psychological, emotional or economic" harms.[70] The European Patients' Forum referred to unvalidated AI tools and inadequate transparency as risks to safety,[6] and also highlighted risks related to incorrect decisions and overdiagnosis.[89]

Guidance and recommendations

Guidance and recommendations to promote safe care were highlighted in several sources. Eight referred to oversight measures including a "human-in-the-loop",[17] monitoring and evaluation,[17,70,81,91] post-market surveillance,[6,95] risk assessment and mitigation,[9,63,81,91,97] regular updates using current data and professional feedback,[17] validation and assessment,[6,81,98] and national safety monitoring systems.[95] Other noted measures included regulation of the safety of AI systems[95] and adherence to regulatory standards for safe, accurate and effective

technologies,[68] as well as training,[81,86,95] use of AI in suitable conditions,[86] and limiting AI to particular uses in specific settings.[91] Alongside the use of safety protocols and frameworks,[63,95] additional measures to promote safe care included high-quality data,[8] informed consent,[63] co-design,[95] and a "whole-of-system approach to safe AI implementation".[96] Finally, characteristics of safe care were described as including a managed approach to AI roll-out and use, protocols for responding to disagreements between an AI system and care provider, robust care pathways and contingency plans, availability of resources to validate systems before use, implementation of clinical reviews, and, where indicated, a phase of "shadow reporting" to facilitate learning.[97]

3.3.8 Theme eight: Integration into care

Twenty-one sources contained concepts captured by the theme "integration into care".[7,9,17,51,57,61,68,71,75,76,79,80,84,86,91,92,96,97,101-103] Ten were academic sources and 11 were grey literature sources (See appendix 6 for detail on the types of evidence supporting each theme).

Conceptualisation of integration into care

The theme of integration into care in the literature refers to the aspects that need to be accounted for to facilitate incorporation of AI systems into existing health and social care systems. This theme incorporates the sub-theme eco-responsibility and sustainability which refers to the importance of prioritising sustainability and considering the environmental impact of AI systems prior to integration into care for current and future generations. It also incorporates the obligation to design AI systems to minimise their environmental consequences, energy consumption, and carbon footprint.

Three sources highlighted the importance of interoperability with existing systems when incorporating AI systems.[7,91,101] Seven sources asserted that stakeholder engagement is key to enabling the integration of AI into existing health and social care systems.[7,86,91,92,96,102,103] Three sources suggested that easy-to-use, straightforward AI systems that are adaptable to existing health and social care settings will be most easily implemented and should therefore be prioritised.[7,92] Two sources asserted that only tested AI systems should be integrated into clinical care and that people using services should never be exposed to untested AI systems.[7,97,101] Three sources highlighted the importance of universality, generalisability, and transferability of AI systems.[7,76,91] so that they can benefit the maximum number of patients and enable operations across services and contexts. Seven sources highlighted the need for AI to add value and improve health and social care in comparison with the use of non-AI systems.[57,61,71,91,96,97,102] If implementation of an AI system is being considered, it should be clear how patients will benefit from new insights that were not previously available.[57,71,91,96] Four

sources highlighted the importance of research to support the successful integration of AI into health and social care.[76,84,96,103] The ACSQHC provided guidance on the value of research for ensuring that AI is fit for purpose and to assess overall its clinical utility pre and post commencement of AI use.[96]

Guidance and recommendations

Guidance and recommendations to promote integration into care were noted in several sources. Six sources highlighted the importance of direct contribution from all relevant stakeholders including the public, patients, and healthcare staff when integrating AI systems.[7,91,92,96,102,103] For example, organisations should facilitate public consultation and incorporate this feedback.[96] In particular, the importance of ensuring diversity of any stakeholders engaged with, to avoid bias or discriminatory outcomes, was stressed.[7,91,103]

Two sources suggested that AI systems with the highest clinical relevance should be prioritised.[57,96] To achieve this, areas of high priority should be identified and documented and service providers and researchers should be liaised with to identify areas of highest needs.[96] Developers should list expected benefits associated with an AI system to clearly demonstrate the additional value that the AI system brings.[71]

Only tools that have been rigorously tested, gone through appropriate approval procedures, have been verified as clinically useful and safe through robust research, and lead to better health outcomes for people who use services should be integrated into the health and social care system.[7] Lekadir et al., suggested that AI systems should be tested in multiple contexts and settings, and external datasets not used in model training should be assessed to ensure generalisability of AI systems.[7]

It was suggested that AI systems will be more widely integrated into clinical care to the extent that they are easy to use to achieve clinical goals.[7] To ensure that users of AI systems are able to efficiently and confidently use AI systems, a concise but sufficient amount of information should be given about AI systems. Too much information can lead to unreadability, too little can lead to lack of trust and transparency, and inability to understand how the AI system generates information.[92]

Several sources commented on adaptability to existing systems. Lekadir et al. suggested that developers should adhere to existing community defined standards such as clinical definitions, medical ontologies, data annotation protocols, and technical standards.[7] IPPOSI, the patient organisation, highlighted that not only do AI systems need to be interoperable with existing systems, but existing systems also need to be adequately digitalised to facilitate the incorporation of AI systems. In particular, they referenced the urgency to implement the national electronic health record system in Ireland.[91]

The Royal Australian and New Zealand College of Radiologists suggested that to achieve optimal integration into clinical care at the system level, new teams with specialisation and expertise in AI systems should be formed.[86] A "centralised, accessible and interoperable" approach should exist.[71,91] AI systems should be monitored and audited to assess their effectiveness and to ensure the AI system is achieving what it was designed to do.[61,97] Governance bodies should enforce regular reviews and oversight, evaluation capacity and infrastructure should be invested in, organisation readiness assessments should be undertaken before adopting AI, health and social care services should be resilient to technology failures, independent evaluations and audits should be incorporated, and assessment results should be shared across health and social care systems.[51,71,76,103]

Conceptualisation of subtheme: Eco-responsibility and sustainability

This theme incorporates the sub-theme eco-responsibility and sustainability, which refers to the importance of prioritising sustainability and considering the environmental impact of AI systems prior to integration into care to enable responsible and safe care for current and future generations. It also acknowledges the obligation to design AI systems to minimise their environmental consequences, energy consumption, and carbon footprint.

Five sources highlighted the environmental burden of AI systems.[17,51,75,79,91,96] The ACSQHC highlighted the significant environmental footprint and large energy consumption of AI systems.[96] Katirai mentioned, in particular, the materials used for hardware such as rare earth metals, and the high carbon emissions from creating and using AI systems and the data centres required for AI functionality.[75] The patient organisation IPPOSI highlighted the importance of considering the health of future generations.[91] In their discussion paper, Seroussi and Zablit highlighted that the health impacts resulting from the carbon footprint of AI systems must be acknowledged and accounted for and that AI systems must contribute substantially more to lives saved than lives harmed through their environmental burden.[79] Jha and colleagues asserted that this responsibility lies with AI developers who have a duty to be environmentally conscious.[17] The WHO highlighted the importance of promoting sustainability and consideration of environmental impacts in AI system design and use.[51]

Three sources mentioned social sustainability.[51,68,75] The WHO asserted that sustainability includes minimisation of disruption to the workplace, including potential job losses due to automation, and the importance of training to aid adaptation of the workforce to new systems.[51]

Guidance and recommendations

Three sources suggested ways sustainability can be achieved.[17,51,79] Seroussi and Zablit suggested that sustainability can be achieved by including eco-design of code,

computation of digital health services' carbon footprint, making services accessible at lower speeds and older generations of devices, and favouring servers that implement energy consumption conservation processes.[79] Jha and colleagues suggested that regular monitoring and updating of AI systems should be implemented to ensure systems are efficient, adequate, and adhere to sustainable development practices. They suggested that AI systems that are efficient and that minimise energy consumption and carbon footprint should be favoured.[17] The WHO suggested that AI systems should adhere to global efforts to reduce the impact on the environment. They suggested that AI systems should be designed to minimise their environmental burden and increase energy efficiency.[51]

With regard to social sustainability, several sources highlighted the importance of training and protecting against job loss and governments' responsibility to anticipate and protect against disruptions to social sustainability.[51,68,75]

3.3.9 Theme nine: Education, training, development, and information provision

Eighteen sources contained concepts captured by the theme "education, training, development, and information provision".[6,9,51,60,61,73,74,76,84,86,89,91,92,94-96,98,103] Five were academic sources and 13 were grey literature sources (See appendix 6 for detail on the types of evidence supporting each theme).

Conceptualisation of education, training, development and information provision

The theme of education, training, development, and information provision encapsulates the essential role of both professional and public education and information provision in supporting informed, confident, responsible use of AI. Public Education and Information Provision is discussed as a sub-theme.

Seven of the sources discussed the type of information and education that staff need to use AI in a safe way.[51,60,73,84,92,94,96] The ACSQHC, based on their scoping review and environmental scan, noted consensus in the literature that training and support for staff is needed prior to the implementation of AI systems.[96] The Australian Health Practitioner Regulation Agency (Ahpra) outlined the importance of professionals understanding the intended use of an AI tool, to inform decisions of when is appropriate to use the outcome of an AI system and of any associated risks, including diagnostic accuracy, data privacy and ethical considerations.[94] MedTech Europe referred to education and training programmes to equip the workforce with skills to maximise the positive impact of AI.[84] Two other sources further outlined that training programmes should be provided for healthcare workers to improve skills in AI and to understand the technical, ethical and legal aspects.[60,96] Interviews about AI clinical decision support systems highlighted that professionals would like an AI literacy programme to focus on evaluating and explaining AI recommendations in resource allocation decisions, developing skills to recognise potential bias in AI recommendations, and building competency to determine when to override AI

recommendations based on person-specific factors.[73] A report from the Medical Protection Society (MPS) Foundation suggested that training could support clinicians to assess whether information provided by AI tools is appropriate for the person using the service.[92] Meanwhile, Ahpra also outlined that, at a minimum, professionals should review product information about an AI tool prior to use, including how it has been trained and tested on populations, where data will be located and how it will be stored, the tool's limitations and clinical contexts where it should not be used.[94] The WHO proposed that any requirement for education and training of professionals should extend beyond clinical care to those working in public health, surveillance, the environment, prevention, protection, education, awareness, diet, nutrition and all other social determinants of health that could be impacted by AI.[51]

Two sources voiced concerns about the risk AI can pose to the future deskilling of professionals as a result of increasing dependence on the AI tool,[51,60] potentially resulting in a situation whereby professionals are incapable of acting if an AI system fails or is compromised.[51]

Guidance and recommendations

Guidance and recommendations related to education, training, development and information provision was outlined in a number of sources, including the types of education and training initiatives needed to enhance professionals' competence when using AI,[60,91,92,96] with some differences in approaches across the sources cited. Suggestions included integrating AI training into medical training programmes and providing continuous education and training,[60,91] developing training for professionals in consultation with the AI developer alongside clinical governance, patient safety and clinical leaders,[96] and training and information for professionals on the use of AI tools provided by AI companies.[92] At a national level, recommendations included planning to develop the current and future workforce with skills needed for using AI and to establish career paths that allow professionals to specialise in AI,[103] as well as proposals for a national education programme.[95] Finally, it was noted that while not all staff may not need to know in-depth about AI, there is a need to explore what practical skills are required by different specialities across healthcare.[95]

Conceptualisation of subtheme: Public education and information provision

The sub-theme public education and information provision encapsulates the importance of educating and informing the public to ensure adequate AI literacy and to enable informed decisions.

The European Patients' Forum noted that health literacy – incorporating digital health literacy and data literacy – is crucial to strengthen knowledge and trust of AI among people who use services, and to support them to exercise their rights while

also realising the benefits of AI in healthcare. It was proposed that public education and health literacy can improve capacity to engage in the development of policy and practice on AI in healthcare.[89] The European Patients' Forum also advocated for increasing the public's education on AI tools and digital literacy to improve confidence in the use of AI, address concerns related to digital hesitancy, and mitigate health inequalities.[6]

Regarding informed consent, analysis of the cited literature suggested there are two aspects of consent to consider: one aspect relates to the right to be sufficiently informed regarding the care one receives in order to make an informed decision relating to the use of AI in care. The other aspect relates to data consent when one's personal data is collected or used by an AI system. The former will be discussed in this section. (The latter relates to the legal requirement to receive informed consent relating to collection of personal data and is discussed under the theme "Privacy".)[66,101] Five articles focused on the importance of enabling people using services to make informed decisions.[61,89,91,96,98] There is no explicit requirement to receive written consent, but, people using services should be sufficiently informed regarding AI use in their care.[61,89,91,92,94,96,98] A report from the Council of Europe asserts that, given the complexity of AI systems it can be challenging to provide plain language explanations in a way that is understandable to those without technical knowledge, however, in order to ensure that a service user can make an informed decision they must be provided with information in an accessible format to them.[98] Notwithstanding this difficulty, the same report asserted that doctors have a responsibility to receive simple explanations regarding AI systems from developers and translate this explanation to service users in a simple yet meaningful format.[98]

Some sources, such as the ACSQHC's international literature review, and Kahraman and colleagues' semi-structured interviews with medical doctors, discussed the possibility of leveraging pre-existing systems and consent procedures.[74,96] Kahraman and colleagues noted that participants in this study felt that consent is already obtained for other procedures, and simply informing the patient that AI systems are being used is a sufficient addition to consent already gathered, whereas others stressed that informing the patient about the use of an AI system was a necessity.[74] This article also highlighted that the doctors interviewed find obtaining signed consent forms significantly burdensome.[74]

Disclosure was discussed in two articles.[61,92] In their review article, Drabiak highlighted mixed opinions in the literature regarding disclosure of AI tool use as part of care; some perspectives favoured disclosure requirements, asserting that physicians have a duty of transparency which includes the duty to disclose use of AI systems; other perspectives favoured non-disclosure, asserting that non-AI tools do not necessarily require disclosure, and that too much disclosure can be

overwhelming and confusing to people using services.[61] A report from the MPS Foundation also suggested that, given current practices for non-AI tools do not necessarily require disclosure, depending on the context, disclosing the use of AI tools should be at the discretion of clinicians.[92]

Guidance and recommendations

Guidance and recommendations to promote public education and information provision were noted in a number of sources. Recommendations included accessible language to facilitate informed decision-making,[96,98] providing information that is comprehensible, in plain language, and clearly outlines the limitations of AI systems and alternative options for care,[96] leveraging pre-existing systems and consent procedures[74,92,96] and promoting AI literacy among people using services and service providers to facilitate informed decision-making.[96] Recommendations also included national engagement campaigns that educate on the goals, benefits and risks of using AI in healthcare,[91] and from a provider perspective, information from regulatory officials and healthcare organisations to guide clinicians' decisions regarding disclosure of AI use.[92]

3.3.10 Theme ten: Data quality

Nine sources contained concepts captured by the theme "data quality".[6,57,64,70,84,89,90,101,102] Three were academic sources and six were grey literature sources (See appendix 6 for detail on the types of evidence supporting each theme).

Conceptualisation of data quality

The theme data quality encapsulates the need for data used to train AI systems to be consistent, standardised, accurate, complete and reliable, and align with privacy and cybersecurity requirements.

All nine sources noted the importance of high-quality data to ensure that the outcomes of an AI system are reliable and trustworthy.[6,57,64,70,84,89,90,101,102] As Reddy et al. noted from their review of the literature, "An AI system is only as good as the data it was derived from. If the data do not reflect the intended purpose, the model predictions are likely to be useless or even harmful".[70] Alelyani, MedTech Europe and the European Patients' Forum further discussed the importance of using high-quality data to develop AI systems. Data should be accurate, standardised, interoperable, unbiased and reliable, as poor quality data can limit the potential of AI to be useful and safe, and can lead to erroneous outcomes including error, or over- or under-diagnosis.[57,84,89] As the ICGP emphasised, no AI tool is 100% accurate.[101] Furthermore, Jeyaraman and colleagues highlighted that one of the challenges to ensuring the quality of AI outputs is hallucinations, where models produce inaccurate or misleading information that appears to be factual or coherent.[64] In

terms of barriers related to data quality, the European Patients' Forum indicated that one of the challenges to developing effective AI algorithms is securing data from health institutions with proper privacy protections.[90]

Guidance and recommendations

Guidance and recommendations relating to data quality were noted in a number of sources. Cited mechanisms included infrastructure that enables a flow of consistent data in standardised formats and with necessary cybersecurity provisions,[84] identifying and mitigating hallucinations,[64] and human oversight to monitor AI outcomes.[6,101] At a European level, it was noted that the European Health Data Space Regulation can contribute to ensuring data is of high quality and suitable for AI purposes.[6]

3.3.11 Theme eleven: Technical robustness and security

Eight sources contained concepts captured by the theme "technical robustness and security".[7,15,57,65,74,91,96,102] Five were academic sources and three were grey literature sources (See appendix 6 for detail on the types of evidence supporting each theme).

Conceptualisation of technical robustness and security

The theme technical robustness and security encapsulates the importance of AI systems being secure against unauthorised access, breaches, attacks, and malicious interference, as well as ensuring their performance under varying conditions.

Three sources pointed to the significance of ensuring that AI systems are secure and robust.[7,57,65] For example, in their qualitative study involving experts in autonomous systems, Alelyani underlined the role of appropriate security measures to protect patient safety and mitigate risks.[57] The reviewed evidence suggested such measures are needed to ensure the functionality, performance, reliability, accuracy and integrity of AI systems.[7,57,65]

Three sources referred to circumstances that potentially implicate the security and robustness of AI systems.[7,57,96] Alelyani cited errors, cyberattacks and unexpected situations, as well as unauthorised access, data breaches, and malicious activity.[57] Regarding the robustness of AI systems, in their international consensus guideline for trustworthy and useable AI, Lekadir et al. highlighted the impact of anticipated and unanticipated variations in data on AI system performance.[7] The Australian Commission on Safety and Quality in Health Care referred specifically to cybersecurity issues, highlighting potential impacts such as privacy issues, data manipulation, disruption to critical infrastructure, and inaccuracies in results and recommendations.[96]

Guidance and recommendations

Guidance and recommendations to promote the technical robustness and security of AI systems were offered by several sources.[7,15,57,74,91,96,102] Referring to cybersecurity, recommendations included taking a risk-based approach, implementing data security plans, ensuring transparency, adhering to regulation, promoting clinicians' knowledge of cybersecurity, employing a user registry, and updating security measures regularly as required.[96] Several sources referenced measures across the AI lifecycle to ensure security and robustness, for example, designing, developing, evaluating and optimising AI systems for robustness against variations in data,[7] keeping security central to the design of AI systems,[102] and validating systems regularly.[15] In addition to robust security measures,[57] other approaches mentioned in the evidence included contingency planning,[57] secure data storage,[74] and regulatory mechanisms to safeguard security.[74,91]

3.3.12 Theme twelve: Human connection

Three sources contained concepts captured by the theme "human connection".[21,73,102] Two were academic sources and one was a grey literature source (See appendix 6 for detail on the types of evidence supporting each theme).

Conceptualisation of human connection

The theme of human connection in the literature refers to the importance of person-centred, caring relationships between patients and health and social care providers.

Three sources highlighted the importance of the relationship between people working in, and people using, health and social care services.[21,73,102] Elgin and Elgin asserted that the therapeutic, human relationship formed between a service provider and user should remain an important part of care. They highlighted the unique advantages that human empathy, instinct, and personal connection bring to the experience of health and social care.[73] They also highlighted the importance of patient values, and Savulescu et al. asserted that paternalism should be minimised.[21] They defined paternalism as making a decision on someone else's behalf or doing what is in someone's best interests, even when that is against the person's will. They further explained that soft paternalism refers to making a decision for someone whose decision-making capacity is hindered whereas hard paternalism refers to making a decision for someone whose decision-making capability is competent.[21]

Guidance and recommendations

Guidance and recommendations to promote human connection were noted in some sources.[21,73] It was suggested that service providers should integrate person-centred values when interpreting AI system recommendations and aid people using services in making the best decision in relation to their health that aligns with their

own values and preferences.[21,73] To minimise paternalism, Savulescu et al. suggested that people who use a service should be involved in decisions, their values and preferences should never be overridden by AI recommendations, and AI systems should be developed with the ability to incorporate patient values. If this is not possible, service providers should act as a mediator to ensure AI outputs are considered as secondary to patient values and decisions.[21]

4. Discussion

4.1. Overview of findings

While existing national strategies in Ireland acknowledge the importance of responsible and safe use of AI, current guidance largely reflects high-level EU guidance that is not specific to health and social care and may lack the specificity needed for real-world application in these settings. This review fills an important gap by consolidating sector-relevant guidance and aligning this evidence with the Irish context through the mapping of themes to established principles for person-centred care and support.

The purpose of this evidence review was to identify concepts and guidance to facilitate the responsible and safe use of AI in health and social care services. Fifty-five relevant sources were identified across the academic and grey literature, representing perspectives including professional associations, national professional or regulatory bodies, patient organisations, academic researchers, industry, and intergovernmental organisations. The majority of the sources had a global focus, while the remainder were from specific geographical regions, including Ireland. A key feature of this review was that many of the included sources explored the use of AI in healthcare, with limited representation of social care perspectives. This means the findings from the review overall take a healthcare focus.

While 12 unique themes were generated, there was considerable interdependence and overlap between themes. That is, guidance and recommendations relevant to one theme may also promote the realisation of another theme. For example, promoting transparency is necessary for oversight, which is central to the theme of human agency and oversight; the theme of inclusivity and non-discrimination is distinct from the theme of upholding people's rights, however, neither could be actualised without the other; ensuring the quality and diversity of data to train AI systems, as outlined in the theme data quality, is also vital to upholding the theme of inclusivity and non-discrimination; finally, adherence to the theme of privacy requires AI systems to be secure, which is a key component of the theme of technical robustness and security.

4.2. Mapping to evidence-based principles for person-centred care and support

A principles-based approach emphasises adherence to broad, fundamental, pre-established principles. Such an approach enables flexibility, adaptability with existing systems and across various contexts, and alignment with long-term objectives. In the Irish context, HIQA has developed a set of principles that work together to achieve person-centred care and support: accountability, a human rights-based approach; safety and wellbeing; and responsiveness.[104,105] These principles are evidence-based, have been developed for use across all health and social care settings and underpin national standards and guidance developed by HIQA.[105] The themes identified in this evidence review have been mapped against the HIQA principles in a way that is consistent with previous approaches in the Irish health and social care system so that services and staff can conceptualise and understand how to use the guidance to ensure person-centred care and support.

The 12 themes were mapped to the four evidence-based principles for person-centred care and support, aiming to situate the international evidence on responsible and safe use of AI according to established principles relevant to the Irish health and social care system.[104,105] Mapping was conducted independently by three members of the review team, with discussion to resolve any discrepancies. Each theme was primarily categorised under the principle that was considered to most reflect its conceptualisation, guidance and recommendations. Due to interdependence among the themes, and also among the principles, some themes were further sub-categorised under other principles where that theme was considered relevant to more than one principle. Results are summarised in table 3 and discussed below, with particular consideration given to the implications for the responsible and safe use of AI tools in health and social care services in Ireland.

Table 3. Matrix showing themes cross-cutting with evidence-based HIQA principles for person-centred care and support

Themes/Principles	Accountability	A human rights-based approach	Safety and wellbeing	Responsiveness
Transparency	**X**	x	x	x
Inclusivity and non-discrimination	x	**X**		
Upholding people's rights	x	**X**		
Data quality	x		x	**X**
Technical robustness and security	x		**X**	x
Human agency and oversight	**X**			
Safe care	x	x	**X**	
Human connection	x	**X**	x	
Responsibility	**X**			
Integration into care	x			**X**
Education, training, development and information provision	x	x		**X**
Privacy	x	**X**	x	

*Note: **X** = primary categorisation; x = sub-categorisation*

4.2.1 Accountability

HIQA conceptualisation

Accountability as defined by HIQA involves services having appropriate governance structures in place to deliver 'high-quality care and support that is consistent, coordinated, and focused on achieving the best outcomes for people using services.'[104] While accountability is a standalone principle, and the cornerstone for the responsible and safe use of AI in health and social care, it is embedded across the three other principles defined in HIQA's Standards Development Framework.[105] Therefore, while three themes – Transparency, Human agency and oversight and Responsibility – were primarily categorised under accountability, the nine other themes were also sub-categorised as relevant to the principle of accountability.

Mapping HIQA conceptualisation to evidence from this review

An accountable service has a governance framework in place, with formalised arrangements including clear lines of accountability at an individual, team and service level so everyone is aware of their roles and responsibilities in the service in relation to, for example, the use of AI.[104] The review emphasised that oversight of AI, in particular continuous and transparent monitoring, is integral to the responsible

and safe use of AI tools within a service. The findings suggest that services should have quality control measures in place, appropriate assessment of tools prior to AI usage, and have regular surveillance and auditing of AI tools once in place.

Several sources in the review noted the importance of preserving clinical expertise. There was caution in the literature regarding fully automating decisions and over-reliance or over-trust on AI systems by clinicians, and the evidence suggested that providers should have the option to override an AI recommendation[6,89,91,92]

An accountable service has an open culture. Service delivery is only within the scope of what it can do safely and effectively, and there is open communication, trust and integrity in terms of how the service is delivered.[104] The evidence in this review highlights the importance of transparency in relation to how AI is used within services.[6,81,82,96] While many sources in the literature highlight that service users should be aware of when AI is used in their care, some differences were also noted, for example, the Australian Medical Association stated the importance of transparency if an AI system is used to inform a diagnosis or treatment recommendation,[81] while the College of Physicians and Surgeons of British Columbia noted that transparency is needed in relation to the extent to which a service provider is relying on an AI system in clinical decision-making.[82]

The review identified that, at a practical level, in order to support transparency within a service there should be record-keeping and a clear audit trail to ensure outputs are traceable, an inventory of all AI tools in use in the service is in place, and information on the use of AI in the service is available in a clear and accessible format for service users.[51,61,66-68,74]

Additional findings also highlighted the importance of explainability, in other words that services must be able to explain some level of the logic and understanding of the reasoning behind how AI produces outputs. The Irish College of General Practitioners noted that it is important that service providers and staff have some understanding of how AI reaches its outcomes.[101] However, the reviewed evidence suggests that the level of understanding necessary may vary according to the nature of contexts and decisions, with strong emphasis on explainability for decisions directly related to care.[73]

As outlined, effective monitoring and oversight is dependent on clearly delineated roles and responsibilities. The College of Physicians and Surgeons of British Columbia emphasised that management are responsible for the acceptance testing and quality assurance of AI tools while clinician's hold responsibility for the interpretation and management of results.[83] Nonetheless, there is consensus in the literature that there should be collective responsibility, whereby all those who are involved in the development and use of AI are accountable and have shared responsibility for its use, thereby avoiding any potential diffusion of responsibility. Individuals should be

assigned responsibility for key aspects of the monitoring of AI tools including quality, compliance, safety and approval.[8]

Finally, there was consensus in the literature that services and service providers must comply with new and existing legislation and policies when implementing AI. The EU AI Act is the EU's regulatory framework for governing AI tools and addressing risk associated with their design, roll-out and use. It is important that services meet their obligations and comply with the EU AI Act and other relevant national and international legislation, regulations, policies, standards and guidelines.[57,71,96,99]

4.2.2 A human rights-based approach

HIQA conceptualisation

A human rights-based approach as defined by HIQA involves 'respecting, protecting and promoting the human rights of the person receiving care and support at all times'.[104] Four themes were primarily categorised under the principle of a human rights-based approach: Privacy, Human connection, Inclusivity and non-discrimination, and Upholding people's rights. Education, training, development and information provision, Transparency, and Safe care were also sub-categorised as relevant to the principle of a human rights-based approach.

Mapping HIQA conceptualisation to evidence from this review

The reviewed evidence puts forward several suggestions for ensuring a human rights-based approach when using AI in health and social care services. Key findings highlighted that use of AI tools should be fair and equitable, without discrimination or favour to any individual or group.[57,67,91,99] Indeed, the findings highlighted that AI tools should be used to minimise inequities and should not worsen existing health inequalities.[67,74,91,99] The evidence emphasised that there should be a focus on minimisation of bias when AI tools are used, with the value of representative data underlined across the literature.[6,57,63,64,70] Results indicated that transparency around data, as well as regular monitoring and auditing, can contribute towards the minimisation of bias.[17,67,86] At a practice level, the evidence suggested that staff should interpret AI outputs through the lens of equity, diversity and inclusion.[82]

The review also highlighted that the accessibility and acceptability of AI tools are necessary to meet diverse needs, with emphasis on inclusive and user-friendly design.[6,57] Engagement with diverse stakeholders and consultation with people who use services were recommended in order to meet various needs when AI tools are used.[77,79,81]

Protection of privacy and safeguarding of personal information were evident in the findings as important considerations for the use of AI tools, particularly given the

potentially sensitive nature of health information.[57,60,62] The review identified that there should be measures in place to protect the right to privacy of people using services, including secure data storage, data minimisation, and compliance with relevant privacy and security regulations.[63,67,74,86,96,101]

This review also identified the significance of human empathy and personal connection in the delivery of care, highlighting that such features of care need to be maintained when AI tools are introduced.[21,73,102] Moreover, the findings point to the need to maintain person-centred values when interpreting recommendations from AI tools, ensuring that decisions informed by AI are aligned with the values and preferences of the person using the service.[21,73]

Finally, results from this review point to several implications for individual autonomy and informed decision-making, as relevant to a human rights-based approach. Respecting autonomy and protecting fundamental principles of healthcare were highlighted among considerations for the use of AI in services.[17,59,62,86,87,91] Regarding informed decisions, the evidence suggests this can be facilitated by the provision of accessible information to people who use services, including information about limitations of AI tools and alternative options.[96] The findings indicate that being transparent with people who use services about the use of AI can support informed decisions,[21,63,96] while transparency also facilitates the promotion of fairness and minimisation of bias in AI tools.[80]

4.2.3 Safety and wellbeing

HIQA conceptualisation

The principle of safety and wellbeing, as defined by HIQA, "refers to how health and social care services work to protect and promote the safety and well-being of people who use services".[104] Two themes were primarily categorised under the principle of safety and wellbeing: Safe care and Technical robustness and security. Transparency, Data quality, Human Connection, and Privacy were also sub-categorised as relevant to the principle of safety and wellbeing.

Mapping HIQA conceptualisation to evidence from this review

The majority of the sources included in the review highlighted the importance of ensuring safe care and mitigating harm when AI is used in the delivery of care. Health and social care services have responsibility to be alert to safety and to respond to any concerns in a person-centred way.[104]

As outlined by Alelyni, safe care ensures no harm or risk from the use of an AI system.[57] Due to the nature of AI, it introduces new and novel challenges that must be considered by services, some of which were addressed in the review, including unexplainable outputs, hallucinations, and risks related to incorrect or over

diagnosis.[89] The review identified measures that should be in place in a service to protect and promote service user safety, including updating safety frameworks and protocols to incorporate the use of AI, human oversight, risk assessment and mitigation, regular monitoring and evaluation of the performance of the tool, and adherence to regulatory standards for safe, accurate and effective use.[6,9,17,70,81,91,95,98]

The review also emphasised the importance of AI systems being secure, robust, and capable of maintaining performance under varying conditions. The evidence highlighted circumstances whereby the security and robustness of an AI system may be challenged including cyberattacks, unauthorised access, data breaches, data manipulation and disruption to critical infrastructure.[57,96] In order to manage and mitigate the aforementioned risk, the literature proposes that services have processes in place including implementing data security plans, adherence to clinical safety guidelines, updating security measures as appropriate, and developing, evaluating and optimising AI systems for robustness against variations in data. It is critical that services have contingency plans in place in the event of an AI risk or failure.[7,15,57,74,91,96,102]

4.2.4 Responsiveness

HIQA conceptualisation

Responsiveness as defined by HIQA relates to how health and social care services are "are organised to deliver coordinated care and support that meets the needs of people using their service".[104] Four themes were primarily categorised under the principle of responsiveness, including Data quality, Integration into care, and Education, training, development and information provision. Transparency and Technical robustness and security were also sub-categorised as relevant to the principle of responsiveness.

Mapping HIQA conceptualisation to evidence from this review

Key to being a responsive service is having staff with the competency, skills and training to provide the highest quality care and support.[104] AI is a new and evolving area and the consensus in the literature is that staff require support, information, training and education prior to using AI tools. Interviews with professionals highlighted how staff want an AI literacy programme to build their competency using AI tools.[73] From the review of the evidence, staff need to know the purpose and use of the AI tool, how to critically evaluate outputs from AI tools, when to use an AI tool in a clinical context, benefits and risks of the AI tool and how data will be managed and stored.[60,73,96]

There was some concern in the literature that clinicians could become dependent on AI and there is a risk that future workforces could become deskilled. Therefore, the

research highlighted that it is important to train staff at all levels and specialities, not just those providing clinical care, and to ensure that the training is relevant to their role and responsibility. Suggestions also included integrating AI training into medical training programmes alongside providing continuous education and development.[60,91]

While the education and training of professionals was a key theme in the review, public education and information was equally prominent in the literature. It is important that staff have the knowledge to provide information to people using services in a clear, accessible and comprehensive manner.[96] The European Patients' Forum outlined how it is important for people using services to have knowledge and to be provided with information regarding the use of AI in their care to understand and have confidence in the use of AI tools and to address any concerns they may have.[6,89] Moreover, the findings highlighted the importance of public education and information to empower people to exercise their rights and to make informed decisions regarding the use of AI.[61,91,92,94,96,98]

A responsive service puts the needs of service users first, is well-coordinated, and works towards a goal of achieving the best possible outcomes for people receiving care.[104] The evidence from this review instils the importance of ensuring the seamless integration of AI tools into existing clinical and operational workflows to reduce impact on service delivery. The evidence suggests that easy-to-use, straightforward AI systems that are interoperable with existing systems should be prioritised.[7,86,92] The review also emphasises that AI tools with the highest clinical relevance should be prioritised, and that AI tools should add value and improve the delivery of care compared to more traditional methods.[57,71,96,101]

In order to ensure that an AI tool will have the best possible outcome for people receiving care, a key finding from this review is the importance of ensuring AI systems have been developed using high-quality data. There is clear consensus in the literature that the data that the AI is developed on should be reliable, accurate, standardised and unbiased as poor quality data can lead to errors or harmful outcomes.[6,70,101]

4.2.5 Mapping to HIQA principles: Conclusion

This mapping exercise revealed that the wide range of concepts identified in the international evidence can be classified according to pre-established, overarching principles of relevance to the Irish health and social care system. Clear and consistent guidance underpinned by these pre-established principles would ensure practical utility across diverse settings and stakeholder groups. Utilising pre-established overarching principles relevant to the Irish context offers a meaningful way to anchor international best practices to the Irish health and social care system.

4.3. Limitations and strengths

This evidence review was undertaken to inform the development of national guidance. While established guidelines for high-quality evidence reviews were closely adhered to,[56,106-109] due to urgency in the system for this guidance to be developed some attenuation to the methodology was needed. This included: searching only one academic database, albeit the most comprehensive and relevant; a targeted rather than expansive search of reputable grey literature sources, meaning that documents and perspectives from less prominent or emerging organisations may have been inadvertently excluded; and minimal citation chaining. The search was also constrained by the accessibility of sources, which may have led to the omission of more recent, unpublished sources or sources that were not available open access. Sources were limited to those published in the English language, which may have introduced geographical bias. Registration of a review protocol would have strengthened adherence to best practice with regard to open science, transparency, replicability, and reproducibility.

Strengths of the review include the systematic and robust methodology that was followed and reported, which ensures transparency and reproducibility. The diverse variety of sources included allowed for a triangulation of perspectives including people who use services, people who work in services, and AI developers. The use of Covidence and NVivo software resulted in a traceable audit trail of decisions made throughout the review. The analyses were conducted rigorously and comprehensively. The involvement of multiple reviewers during screening, extraction and coding, and the engagement of the wider project team at key stages of the analysis, including involvement of a librarian, also strengthens the review.

4.4. Gaps and future research

This evidence review identified several gaps in the literature, as well as avenues for future research. Firstly, the majority of evidence identified through this review included theoretical assertions regarding the responsible and safe use of AI in healthcare. There was a dearth of evidence-based testing of these concepts for effectiveness, practical utility, and usefulness to people using services and staff working in services. Future research should validate these concepts. Secondly, while identification and synthesis of international evidence provides a strong foundation, national guidance should be co-developed with people using and working in health and social care services. This would enable the translation of the concepts and guidance into implementable practices, and determine their feasibility and acceptability in real-world settings. Related to this, there was a lack of empirical research and or discussion regarding the suitability of the concepts to the needs of vulnerable or marginalised groups, and whether there may be unique disadvantages to incorporating such concepts for groups with protected characteristics. Further

research is required to examine how AI systems affect individuals and groups who may be vulnerable in the Irish context. Engagement with individuals with lived experience and advocates would ensure any future guidance addresses the diverse needs of the Irish population.

4.5. Conclusion and next steps

This evidence review highlights concepts and guidance to facilitate the responsible and safe use of AI in health and social care. Through synthesis of 71 concepts identified across 55 international evidence sources, generation of 12 unique themes, and alignment of these themes with established principles for person-centred care and support, the review provides a valuable, evidence-based foundation for developing national guidance for the responsible and safe use of AI in health and social care in Ireland. Overall, the findings demonstrate the range of distinct but interconnected factors that need to be considered to guide the responsible and safe use of AI in services, from transparency, human agency and oversight, and responsibility, to privacy, inclusivity and non-discrimination, upholding rights and human connection, as well as safety, robustness and security, data quality, integration into care, and education, training, development and information.

The review forms one part of an evidence-based and collaborative process to develop *National Guidance for the Responsible and Safe Use of AI in Health and Social Care Services* in Ireland. While the reviewed sources offer useful guidance and recommendations from international settings, further engagement with those who use and work in health and social care services in Ireland will be key to ensuring the national guidance is developed in a way that reflects real-world complexities and diverse needs of people who use services. As part of this process, HIQA engaged with relevant stakeholders through a steering group, a co-production working group, a public scoping consultation, focus groups and interviews to inform the development of the draft guidance. HIQA is also undertaking a public consultation on the draft guidance. The findings from this evidence review, together with insights gathered from stakeholders in the Irish health and social care context, will inform the development of the National Guidance.

Key terms used in the evidence review

AI lifecycle: The series of stages an AI tool goes through, from initial design and development to deployment and monitoring.

Artificial Intelligence (AI; AI tool): A machine-based system capable of operating autonomously and producing outputs like predictions, recommendations, or decisions based on input data.

Black Box AI: An AI tool that can be viewed in terms of its inputs and outputs without any knowledge of its internal workings.

Deep learning: A subfield of machine learning that uses multi-layered artificial neural networks to learn patterns within datasets.

Deployer: As per the AI Act, a deployer is defined as "any natural or legal person, including a public authority, agency or other body, using an AI system under its authority, except where the AI system is used in the course of a personal non-professional activity".

Deployment: A stage in the AI lifecycle where a tool is integrated into real-world environments, making it operational for users.

Generative AI: A type of AI that can create new content, such as text, images, or videos, by learning patterns and structures from large amounts of data. It differs from natural language processing in that it can create new content, not just analyse or understand existing data. Examples include chatbots such as ChatGPT.

Machine learning: A sub-field of AI which focuses on development of tools that are able to learn and adapt without following explicit instructions, imitating the way that humans learn, gradually improving their accuracy, by using algorithms and statistical models to analyse and draw inferences from patterns in data.

Narrow AI: AI tools designed for specific tasks or domains rather than general reasoning or learning across domains, for example, a voice assistant like Siri.

Natural Language Processing: A sub-field of AI that helps computers understand, interpret and use human language. It enables computers to read, write, and interpret text or speech in a way that makes sense to people. It can be used to transcribe clinician notes.

Training data: The data required to train, or "teach", a machine learning algorithm when developing a model.

5. References

1. Department of Public Expenditure, Infrastructure, Public Service Reform and Digitalisation. Guidelines for the Responsible Use of AI in the Public Service. 2025. Available from: https://www.gov.ie/en/department-of-public-expenditure-infrastructure-public-service-reform-and-digitalisation/publications/guidelines-for-the-responsible-use-of-ai-in-the-public-service/

2. National Institute for Health and Care Research. Artificial intelligence: 10 promising interventions for healthcare. 2023. Available from: https://evidence.nihr.ac.uk/collection/artificial-intelligence-10-promising-interventions-for-healthcare/

3. Botha, N. N., Segbedzi, C. E., Dumahasi, V. K., Maneen, S., Kodom, R. V., Tsedze, I. S., et al. Artificial intelligence in healthcare: a scoping review of perceived threats to patient rights and safety. Archives of Public Health. 2024;82(1):188. https://doi.org/10.1186/s13690-024-01414-1

4. Karimian, G., Petelos, E. & Evers, S. M. A. A. The ethical issues of the application of artificial intelligence in healthcare: a systematic scoping review. AI and Ethics. 2022;2(4):539-51. https://doi.org/10.1007/s43681-021-00131-7

5. European Parliamentary Research Service. Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts. 2022. Available from: https://www.europarl.europa.eu/stoa/en/document/EPRS_STU(2022)729512

6. European Patients' Forum. Artificial Intelligence in Healthcare: Advancing patient-centric care through co-design and responsible implementation. 2023. Available from: https://www.eu-patient.eu/news/latest-epf-news/2023/artificial-intai-in-healthcare-advancing-patient-centric-care-through-co-design-and-responsible-implementation/#:~:text=This%20position%20paper%20explores%20the%20applications%2C%20benefits%2C%20and,recommendations%20for%20a%20responsible%20deployment%20of%20AI%20solutions.

7. Lekadir, K., Frangi, A. F., Porras, A. R., Glocker, B., Cintas, C., Langlotz, C. P., et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. BMJ. 2025;388:e081554. https://doi.org/10.1136/bmj-2024-081554

8. MedTech Europe. Trustworthy Artificial Intelligence in healthcare. 2019. Available from: https://www.medtecheurope.org/resource-library/medtech-europe-views-on-the-artificial-intelligence-act/

9. Organisation for Economic Co-operation and Development (OECD). OECD AI Principles. 2024. Available from: https://www.oecd.org/en/topics/sub-issues/ai-principles.html

10. HOPE - European Hospital and Healthcare Federation. HOPE Position Paper on Artificial Intelligence. 2021. Available from: https://dev.hospitalhealthcare.com/wp-content/uploads/2023/01/Position-on-Artificial-Intelligence.pdf

11. Artificial Intelligence Act (Regulation (EU) 2024/1689). European Union. 2024. Available from: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.

12. El Morr, C., Ozdemir, D., Asdaah, Y., Saab, A., El-Lahib, Y. & Sokhn, E. S. AI-based epidemic and pandemic early warning systems: A systematic scoping review. Health Informatics Journal. 2024;30(3):14604582241275844. https://doi.org/10.1177/14604582241275844

13. Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K. & Tekade, R. K. Artificial intelligence in drug discovery and development. Drug Discovery Today. 2021;26(1):80-93. https://doi.org/10.1016/j.drudis.2020.10.010

14. Yim, D., Khuntia, J., Parameswaran, V. & Meyers, A. Preliminary Evidence of the Use of Generative AI in Health Care Clinical Services: Systematic Narrative Review. JMIR Med Inform. 2024;12:e52073. https://doi.org/10.2196/52073

15. Saenz, A. D., Centi, A., Ting, D., You, J. G., Landman, A. & Mishuris, R. G. Establishing responsible use of AI guidelines: a comprehensive case study for healthcare institutions. npj Digital Medicine. 2024;7(1):348. https://doi.org/10.1038/s41746-024-01300-8

16. Lee, Y. B., Kim, G., Jun, J. E., Park, H., Lee, W. J., Hwang, Y. C., et al. An Integrated Digital Health Care Platform for Diabetes Management With AI-Based Dietary Management: 48-Week Results From a Randomized Controlled Trial. Diabetes Care. 2023;46(5):959-66. https://doi.org/10.2337/dc22-1929

17. Jha, D., Durak, G., Sharma, V., Keles, E., Cicek, V., Zhang, Z., et al. A conceptual framework for applying ethical principles of AI to medical practice. Bioengineering. 2025;12(2):180. https://doi.org/10.3390/bioengineering12020180

18. Garriga, R., Mas, J., Abraha, S., Nolan, J., Harrison, O., Tadros, G., et al. Machine learning model to predict mental health crises from electronic health records. Nature Medicine. 2022;28(6):1240-8. https://doi.org/10.1038/s41591-022-01811-5

19. Conway SPHERE Research Group. AI_PREMie. (no date). Accessed Available from: https://www.ucd.ie/conwaysphere/research/ourresearch/aipremie/

20. Gawande, M. S., Zade, N., Kumar, P., Gundewar, S., Weerarathna, I. N. & Verma, P. The role of artificial intelligence in pandemic responses: from epidemiological modeling to vaccine development. Molecular Biomedicine. 2025;6(1):1. https://doi.org/10.1186/s43556-024-00238-3

21. Savulescu, J., Giubilini, A., Vandersluis, R. & Mishra, A. Ethics of artificial intelligence in medicine. Singapore Medical Journal. 2024;65(3):150-8. https://doi.org/10.4103/singaporemedj.SMJ-2023-279

22. Cambridge Cancer Research Hospital. Speeding up breast cancer diagnosis using AI. 2023. Available from: https://www.cambridgecancer.org.uk/news/speeding-up-breast-cancer-diagnosis-using-ai/

23. Randhawa, G., Ferreyra, M., Ahmed, R., Ezzat, O. & Pottie, K. Using machine translation in clinical practice. Canadian Family Physician. 2013;59(4):382-3.

24. Atee, M., Hoti, K., Parsons, R. & Hughes, J. A novel pain assessment tool incorporating automated facial analysis: interrater reliability in advanced dementia. Clinical Interventions in Aging. 2018(13):1245-58. https://doi.org/10.2147/CIA.S168024

25. Atee, M., Hoti, K. & Hughes, Jeffery D. Psychometric Evaluation of the Electronic Pain Assessment Tool: An Innovative Instrument for Individuals with Moderate-to-Severe Dementia. Dementia and Geriatric Cognitive Disorders. 2018;44(5-6):256-67. https://doi.org/10.1159/000485377

26. Zargarzadeh, S., Mirzaei, M., Ou, Y. & Tavakoli, M. From Decision to Action in Surgical Autonomy: Multi-Modal Large Language Models for Robot-Assisted Blood Suction. ArXiv. 2024:2408.07806v2. https://doi.org/10.48550/arXiv.2408.07806

27. Sørensen, L., Johannesen, D. T. & Johnsen, H. M. Humanoid robots for assisting people with physical disabilities in activities of daily living: A scoping review. Assistive Technology. 2025;37(3):203-19. https://doi.org/10.1080/10400435.2024.2337194

28. Subramaniam, S., Faisal, A. I. & Deen, M. J. Wearable Sensor Systems for Fall Risk Assessment: A Review. Frontiers in Digital Health. 2022;Volume 4 - 2022. https://doi.org/10.3389/fdgth.2022.921506

29. Health Service Executive. Digital Health Strategic Implementation Roadmap. 2024. Available from: https://about.hse.ie/api/v2/download-file/file_based_publications/Digital_Health_Strategic_Implementation_Roadmap.pdf/

30. Department of Health. Digital for Care: A Digital Health Framework for Ireland 2024-2030. 2024. Available from: https://www.gov.ie/en/department-of-health/publications/digital-for-care-a-digital-health-framework-for-ireland-2024-2030/

31. Baldassarre, M. T., Caivano, D., Fernández Nieto, B., Gigante, D. & Ragone, A. Fostering Human Rights in Responsible AI: A Systematic Review for Best Practices in Industry. IEEE Transactions on Artificial Intelligence. 2025;6(2):416-31. https://doi.org/10.1109/TAI.2024.3394389

32. Khan, A. A., Badshah, S., Liang, P., Khan, B., Waseem, M., Niazi, M., et al. Ethics of AI: A Systematic Literature Review of Principles and Challenges. Proceedings of

the 26th International Conference on Evaluation and Assessment in Software Engineering. 2021:383-92. https://doi.org/10.1145/3530019.3531329

33. Batool, A., Zowaghi, D, Bano, M. Responsible AI Governance: A Systematic Literature Review. ArXiv. 2023:2401.10896 https://doi.org/10.48550/arXiv.2401.10896

34. Goellner, S., Tropmann-Frick, M. & Brumen, B. Responsible Artificial Intelligence: A Structured Literature Review. ArXiv. 2024:2403.06910. https://doi.org/10.48550/arXiv.2403.06910

35. Mohsin Khan, M., Shah, N., Shaikh, N., Thabet, A., Alrabayah, T. & Belkhair, S. Towards secure and trusted AI in healthcare: A systematic review of emerging innovations and ethical challenges. International Journal of Medical Informatics. 2025;195:105780. https://doi.org/10.1016/j.ijmedinf.2024.105780

36. Ayorinde, A., Mensah, D. O., Walsh, J., Ghosh, I., Ibrahim, S. A., Hogg, J., et al. Health Care Professionals' Experience of Using AI: Systematic Review With Narrative Synthesis. Journal of Medical Internet Research. 2024;26:e55766. https://doi.org/10.2196/55766

37. Ahmed, M. I., Spooner, B., Isherwood, J., Lane, M., Orrock, E. & Dennison, A. A Systematic Review of the Barriers to the Implementation of Artificial Intelligence in Healthcare. Cureus. 2023;15(10):e46454. https://doi.org/10.7759/cureus.46454

38. Nasarian, E., Alizadehsani, R., Acharya, U. R. & Tsui, K.-L. Designing Interpretable ML System to Enhance Trustworthy AI in Healthcare: A Systematic Review of the Last Decade to A Proposed Robust Framework (Preprint Version 2). Research Square. 2023. https://doi.org/10.21203/rs.3.rs-3626164/v2

39. Trocin, C., Mikalef, P., Papamitsiou, Z. & Conboy, K. Responsible AI for Digital Health: a Synthesis and a Research Agenda. Information Systems Frontiers. 2023;25(6):2139-57. https://doi.org/10.1007/s10796-021-10146-4

40. Chustecki, M. Benefits and Risks of AI in Health Care: Narrative Review. Interactive Journal of Medical Research. 2024;13:e53616. https://doi.org/10.2196/53616

41. Siala, H. & Wang, Y. SHIFTing artificial intelligence to be responsible in healthcare: A systematic review. Social Science & Medicine. 2022;296:114782. https://doi.org/10.1016/j.socscimed.2022.114782

42. Tang, L., Li, J. & Fantus, S. Medical artificial intelligence ethics: A systematic review of empirical studies. DIGITAL HEALTH. 2023;9:20552076231186064. https://doi.org/10.1177/20552076231186064

43. Chen, F., Wang, L., Hong, J., Jiang, J. & Zhou, L. Unmasking bias in artificial intelligence: a systematic review of bias detection and mitigation strategies in

electronic health record-based models. Journal of the American Medical Informatics Association. 2024;31(5):1172-83. https://doi.org/10.1093/jamia/ocae060

44. d'Elia, A., Gabbay, M., Rodgers, S., Kierans, C., Jones, E., Durrani, I., et al. Artificial intelligence and health inequities in primary care: a systematic scoping review and framework. Family Medicine and Community Health. 2022;10(Suppl 1):e001670. https://doi.org/10.1136/fmch-2022-001670

45. Marko, J. G. O., Neagu, C. D. & Anand, P. B. Examining inclusivity: the use of AI and diverse populations in health and social care: a systematic review. BMC Medical Informatics and Decision Making. 2025;25(1):57. https://doi.org/10.1186/s12911-025-02884-1

46. Department of Enterprise, Trade and Employment. AI - Here for Good: A National Artificial Intelligence Strategy for Ireland. 2021. Available from: https://enterprise.gov.ie/en/publications/national-ai-strategy.html

47. Department of Enterprise, Trade and Employment. AI - Here for Good: Progress Report on the National AI Strategy. 2023. Available from: https://www.gov.ie/en/department-of-enterprise-tourism-and-employment/publications/progress-report-on-implementation-of-the-national-ai-strategy-ai-here-for-good/

48. European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC. Study on the deployment of AI in healthcare – Final report. 2025. Available from: https://doi.org/10.2875/2169577

49. Oxford Languages. Guidance. Accessed Available from: https://www.bing.com/search?qs=HS&pq=guidance+&sk=CSYN1MT11AS2&sc=18-9&q=guidance+definition&cvid=c410f3c2feb94db89c48b5b0f0f0735b&gs_lcrp=EgRlZGdlKgkIABBFGDsY-QcyCQgAEEUYOxj5BzIGCAEQABhAMgYIAhBFGDkyBggDEAAYQDIGCAQQABhAMgYIBRAAGEAyBggGEEUYPDIGCAcQRRg8MgYICBBFGD0yCAgJEOkHGPIHMggIChDpBxj8VdIBCDE3MzJqMGo0qAIIsAIB&FORM=ANAB01&PC=U531

50. Cambridge Dictionary. Meaning of concept in English. Accessed Available from: https://dictionary.cambridge.org/dictionary/english/concept

51. World Health Organization. Ethics and Governance of Artificial Intelligence for health. 2021. Available from: https://iris.who.int/bitstream/handle/10665/341996/9789240029200-eng.pdf?sequence=1&isAllowed=y

52. Capacity4dev. Design, Results and Reporting. 2025. Available from: https://capacity4dev.europa.eu/groups/design-results-and-reporting/info/ec-official-guidelines_en

53. European Commission: Directorate-General for Communications Networks, Content and Technology and High-Level Expert Group on Artificial Intelligence.

Ethics Guidelines for Trustworthy AI. 2019. Available from: https://data.europa.eu/doi/10.2759/346720

54. McGowan, J., Sampson, M., Salzwedel, D. M., Cogo, E., Foerster, V. & Lefebvre, C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. Journal of Clinical Epidemiology. 2016;75:40-6. https://doi.org/10.1016/j.jclinepi.2016.01.021

55. Pawliuk, C., Brown, H. L., Widger, K., Dewan, T., Hermansen, A. M., Grégoire, M. C., et al. Optimising the process for conducting scoping reviews. BMJ Evididence Based Medicine. 2021;26(6):312. https://doi.org/10.1136/bmjebm-2020-111452

56. Pollock, D., Peters, M. D. J., Khalil, H., McInerney, P., Alexander, L., Tricco, A. C., et al. Recommendations for the extraction, analysis, and presentation of results in scoping reviews. JBI Evidence Synthesis. 2023;21(3). 10.11124/JBIES-22-00123

57. Alelyani, T. Establishing trust in artificial intelligence-driven autonomous healthcare systems: an expert-guided framework. Frontiers in Digital Health. 2024;6. https://doi.org/10.3389/fdgth.2024.1474692

58. Arbelaez Ossa, L., Milford, S. R., Rost, M., Leist, A. K., Shaw, D. M. & Elger, B. S. AI through ethical lenses: a discourse analysis of guidelines for AI in healthcare. Science and Engineering Ethics. 2024;30(3):24. https://doi.org/10.1007/s11948-024-00486-0

59. Armitage, R. C. Implications of large language models for clinical practice: ethical analysis through the Principlism Framework. Journal of Evaluation in Clinical Practice. 2025;31(1):e14250. https://doi.org/10.1111/jep.14250

60. Corfmat, M., Martineau, J. T. & Régis, C. High-reward, high-risk technologies? An ethical and legal account of AI development in healthcare. BMC Medical Ethics. 2025;26(1):4. https://doi.org/10.1186/s12910-024-01158-1

61. Drabiak, K. Leveraging law and ethics to promote safe and reliable AI/ML in healthcare. Frontiers in Nuclear Medicine. 2022;2:983340. https://doi.org/10.3389/fnume.2022.983340

62. Elendu, C., Amaechi, D. C., Elendu, T. C., Jingwa, K. A., Okoye, O. K., John Okah, M., et al. Ethical implications of AI and robotics in healthcare: A review. Medicine (Baltimore). 2023;102(50):e36671. https://doi.org/10.1097/md.0000000000036671

63. Harishbhai Tilala, M., Kumar Chenchala, P., Choppadandi, A., Kaur, J., Naguri, S., Saoji, R., et al. Ethical considerations in the use of Artificial Intelligence and machine learning in health care: a comprehensive review. Cureus. 2024;16(6):e62443. https://doi.org/10.7759/cureus.62443

64. Jeyaraman, M., Balaji, S., Jeyaraman, N. & Yadav, S. Unravelling the ethical enigma: Artificial Intelligence in healthcare. Cureus. 2023;15(8):e43262. https://doi.org/10.7759/cureus.43262

65. Kim, M., Sohn, H., Choi, S. & Kim, S. Requirements for trustworthy Artificial Intelligence and its application in healthcare. Healthcare Informatics Research. 2023;29(4):315-22. https://doi.org/10.4258/hir.2023.29.4.315

66. Maccaro, A., Stokes, K., Statham, L., He, L., Williams, A., Pecchia, L., et al. Clearing the fog: a scoping literature review on the ethical issues surrounding Artificial Intelligence-based medical devices. Journal of Personalised Medicine. 2024;14(5):443. https://doi.org/10.3390/jpm14050443

67. Marques, M., Almeida, A. & Pereira, H. The medicine revolution through Artificial Intelligence: ethical challenges of machine learning algorithms in decision-making. Cureus. 2024;16(9):e69405. https://doi.org/10.7759/cureus.69405

68. Mennella, C., Maniscalco, U., De Pietro, G. & Esposito, M. Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. Heliyon. 2024;10(4):e26297. https://doi.org/10.1016/j.heliyon.2024.e26297

69. Prakash, S., Balaji, J. N., Joshi, A. & Surapaneni, K. M. Ethical conundrums in the application of Artificial Intelligence (AI) in healthcare: a scoping review of reviews. Journal of Personalised Medicine. 2022;12(11):1914. https://doi.org/10.3390/jpm12111914

70. Reddy, S., Rogers, W., Makinen, V.-P., Coiera, E., Brown, P., Wenzel, M., et al. Evaluation framework to guide implementation of AI systems into healthcare settings. BMJ Health & Care Informatics. 2021;28(1):e100444. https://doi.org/10.1136/bmjhci-2021-100444

71. Sousa-Pinto, B., Marques-Cruz, M., Neumann, I., Chi, Y., Nowak, A. J., Reinap, M., et al. Guidelines International Network: Principles for use of Artificial Intelligence in the health guideline enterprise. Annals of Internal Medicine. 2025;178(3):408-15. https://doi.org/10.7326/annals-24-02338

72. Vandemeulebroucke, T. The ethics of artificial intelligence systems in healthcare and medicine: from a local to a global perspective, and back. Pflugers Archiv - European Journal of Physiology. 2025;477(4):591-601. https://doi.org/10.1007/s00424-024-02984-3

73. Elgin, C. Y. & Elgin, C. Ethical implications of AI-driven clinical decision support systems on healthcare resource allocation: a qualitative study of healthcare professionals' perspectives. BMC Medical Ethics. 2024;25(1):148. https://doi.org/10.1186/s12910-024-01151-8

74. Kahraman, F., Aktas, A., Bayrakceken, S., Çakar, T., Tarcan, H. S., Bayram, B., et al. Physicians' ethical concerns about artificial intelligence in medicine: a

qualitative study: "The final decision should rest with a human". Frontiers in Public Health. 2024;12:1428396. https://doi.org/10.3389/fpubh.2024.1428396

75. Katirai, A. The ethics of advancing artificial intelligence in healthcare: analyzing ethical considerations for Japan's innovative AI hospital system. Frontiers in Public Health. 2023;11:1142062. https://doi.org/10.3389/fpubh.2023.1142062

76. Zuchowski, L. C., Zuchowski, M. L. & Nagel, E. A trust based framework for the envelopment of medical AI. npj Digital Medicine. 2024;7(1):230. https://doi.org/10.1038/s41746-024-01224-3

77. Tierney, A. A., Reed, M. E., Grant, R. W., Doo, F. X., Payán, D. D. & Liu, V. X. Health equity in the era of large language models. The American Journal of Managed Care. 2025;31(3):112-7. https://doi.org/10.37765/ajmc.2025.89695

78. Rose, S. L. & Shapiro, D. An ethically supported framework for determining patient notification and informed consent practices when using Artificial Intelligence in health care. CHEST. 2024;166(3):572-8. https://doi.org/10.1016/j.chest.2024.04.014

79. Seroussi, B. & Zablit, I. Implementation of digital health ethics: a first step with the adoption of 16 European ethical principles for digital health. Studies in Health Technology and Informatics. 2024;310:1588-92. https://doi.org/10.3233/shti231331

80. Upadhyay, U., Gradisek, A., Iqbal, U., Dhar, E., Li, Y. C. & Syed-Abdul, S. Call for the responsible artificial intelligence in the healthcare. BMJ Health & Care Informatics. 2023;30(1):e100920. https://doi.org/10.1136/bmjhci-2023-100920

81. Australian Medical Association. Artificial Intelligence in Healthcare. 2023. Available from: https://www.ama.com.au/articles/artificial-intelligence-healthcare

82. College of Physicians and Surgeons of British Columbia. Ethical Principles for Artificial Intelligence in Medicine. 2024. Available from: https://www.cpsbc.ca/files/pdf/IG-Artificial-Intelligence-in-Medicine.pdf

83. College of Physicians and Surgeons of British Columbia. Artificial Intelligence in Diagnostic Services. 2025. Available from: https://www.cpsbc.ca/files/pdf/DAP-PS-Artificial-Intelligence-in-Diagnostic-Services.pdf

84. MedTech Europe. Artificial Intelligence in MedTech: Delivering on the promise of better healthcare in Europe. 2019. Available from: https://www.medtecheurope.org/resource-library/ai-in-medtech-delivering-on-the-promise-of-better-healthcare-in-europe/

85. The Care Workers Charity. Care Workers' Guidance and Statement of Expectations on the Responsible Use of AI and Particularly Generative AI in Adult Social Care. 2024. Available from: https://www.thecareworkerscharity.org.uk/wp-content/uploads/2024/09/Careworkers-guidance-and-statement-of-expectations-on-the-responsible-use-of-AI-and-particularly-generative-AI-in-adult-social-care.pdf

86. The Royal Australian and New Zealand College of Radiologists. Ethical Principles for Artificial Intelligence for Medicine Version 2. 2023. Available from: https://www.ranzcr.com/whats-on/news-media/ethical-principles-for-artificial-intelligence-for-medicine-version-2?searchword=Artificial%20intelligence

87. The Royal Australian and New Zealand College of Radiologists. Autonomous Artifical Intelligence Position Statement. 2024. Available from: https://www.ranzcr.com/college/document-library/autonomous-ai-position-paper

88. The Royal Australian and New Zealand College of Radiologists. Generative Artificial Intelligence and Large Language Models. 2024. Available from: https://www.ranzcr.com/college/document-library/generative-ai?searchword=Artificial%20intelligence

89. European Patients' Forum. Public consultation on the White Paper on Artificial Intelligence. 2020. Available from: https://www.eu-patient.eu/globalassets/documents/1.-ai-white-paper_consultation-response_epf_statement-final.pdf

90. European Patients' Forum. Artificial Intelligence in Healthcare from a Patients Perspective. 2022. Available from: https://www.eu-patient.eu/globalassets/report-ai-1612---del-castillo-and-nicholas.pdf

91. Irish Platform for Patient Organisations, Science and Industry (IPPOSI). Citizen's Jury: Artificial Intelligence in healthcare in Ireland. 2025. Available from: https://ipposi.ie/wp-content/uploads/2025/02/Citizens-Jury-Verdict-Report-21.02.25.pdf

92. Medical Protection Society (MPS) Foundation. Avoiding the AI 'Off-Switch'. 2025. Available from: https://www.thempsfoundation.org//docs/foundationlibraries/foundation-default-library/white-papers/ai-white-paper_the-mps-foundation.pdf

93. Australian Medical Association. AMA submission to the Therapeutic Goods Administration consultation on clarifying and strengthening the regulation of Artificial Intelligence (AI). 2024. Available from: https://www.ama.com.au/articles/ama-submission-tga-consultation-clarifying-and-strengthening-regulation-artificial

94. Australian Health Practitioner Regulation Agency (Ahpra). Meeting your Professional Obligations when using Artificial Intelligence in Healthcare. 2024. Accessed Available from: https://www.ahpra.gov.au/Resources/Artificial-Intelligence-in-healthcare.aspx

95. Australian Alliance for Artificial Intelligence in Healthcare. A Roadmap for Artificial Intelligence in Healthcare for Australia. 2021. Available from: https://aihealthalliance.org/wp-content/uploads/2021/12/AAAiH_Roadmap_1Dec2021_FINAL.pdf

96. Australian Commission on Safety and Quality in Health Care. AI Implementation in Hospitals: Legislation, Policy, Guidelines and Principles, and Evidence about Quality and Safety. 2024. Available from: https://www.safetyandquality.gov.au/sites/default/files/2025-08/ai-implementation-in-hospitals-legislation-policy-guidelines-and-principles-and-evidence-about-quality-and-safety.pdf

97. Care Quality Commission. Using machine learning in diagnostic services: CQC's regulatory sandbox report. 2020. Available from: https://www.cqc.org.uk/sites/default/files/20200324%20CQC%20sandbox%20report_machine%20learning%20in%20diagnostic%20services.pdf

98. Council of Europe. The impact of artificial intelligence on the doctor-patient relationship. 2022. Available from: https://www.coe.int/en/web/human-rights-and-biomedicine/report-impact-of-ai-on-the-doctor-patient-relationship

99. Council of Europe. The application of artificial intelligence in healthcare and its impact on the patient-doctor relationship. 2024. Available from: https://rm.coe.int/prems-172924-gbr-2007-artificial-intelligence-16x24-web/1680b30e0b

100. European Commission Joint Research Centre. Artificial intelligence for healthcare and well-being during exceptional times. 2023. Available from: https://ai-watch.ec.europa.eu/publications/artificial-intelligence-healthcare-and-well-being-during-exceptional-times_en

101. Irish College of General Practitioners. The Use of Artificial Intelligence in Irish General Practice. 2025. Available from: https://www.irishcollegeofgps.ie/LinkClick.aspx?fileticket=rGiNT9WwkkA%3D&portalid=0

102. National Health Service. Artificial Intelligence: How to get it right. 2019. Available from: https://webarchive.nationalarchives.gov.uk/ukgwa/20241101055112/https://transform.england.nhs.uk/ai-lab/explore-all-resources/understand-ai/artificial-intelligence-how-get-it-right/

103. Thornton, N., Hardie, T., Horton, T. & Gerhold, M. Priorities for an AI in health care strategy. 2024. Available from: https://www.health.org.uk/reports-and-analysis/briefings/priorities-for-an-ai-in-health-care-strategy

104. Health Information and Quality Authority (HIQA). Evidence review to inform the development of a set of principles to underpin future national standards for health and social care services. 2021. Available from: https://www.hiqa.ie/sites/default/files/2021-02/Evidence-review-for-development-of-principles-underpining-future-national-standards.pdf

105. Health Information and Quality Authority (HIQA). Standards Development Framework: a principles-based approach. 2021. Available from:

https://www.hiqa.ie/sites/default/files/2021-09/Standards-Development-Framework-a-principles-based-approach.pdf

106. Peters, M. D. J., Marnie, C., Tricco, A. C., Pollock, D., Munn, Z., Alexander, L., et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evidence Implementation. 2021;19(1):3-10. https://doi.org/10.1097/xeb.0000000000000277

107. Munn, Z., Pollock, D., Khalil, H., Alexander, L., Mclnerney, P., Godfrey, C. M., et al. What are scoping reviews? Providing a formal definition of scoping reviews as a type of evidence synthesis. JBI Evidence Synthesis. 2022;20(4):950-2. https://doi.org/10.11124/jbies-21-00483

108. Pollock, D., Evans, C., Menghao Jia, R., Alexander, L., Pieper, D., Brandão de Moraes, É., et al. "How-to": scoping review? Journal of Clinical Epidemiology. 2024;176:111572. https://doi.org/10.1016/j.jclinepi.2024.111572

109. Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Annals of Internal Medicine. 2018;169(7):467-73. https://doi.org/10.7326/M18-0850

6. Appendices

Appendix 1. Eligibility criteria

	Inclusion Criteria	Exclusion Criteria
Concept	• Guidance, frameworks, principles, related to facilitating at least one of the following types of use of AI: - Safe - Responsible - Ethical - Trustworthy.	• Papers focused on AI use in settings outside of health or social care • Technical development, implementation or deployment of AI tools • Legal advice on how to comply with EU AI Act • Any document that is promotional rather than policy driven • Articles that deal with a specific area of healthcare or a specific illness • Articles that deal specifically with mental health
Context	• AI use in health or social care settings including, but not limited to: hospital care, ambulance services, community care, primary care, and general practice and children's services.	• Papers focused on AI use in settings outside of health or social care settings.

Evidence source	• Peer review Academic articles; Evidence based reviews; policy papers; consensus statements; expert opinion pieces; frameworks; guidance as available in the database Medline. • Grey literature National or international AI health strategies; policy reports; governance frameworks; principles; official guidance/guidelines; position statements from reputable organisations including government bodies, international health organisations, regulatory agencies, professional organisations as available in the websites identified in appendix 3.	• Purely theoretical papers (not focused on concrete guidance, frameworks, governance discussions) • Opinion pieces without policy or regulatory focus; individual hospital policies (unless part of national/international strategies) • Unofficial documents (blog posts, news articles, individual opinion pieces) • Primary sources that are already incorporated into an included evidence review will be excluded.
Timeframe	• Grey literature Date of publication is from 2019 to 30 March 2025. In 2019, the EU Ethical Guidance on Trustworthy AI[53] proposed a set of seven key requirements that AI systems should meet in orders to be deemed trustworthy. This was a key turning point and we feel that literature after the release of this guidance is most relevant to inform a national framework for responsible AI use in the Irish context.	• Date of publication is prior to the dates stated in the inclusion criteria. • The search will be conducted in March 2025 and any literature published after the 31 March will not be included in the review.

	• Peer review Date of publication is from 2022 to 30 March 2025. Literature before 2022 is unlikely to consider the EU Ethical Guidance on Trustworthy AI.[53] Due to the time taken to publish articles in peer-reviewed journals, articles tend to be published approximately a year after the research was undertaken. Thus, we will exclude 2020 as this likely excludes research that predates the release of the EU ethical Guidance on trustworthy AI.[53]	
Geographical scope	• The geographical scope we are considering is global but we will prioritise regions with established AI governance for the grey literature search (EU, UK, Canada, Australia).	
Other criteria	• Full texts must be available online (open access) or through HIQA's library. Text must be available in English.	

Appendix 2. Search strategy

A. Medline Complete search

Databases		Number of results	Date searched	
MEDLINE via Ebscohost		1784	3 March 2025	
#	Query	Limiters/ Expanders	Last Run Via	Results
S10	S4 AND S8	Limiters - Publication Date: 20220101-20251231	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	1784
S9	S4 AND S8		Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	2,221
S8	S5 OR S6 OR S7		Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	566,873
S7	TI (("Responsible" OR "Safe" OR "Ethic*" OR "Trust*" OR "Equit*" OR "Principle*")		Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	241,094
S6	(MH "Trust")		Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	2,575
S5	MH "Ethics+"		Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	159,807

S4	S1 OR S2 OR S3		Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	407,694
S3	AB (("AI" OR "Artificial Intelligenc*" OR "Machine Learning" OR "Deep Learning" OR "Chat GPT" OR "Neural Network") OR TI ("AI" OR "Artificial Intelligenc*" OR "Machine Learning" OR "Deep Learning" OR "Chat GPT" OR "Neural Network") OR ADJ2 (artificial OR intelligen* OR generative OR algorithm OR machine OR learning OR neural OR network*))		Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	308,861
S2	MH "Machine Learning"		Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	86,787
S1	MH "Artificial Intelligence+"		Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	225,799

B. Grey literature search terms

Concept 1: Artificial Intelligence	Concept 2: Ethics	Concept 3: Health & Social Care	Concept 4: Document type
"AI"	Ethic*	"Health Care"	"Framework*"
"Artificial Intelligence"	Trust*	"Social Care"	"Principle*"
"Generative Artificial Intelligence"	Responsibl*	"Health and social care"	"Guidance*"
"Machine Learning"	Princip*	"Delivery of health care"	
"Deep Learning"	Safe*	"Quality of health care"	
"Neural Network"	"Clinical decision making"		
"Large language model"	Accessibl*		
Algorithms			

Appendix 3. Grey literature websites searched

Country	Organisation/body	URL
United Kingdom	Care Quality Commission	https://www.cqc.org.uk/
United Kingdom	NICE National Institute for Health and Care Excellence	https://www.nice.org.uk/
United Kingdom	National Health Service (NHS)	https://transform.england.nhs.uk/ai-lab/
United Kingdom	The Health Foundation	https://www.health.org.uk/
United Kingdom	Care Workers Charity	https://www.thecareworkerscharity.org.uk/
Global	OECD	https://oecd.ai/en/
Global	WHO	https://www.who.int/
Global	International Organisation for Standards (ISO)	https://www.iso.org/home.html
Europe	European Commission	https://commission.europa.eu/index_en
Europe	European Patients Forum	https://www.eu-patient.eu/
Europe	European Hospital and Healthcare Federation	https://hope.be/
Europe	European Medicines Agency (EMA)	https://www.ema.europa.eu/en
Europe	Council of Europe	https://www.coe.int/en/web/portal/
Europe	MedTech	https://www.medtecheurope.org/
Australia	Australian Alliance for Artificial Intelligence in Healthcare	https://aihealthalliance.org/
Australia	Australian Government	https://www.health.gov.au
Australia	Australian Health Practitioner Regulation Agency & National Boards	https://www.ahpra.gov.au/
Australia	Australian Commission on Safety and Quality in Healthcare	https://www.safetyandquality.gov.au/

Australia	Royal Australian and New Zealand College of Radiologists	https://www.ranzcr.com/
Australia	Healthlink	https://www.healthlink.com.au/
Australia	National Health and Medicinal Research Council	https://www.nhmrc.gov.au/
Canada	College of Physicians and Surgeons of British Columbia	https://www.cpsbc.ca/
Canada	Health Canada	https://www.canada.ca/en/health-canada.html
Canada	Canadian Institute for Health Information	https://www.cihi.ca/en

Appendix 4. Characteristics of sources included in the evidence review

A. Academic Sources

#	Title	Author	Year	Location	Aim	Study design and methods	Underpinning framework
1	Establishing trust in artificial intelligence-driven autonomous healthcare systems: an expert-guided framework.	Alelyani	2024	Global	To design a framework to contribute to the advancement of trustworthy assessment practices in the field of AI in healthcare systems.	Qualitative study involving semi-structured interviews with 15 experts in autonomous systems technologies in healthcare	The interview guide was developed based on measures outlined by The National Institute of Standards and Technology
2	Implications of Large Language Models for Clinical Practice: Ethical Analysis Through the Principlism Framework.	Armitage	2024	Global	To consider the ethical implications of LLMs for medical practitioners in their delivery of clinical care through the ethical framework of principlism.	Review article	Principlism framework (Beauchamp and Childress)
3	Ethical implications of AI and robotics in healthcare: A review.	Elendu et al.	2023	Global	To examine and analyse the ethical implications of AI and robotics in healthcare. To provide a comprehensive understanding of the challenges and	Narrative review	-

#	Title	Author	Year	Location	Aim	Study design and methods	Underpinning framework
					opportunities associated with AI.		
4	High-reward, high-risk technologies? An ethical and legal account of AI development in healthcare.	Corfmat et al.	2025	Global	To summarise how to approach the challenges of AI from an ethical and legal perspective. To suggest improvements to help healthcare professionals better navigate the AI wave.	Literature review	-
5	Leveraging law and ethics to promote safe and reliable AI/ML in healthcare.	Drabiak	2022	United States	To describe how the law provides a mechanism to promote safety and reliability of AI systems. To provide an overview of potential areas of liability. To summarise strategies to minimise risk and promote safe and reliable AI.	Review article	Reviews FDA regulations
6	Ethical implications of AI-driven clinical decision support systems on healthcare	Elgin & Elgin	2024	Turkey	To explore healthcare Professionals' perspectives on the ethical implications of using AI Clinical Decision Support Systems for	Qualitative study involving semi-structured interviews with 23 healthcare professionals, including physicians, nurses,	The interview protocol was informed by existing theoretical frameworks on algorithmic ethics in healthcare and contemporary debates

#	Title	Author	Year	Location	Aim	Study design and methods	Underpinning framework
	resource allocation: a qualitative study of healthcare professionals' perspectives.				healthcare resource allocation.	administrators, and medical ethicists	about value considerations in medical AI implementation
7	Unraveling the Ethical Enigma: Artificial Intelligence in Healthcare.	Jeyaraman et al.	2023	Global	To discuss the ethical concerns and difficulties posed related to the use of AI in healthcare, in particular publicly accessible LLMs.	Review article	-
8	A Conceptual Framework for Applying Ethical Principles of AI to Medical Practice	Jha et al.	2025	Global	To examine the ethical implications of Ai based healthcare technologies. To provide structured guidelines for responsible AI deployment.	Conceptual framework	-
9	Physicians' ethical concerns about artificial intelligence in medicine: a qualitative study: "The final	Kahraman et al.	2024	Turkey	To investigate the acceptability of AI in Medicine and to elucidate any technical and scientific, as well as social.	Qualitative study involving semi-structured interviews with 25 medical doctors	-

#	Title	Author	Year	Location	Aim	Study design and methods	Underpinning framework
	decision should rest with a human"						
10	The ethics of advancing artificial intelligence in healthcare: analyzing ethical considerations for Japan's innovative AI hospital system.	Katirai	2023	Japan	To identify the extent to which Japan's Cross-Ministerial Strategic Innovation Promotion Program (SIP) for an "Innovative AI Hospital System" addressed ethical considerations set out in the World Health Organization's 2021 Guidance on the Ethics and Governance of Artificial Intelligence for Health.	Content analysis of a single published document using a framework informed by international guidance	World Health Organization's 2021 Guidance on the Ethics and Governance of Artificial Intelligence for Health
11	Requirements for Trustworthy Artificial Intelligence and its Application in Healthcare	Kim et al.	2023	South Korea	To review the requirements for trustworthy AI and examine the current status of its application and related policy initiatives in healthcare.	Review article	-
12	FUTURE-AI: international consensus	Lekadir et al.	2025	Global	To define international guidelines for trustworthy healthcare AI and deliver	International consensus process using a modified delphi method	-

#	Title	Author	Year	Location	Aim	Study design and methods	Underpinning framework
	guideline for trustworthy and deployable artificial intelligence in healthcare				the first structured and holistic guideline for trustworthy and ethical AI in healthcare.	with 117 stakeholders including clinicians, data scientists, computer engineers, AI scientists, healthcare practitioners, ethicists, social scientists, legal experts, industry professionals, patient advocates, regulatory experts	
13	Clearing the Fog: A Scoping Literature Review on the Ethical Issues Surrounding Artificial Intelligence-Based Medical Devices	Maccaro et al.	2024	Global	To capture the progression of the ethical and legal debate and the proposed ethical frameworks available concerning the use of AI based medical technologies; to produce a coherent ethical framework for AI-based medical technologies.	Scoping review	-
14	The Medicine Revolution Through Artificial Intelligence:	Marques et al.	2024	Global	To discuss the ethical issues of AI algorithms used mainly in data management, diagnosis,	Narrative review	-

#	Title	Author	Year	Location	Aim	Study design and methods	Underpinning framework
	Ethical Challenges of Machine Learning Algorithms in Decision-Making.				intervention, and decision-making processes.		
15	Ethical and regulatory challenges of AI technologies in healthcare: A narrative review.	Mennella et al.	2024	Global	To provide a comprehensive overview of the existing evidence concerning AI technologies, examining both technical aspects and regulatory considerations.	Narrative review	World Health Organization's 2021 Guidance on the Ethics and Governance of Artificial Intelligence for Health; the EU AI Act and other and other cross-sector and health-specific regulations
16	AI Through Ethical Lenses: A Discourse Analysis of Guidelines for AI in Healthcare.	Arbelaez Ossa et al.	2024	Global	To analyse how guidelines construct, articulate, and frame AI ethics for healthcare. To critically interpret these guidelines' underlying ideologies.	Review with discourse analysis of published guidelines	-
17	Ethical Conundrums in the Application of Artificial Intelligence (AI)	Prakash et al.	2022	Global	To ascertain the ethical concerns of AI applications in healthcare, to identify the knowledge gaps and provide recommendations	Scoping review of reviews	-

#	Title	Author	Year	Location	Aim	Study design and methods	Underpinning framework
	in Healthcare-A Scoping Review of Reviews.				for an ethical and legal framework.		
18	Evaluation framework to guide implementation of AI systems into healthcare settings	Reddy et al.	2021	Global	To create a framework that assesses real-world systems of AI in healthcare.	Literature review, consensus approach and expert review by eight-member international panel with expertise in medicine, data science, healthcare policy, biomedical research and healthcare commissioning	Principles for translational research; Health Technology Assessment
19	An Ethically Supported Framework for Determining Patient Notification and Informed Consent Practices When Using Artificial Intelligence in Health Care.	Rose & Shapiro	2024	United States	To provide guidance on when hospital leaders should tell patients about the use of AI in their care.	Case example to test evaluation guidance	-

#	Title	Author	Year	Location	Aim	Study design and methods	Underpinning framework
20	Establishing responsible use of AI guidelines: a comprehensive case study for healthcare institutions.	Saenz et al.	2024	United States	To present a comprehensive case study for the responsible integration of artificial intelligence into healthcare settings. To propose a set of guidelines emphasizing 8 principles.	Expert consensus using delphi-like methodology with 18 people from an integrated academic healthcare system, including leaders from informatics, legal, research, data analytics, privacy, patient experience, equity, quality, and safety; case study.	-
21	Ethics of artificial intelligence in medicine.	Savulescu et al.	2024	Singapore	To review the main ethical issues that arise from the use of AI technologies in medicine.	Review article	-
22	Implementation of Digital Health Ethics: A First Step with the Adoption of 16 European Ethical Principles for Digital Health.	Seroussi & Zablit	2024	Europe	To shape digital health ethics, and define ethical rules and policies.	Discussion paper	Principlism framework (Beauchamp and Childress); digital ethics

#	Title	Author	Year	Location	Aim	Study design and methods	Underpinning framework
23	Guidelines International Network: Principles for Use of Artificial Intelligence in the Health Guideline Enterprise.	Sousa-Pinto et al.	2025	Global	To propose a set of principles for the development and use of AI tools or processes to support the health guideline enterprise.	Scoping review and consensus-based process involving working group with diverse backgrounds in evidence-based healthcare, guideline development, data science, and AI; European Patients Forum were involved to contribute to the suggested principles	-
24	Health equity in the era of large language models.	Tierney et al.	2025	United States	To summarise the promises and challenges of LLMs for health equity.	Summary and grouping paper	Professional, national, and international guidelines to address equity-related challenges in the US
25	Ethical Considerations in the Use of Artificial Intelligence and Machine Learning in Health Care: A	Harishbhai Tilala et al.	2024	Global	To delve into the ethical dimensions of AI and ML in health care, examining the ethical principles that underpin responsible AI deployment, and exploring the ethical dilemmas	Narrative review	Foundational ethical principles that have guided healthcare practice (beneficence, non-maleficence, autonomy, justice)

#	Title	Author	Year	Location	Aim	Study design and methods	Underpinning framework
	Comprehensive Review.				inherent in the use of these technologies.		
26	Call for the responsible artificial intelligence in the healthcare.	Upadhyay et al.	2023	Global	To highlight the complexities and potentials of Ai in healthcare. To emphasise the necessity of 5 principles when developing and implementing AI.	Appeal article	-
27	The ethics of artificial intelligence systems in healthcare and medicine: from a local to a global perspective, and back.	Vandemeule broucke	2024	Global	To outline the global impacts, including environmental and societal impacts, of AI systems in healthcare and medicine. Integrates ethical issues occurring within local health and medical settings with those occurring in social and environmental contexts in which these settings are located.	Review article	Contrasts a local isolationist ethical approach (Principlist approach) with a global approach to the ethics of AI-systems in healthcare and medicine
28	A trust based framework for the envelopment of medical AI.	Zuchowski et al.	2024	Europe	To develop a novel legal, social and regulatory envelopment of medical AI that is explicitly based on	Framework	Draws on the understanding of the relationships between trust and reliance put

#	Title	Author	Year	Location	Aim	Study design and methods	Underpinning framework
					the preservation of trust between patients and medical professionals; develop a framework for the legal, social, and regulatory envelopment of medical AI.		forward by Baier (1986) and McGreer & Pettit (2017)

B. Grey literature sources

#	Title	Author	Year	Location	Type of organisation	Type of source	Aim	Approach/method (if reported)	Relevant underpinning conceptual, legal, theoretical, or policy frameworks
29	A Roadmap for Artificial Intelligence in Healthcare for Australia	Australian Alliance for Artificial Intelligence in Healthcare	2021	Australia	Professional association	Report	To identify current gaps in Australia's capability to translate AI into effective and safe clinical services. To provide guidance on key issues such as workforce, industry capability, implementation, regulation, and cyber security. To provide recommendations across priority areas.	Community consultation, workshops, a national survey with 152 stakeholder organisations and individuals from healthcare, working group consultation, industry advisory group consultation	Builds on the extensive work that has already been undertaken nationally and internationally, including existing national frameworks and policies that relate to AI.
30	AI Implementation in Hospitals: Legislation, Policy, Guidelines and Principles, and Evidence about Quality and Safety	Australian Commission on Safety and Quality in Health Care	2024	Australia	Governmental Agency	Report	To identify principles that enable the responsible and safe implementation of AI in healthcare.	Scoping review and environmental/policy scan	-
31	Meeting your professional obligations when using Artificial Intelligence in healthcare	Australian Health Practitioner Regulation Agency (Ahpra)	2024	Australia	National professional or regulatory body	Webpage	To identify key principles to highlight existing professional obligations that apply when health practitioners use AI in their practice.	-	-

#	Title	Author	Year	Location	Type of organisation	Type of source	Aim	Approach/method (if reported)	Relevant underpinning conceptual, legal, theoretical, or policy frameworks
32	Artificial Intelligence in Healthcare	Australian Medical Association (AMA)	2023	Australia	Professional association	Position paper	To outline the AMA's position on the application of AI and AI tools in healthcare.	-	-
33	AMA submission to the Therapeutic Goods Administration consultation on clarifying and strengthening the regulation of Artificial Intelligence (AI)	Australian Medical Association (AMA)	2024	Australia	Professional association	Submission to consultation	To clarify and strengthen the regulation of AI – a submission by the AMA to Therapeutic Goods Administration.	Review	The AMA supports the TGA's approach to regulating AI products that fit the definition of a medical device under the Therapeutic Goods Act 1989. The AMA supports the TGA's classification for medical devices based on risk, a system that works effectively to promote clinician and patient safety in healthcare. The AMA approves of the Therapeutic Goods Administration's

#	Title	Author	Year	Location	Type of organisation	Type of source	Aim	Approach/method (if reported)	Relevant underpinning conceptual, legal, theoretical, or policy frameworks
									intent to align the legislative and regulatory framework for therapeutic goods with the intent of the Department of Industry Science and Resources's proposed mandatory guardrails for AI
34	Using machine learning in diagnostic services: CQC's regulatory sandbox report	Care Quality Commission (CQC)	2020	United Kingdom	National professional or regulatory body	Report	To present the findings from the Care Quality Commission's regulatory sandbox pilot.	Regulatory sandboxing focusing on the use of machine learning applications for diagnostic purposes in healthcare services, involving seven technology suppliers and their NHS partners who were delivering machine learning applications in diagnostic pathway	-
35	Ethical Principles for Artificial Intelligence in Medicine	College of Physicians and Surgeons of British Columbia	2024	Canada	National professional or regulatory body	Position paper	To provide an interim guidance document to promote the safe, effective, and ethical utilisation of AI tools in	-	-

#	Title	Author	Year	Location	Type of organisation	Type of source	Aim	Approach/method (if reported)	Relevant underpinning conceptual, legal, theoretical, or policy frameworks
							healthcare and protect public safety.		
36	Artificial Intelligence in Diagnostic Services	College of Physicians and Surgeons of British Columbia	2025	Canada	National professional or regulatory body	Position paper	To provide the CPSBC's position on introducing AI in diagnostic services to act as guidance for stakeholders while policies and procedures are being developed.	-	-
37	The impact of artificial intelligence on the doctor-patient relationship	Council of Europe	2022	Europe	Intergovernmental organisation	Report	To outline the impact of AI on the doctor-patient relationship.	-	-
38	The application of artificial intelligence in healthcare and its impact on the patient-doctor relationship	Council of Europe	2024	Europe	Intergovernmental organisation	Report	To support decision makers, health providers, health professionals and patients (including patient associations) to consider how AI systems are used in healthcare, having regard to their human rights implications; develop and strengthen the therapeutic relationship, especially in supporting doctors and, where	Drafting group meetings, expert exchange, consultation	Selected human rights principles of particular relevance to the therapeutic relationship, namely consent (Article 5 of the European Convention on Human Rights and Biomedicine, or the Oviedo Convention), professional

#	Title	Author	Year	Location	Type of organisation	Type of source	Aim	Approach/method (if reported)	Relevant underpinning conceptual, legal, theoretical, or policy frameworks
							appropriate, other healthcare professionals in promoting the agency and autonomy of patients, patient welfare and equitable access to healthcare.		standards (Article 4, Oviedo Convention), private life and right to information (Article 10, Oviedo Convention) and equitable access to healthcare (Article 3, Oviedo Convention)
39	Artificial intelligence for healthcare and well-being during exceptional times - European Commission	European Commission Joint Research Centre	2023	Europe	Other: Research centre	Report	To provide a detailed state of the art of the current and near-future applications of AI in medicine, healthcare and wellbeing.	Evidence review	-
40	Public consultation on the White Paper on Artificial Intelligence: EPF's Response & Accompanying Statement	European Patients Forum	2020	Europe	Patient organisation	Participatory report	To outline EPF's response to the European Commission's white paper on AI consultation.	Consultation process with European Patients Forum (EPF) members and EPF Digital Health Working group	-
41	Artificial Intelligence in Healthcare from a Patient's Perspective	European Patients Forum (EPF)	2022	Europe	Patient organisation	Participatory report	To present findings from explorative research conducted to understand	Engagements including two webinars, micro survey, in-depth interviews with 16-	-

#	Title	Author	Year	Location	Type of organisation	Type of source	Aim	Approach/method (if reported)	Relevant underpinning conceptual, legal, theoretical, or policy frameworks
							the opportunities and challenges of deploying AI in the health sector from a patient perspective.	18 deployment actors including patients, patient representatives, technologists, researchers, and AI policy experts.	
42	Artificial Intelligence in healthcare: advancing patient-centric care through co-design and responsible implementation	European Patients Forum (EPF)	2023	Europe	Patient organisation	Participatory report	To explore the applications, benefits, and challenges associated with AI in healthcare from a patient perspective. To provide key recommendations for responsible deployment of AI solutions.	Survey with 146 patient organisations and individual patient advocates; discussions with participants of a bootcamp on AI; consultation with working group and secretariat	
43	The Use of Artificial Intelligence in Irish General Practice	Irish College of General Practitioners (ICGP)	2025	Ireland	National professional or regulatory body	Report	To outline potential uses of AI in general practice. To outline risks and problems associated with the use of AI.	-	-
44	Citizens' Jury: Artificial Intelligence in Healthcare in Ireland	Irish Platform for Patient Organisations, Science and Industry (IPPOSI)	2025	Ireland	Patient organisation	Participatory report	To record the outputs from the IPPOSI citizen's jury discussion process to develop recommendations on the future of AI in healthcare in Ireland	Citizens' jury with 24 members of the public	-

#	Title	Author	Year	Location	Type of organisation	Type of source	Aim	Approach/method (if reported)	Relevant underpinning conceptual, legal, theoretical, or policy frameworks
45	Avoiding the AI 'Off-Switch': Make AI Work for Clinicians to Deliver for Patients	Medical Protection Society (MPS) Foundation	2025	United Kingdom	Research institute/academic paper	White paper	To consider the impact of AI decision-support tools on clinicians, identifying tools that support clinicians to serve patients and those that increase clinician stress and workload.	Simulated consultations with 21 clinicians using AI prototype decision support tools. Post simulation interviews and surveys with clinicians and actor patients. Peer feedback on clinician performance during the simulation by 28 independent clinicians.	
46	Trustworthy Artificial Intelligence (AI) in healthcare	MedTech Europe	2019	Europe	Other: Industry trade association	Position paper	To provide the medtech industry's response and contribution to the topic of ethical AI	-	Ethics Guidelines for Trustworthy Artificial Intelligence (AI) defined by the European Commission High-Level Expert Group on AI
47	Trustworthy Artificial Intelligence (AI) in healthcare	MedTech Europe	2019	Europe	Other: Industry trade association	Position paper	To address the challenges at the European level that impede the deployment of AI in healthcare. To recommend specific policy actions to make healthcare better and safer, improve access	-	Policy measures specific to healthcare and medical technology perspective as relevant to AI in Europe; Ethical Guidelines and

#	Title	Author	Year	Location	Type of organisation	Type of source	Aim	Approach/method (if reported)	Relevant underpinning conceptual, legal, theoretical, or policy frameworks
							and outcomes, empower patients and citizens with information, and make healthcare delivery more efficient.		Policy and Investment Recommendations published in 2019 by the European Commission's High-Level Expert Group on AI
48	AI: How to get it right report	National Health Service (NHS)	2019	United Kingdom	Other: National public body	Report	To provide an overview of AI within the health and social care system and to outline where in the system AI technology can be used and the policy work that is, and will need to be done, to ensure AI is used in a safe, effective, and ethically acceptable manner.	State of the nation survey and international horizon scan of evidence	The Code of Conduct for Data-Driven Health and Care Technology and Nuffield Council on Bioethics' principles for data initiatives
49	Laying the foundations for artificial intelligence in health	Organisation for Economic Co-operation and Development (OECD)	2021	Global	Intergovernmental organisation	Report	To provide an overview of the background and current state of AI in health and to propose areas for future exploration by policy makers to advance the	Desktop review, stakeholder consultation	OECD Principles on AI

#	Title	Author	Year	Location	Type of organisation	Type of source	Aim	Approach/method (if reported)	Relevant underpinning conceptual, legal, theoretical, or policy frameworks
							future of responsible AI in health.		
50	Ethical Principles for Artificial Intelligence for Medicine Version 2	Royal Australian and New Zealand College of Radiologists (RANZCR)	2023	Australia and New Zealand	Professional association	Position paper	To develop ethical principles to inform the development of professional and practice standards regarding AI in medicine to guide all stakeholders involved in research or deployment of AI.	Consultation process	Complements existing medical ethical frameworks, which may not adequately address the issues likely to emerge from use of AI in medicine
51	Generative Artificial Intelligence and Large Language Models	Royal Australian and New Zealand College of Radiologists (RANZCR)	2024	Australia and New Zealand	Professional association	Position paper	To assist RANZCR, its staff, fellows, members, and other individuals to understand the risks and benefits of using AI.	-	-
52	Autonomous Artificial Intelligence Position Statement	Royal Australian and New Zealand College of Radiologists (RANZCR)	2024	Australia and New Zealand	Professional association	Position paper	To outline the position of RANZCR on the use of AI in medicine.	-	-
53	Care Workers' Guidance and Statement of Expectations on the Responsible use of AI and Particularly	The Careworkers Charity	2024	United Kingdom	Other: Charity	Position paper	To outline key principles for the responsible use of generative Artificial Intelligence (AI) in adult social care, offering valuable insights for	Roundtable discussion with frontline care workers	-

#	Title	Author	Year	Location	Type of organisation	Type of source	Aim	Approach/method (if reported)	Relevant underpinning conceptual, legal, theoretical, or policy frameworks
	Generative AI in Adult Social Care						employers, AI developers, policy makers, local authorities, regulators, and care workers.		
54	Priorities for an AI in healthcare strategy	The Health Foundation	2024	United Kingdom	Research institute	Report	To develop a dedicated strategy for AI in healthcare.	-	-
55	Ethics and Governance of Artificial Intelligence for Health	World Health Organisation (WHO)	2021	Global	Intergovernmental organisation	Report	To provide a set of ethical principles for the use of AI in healthcare.	Expert group on Ethics and Governance of AI for Health, involving 20 experts in public health, medicine, law, human rights, technology and ethics	

Appendix 5. Codes grouped into final themes

71 original codes	Refined codes (round 1) 43	Refined codes (round 2) 20	Final themes (round 3) 12	Number of sources
Accessibility	Equity	Inclusivity and non-discrimination	**Inclusivity and non-discrimination**	35/55
Equity				
Minimisation of bias	Minimisation of bias			
Co-design	Co-design			
User experience and acceptance				
Fairness	Fairness			
Justice				
Inclusivity	Inclusivity			
Public benefit	Public benefit	Upholding people's rights	**Upholding people's rights**	25/55
Autonomy	Autonomy			
Human Rights	Human rights			
Application of human values				
Ethics				
Respect human dignity	Respect human dignity			
Governance	Governance	Oversight	**Human agency and oversight**	27/55
Model validation and performance	Model validation and performance			
Oversight	Oversight			
Assessment	Assessment			
Information not recommendation	Human in the loop	Human in the Loop		
Human in the loop				

Non-equivalence to professionals				
Interoperability	Interoperability	Integration into clinical care	**Integration into care**	21/55
Responsiveness	Responsiveness			
Usability				
Preplanning				
Teamwork				
Well-led				
Universality	Universality			
Ambient AI	Additionality	Additionality		
Additionality				
Clinical relevance				
Effective				
Eco-responsibility	Eco-responsibility	Eco-responsibility and sustainability		
Sustainability				
Innovative	Innovative	Research		
Investment	Investment			
Development	Development			
Research	Research			
Robustness	Robustness	Security	**Technical robustness and security**	8/55
Security	Security			
Beneficence	Safety	Safe care	**Safe care**	23/55
Safety				
Non-maleficence				
Catastrophic dual use				
Traceability	Transparency	Transparency	**Transparency**	37/55

Transparency				
Credibility				
Explainability	Explainability			
Interpretability				
Caring	Caring	Trust		
Obsolescence	Obsolescence			
Trust	Trust			
Upskilling	Upskilling	Education, Training, and Development of Staff	**Education, training, development, and information provision**	18/55
Understanding				
Decision making				
Informed Consent	Informed Consent	Informed decisions		
Integrity				
Education	Education	Public Education and Information		
Accountability	Accountability	Responsibility	**Responsibility**	27/55
Fidelity/professional faithfulness	Fidelity/professional faithfulness			
Responsibility	Responsibility			
Compliance	Compliance	Compliance		
Regulatory compliance				
Professional Standards	Professional standards			
Person-centred care	Person-centred care	Human connection	**Human connection**	3/55
Paternalism				
Privacy	Privacy	Privacy	**Privacy**	29/55
Confidentiality				
Protection of Data	Protection of health data			

Self-determination regarding health data				
Data quality	Data quality	Data quality	**Data quality**	9/55

Appendix 6. Types of sources and evidence supporting each theme

Theme	Academic sources	Grey literature sources
Transparency (n=37)	• Interviews with experts in autonomous systems technologies in healthcare[57] • 4 narrative reviews[62,63,67,68] • 4 evidence reviews[21,61,64,65] • Interviews with healthcare professionals[73] • 2 frameworks[17,76] • Interviews with medical doctors[74] • International consensus process[7] • Scoping review[66] • Literature review, consensus approach and expert review[70] • Expert consensus process[15] • Discussion paper[79] • Scoping review and consensus process[71] • Summary and grouping paper[77] • Appeal article[80]	• Scoping review and environmental/policy scan from a governmental agency (Australian Commission on Safety and Quality in Health Care)[96] • Webpage from a national professional or regulatory body (Ahpra)[94] • 4 position papers from professional associations (Australian Medical Association; RANZCR), a national professional or regulatory body (College of Physicians and Surgeons of British Columbia), and an industry trade association in Europe (MedTech Europe)[8,81,82,86] • Report from a regulatory sandboxing pilot conducted by a national professional or regulatory body (Care Quality Commission)[97] • 3 reports from intergovernmental organisations (Council of Europe; OECD; WHO)[9,51,98] • Evidence review conducted by a European research centre (European Commission Joint Research Centre)[100] • 3 participatory reports from patient organisations (European Patients' Forum; IPPOSI)[6,89,91] • Report from a national professional or regulatory body in Ireland (ICGP)[101] • Report from a national public body (NHS)[102]

Theme	Academic sources	Grey literature sources
Inclusivity and non-discrimination (n=35)	• Interviews with experts in autonomous systems technologies in healthcare[57] • 5 evidence reviews[21,59,61,64,65] • 4 narrative reviews[62,63,67,68] • 2 literature reviews[60,70] • Conceptual framework[17] • Interviews with medical doctors[74] • Content analysis[75] • International consensus process[7] • 1 scoping review[66] • Discussion paper[79] • Summary and grouping paper[77]	• Scoping review and environmental/policy scan from a governmental agency (Australian Commission on Safety and Quality in Health Care)[96] • 2 position papers from professional associations (Australian Medical Association; RANZCR)[81,86] • 2 reports from a national professional or regulatory body (Irish College of General Practitioners; College of Physicians and Surgeons of British Columbia)[82,101] • 4 reports from intergovernmental organisations (Council of Europe; OECD; WHO)[9,51,98,99] • Report from a research centre (European Commission Joint Research Centre)[100] • 4 participatory reports from patient organisations (European Patients Forum; IPPOSI)[6,89-91] • White paper from a research institute (MPS Foundation)[92] • Report from a national public body (NHS)[102]
Privacy (n=29)	• Interviews with experts in AI systems technologies in healthcare[57] • 4 narrative reviews[62,63,67,68] • 2 literature reviews[60,70] • Interviews with healthcare professionals[73] • 3 review articles[21,64,65] • Conceptual framework[17] • Scoping review[66]	• Report informed by workshops, consultations and national survey from a professional association (Australian Alliance for Artificial Intelligence in Healthcare)[95] • Scoping review and environmental/policy scan from a governmental agency (Australian Commission on Safety and Quality in Health Care)[96] • Webpage from a national professional or regulatory body (Ahpra)[94]

Theme	Academic sources	Grey literature sources
	• Framework[76]	• 2 position papers by professional associations (Australian Medical Association, RANZCR)[81,86] • 2 reports from national professional or regulatory bodies (ICGP; College of Physicians and Surgeons of British Columbia)[82,101] • 3 reports from intergovernmental organisations (Council of Europe; WHO)[51,98,99] • 2 reports by a research institute/centre (European Commission Joint Research Centre; The Health Foundation)[100,103] • 2 participatory reports from patient organisations (European Patients' Forum; IPPOSI)[6,91] • Report from a national public body (NHS)[102]
Human agency and oversight (n=27)	• Interviews with experts in autonomous systems technologies in healthcare[57] • Interviews with medical doctors[74] • Literature review[60] • Scoping review and consensus process[71] • Review article[61] • Narrative review[68] • Summary and grouping paper[77] • Framework[76]	• Report informed by workshops, consultations and national survey from a professional association (Australian Alliance for Artificial Intelligence in Healthcare)[95] • Scoping review and environmental/policy scan from a governmental agency (Australian Commission on Safety and Quality in Health Care)[96] • Report from a regulatory sandboxing pilot by a national professional or regulatory body in the UK (Care Quality Commission)[97] • 5 position papers from a national professional or regulatory body in Canada (College of Physicians and Surgeons of British Columbia) a European industry trade association

Theme	Academic sources	Grey literature sources
		(MedTech Europe) and a professional association in New Zealand and Australia (RANZCR)[8,82-84,86] • 3 reports from intergovernmental organisations (OECD; WHO; Council of Europe)[9,51,98] • 2 reports from a research institute/centre (European Commission Joint Research centre; The Health Foundation)[100,103] • 4 participatory reports from patient organisations (IPPOSI; European Patients' Forum)[6,89-91] • White paper involving simulated consultations, interviews, surveys, and peer feedback conducted by a research institute in the United Kingdom (MPS Foundation)[92] • Report from a national public body (NHS)[102]
Responsibility (n=27)	• Interviews with experts in autonomous systems technologies in healthcare[57] • 3 narrative reviews[62,67,68] • Literature review[60] • 3 evidence reviews[15,21,61,64] • Interviews with healthcare professionals[73] • 2 frameworks[17,76] • Content analysis[75] • Scoping review[66]	• Scoping review and environmental/policy scan from a governmental agency (Australian Commission on Safety and Quality in Health Care)[96] • Webpage from a national professional or regulatory body (Ahpra)[94] • 5 position papers from two professional associations (Australian Medical Association; RANZCR), a national professional or regulatory body (College of Physicians and Surgeons of British Columbia), a European industry trade association (MedTech Europe) and a charity (The Careworkers' Charity)[8,81,83,85,86]

Theme	Academic sources	Grey literature sources
	• Scoping review and consensus process[71]	• 3 reports from intergovernmental organisations (Council of Europe; OECD; WHO)[9,51,99] • 2 participatory reports from patient organisations (European Patients' Forum; IPPOSI)[6,91] • Report from a national professional or regulatory body (ICGP) [101]
Upholding people's rights (n=25)	• 2 evidence reviews[21,59] • 3 narrative reviews[62,63,68] • 1 literature review[60] • 2 frameworks[17,76] • Interviews with medical doctors[74] • Content analysis[75] • Scoping review[66] • Expert consensus process[15] • Discussion paper[79] • Scoping review and consensus process[71]	• Scoping review and environmental/policy scan from a governmental agency (Australian Commission on Safety and Quality in Health Care)[96] • Webpage from a national professional or regulatory body (Ahpra)[94] • 2 reports from professional associations (Australian Medical Association; RANZCR)[81,86] • Position paper from a professional association (RANZCR)[87] • 3 reports from intergovernmental organisations (OECD; WHO; Council of Europe)[9,51,98] • 3 participatory reports from patient organisations (IPPOSI; European Patients' Forum)[6,89,91]
Safe care (n=23)	• Interviews with experts in autonomous systems technologies in healthcare[57] • 2 evidence reviews[21,59] • Conceptual framework[17] • Interviews with medical doctors[74] • Content analysis[75]	• Report informed by workshops, consultations and national survey from a professional association (Australian Alliance for Artificial Intelligence in Healthcare)[95] • Scoping review and environmental/policy scan from a governmental agency (Australian Commission on Safety and Quality in Health Care)[96]

Theme	Academic sources	Grey literature sources
	• 3 narrative reviews[63,67,68] • Literature review, consensus approach and expert review[70] • Expert consensus process[15]	• 3 position papers from two professional associations (Australian Medical Association; RANZCR), a European industry trade association (MedTech Europe), and a professional association (RANZCR)[8,81,86] • Report from a regulatory sandboxing pilot conducted by a national professional or regulatory body (Care Quality Commission)[97] • 3 reports from an intergovernmental organisation (OECD; Council of Europe; WHO)[9,51,98] • 3 participatory reports from patient organisations (IPPOSI; European Patients' Forum)[6,89,91]
Integration into clinical care (n=21)	• Interviews with experts in autonomous systems technologies in healthcare[57] • Review article[61] • 2 frameworks[17,76] • Content analysis[75] • International consensus process[7] • Narrative review[68] • Discussion paper[79] • Scoping review and consensus process[71] • Appeal article[80]	• Scoping review and environmental/policy scan from a governmental agency (Australian Commission on Safety and Quality in Health Care)[96] • 2 reports from national professional or regulatory bodies (Care Quality Commission; ICGP)[97,101] • Citizen's jury involving members of the public conducted by an Irish patient organisation (IPPOSI)[91] • White paper involving simulated consultations, interviews, surveys, and peer feedback conducted by a research institute (MPS Foundation)[92] • Report from a national public body (NHS)[102] • 2 reports from intergovernmental organisations (OECD; WHO)[9,51]

Theme	Academic sources	Grey literature sources
		• 2 position papers from a professional association (RANZCR) and a European industry trade association (MedTech Europe)[84,86] • Report from a research institute (The Health Foundation)[103]
Education, training, development and information provision (n=18)	• Literature review[60] • Review article[61] • Interviews with healthcare professionals[73] • Interviews with medical doctors[74] • Framework[76]	• Report informed by workshops, consultations and national survey from a professional association (Australian Alliance for Artificial Intelligence in Healthcare)[95] • Scoping review and environmental/policy scan from a governmental agency (Australian Commission on Safety and Quality in Health Care)[96] • Webpage from a national professional or regulatory body (Ahpra)[94] • 3 reports from intergovernmental organisations (Council of Europe; OECD; WHO)[9,51,98] • 3 participatory reports from patient organisations (European Patients' Forum; IPPOSI)[6,89,91] • White paper involving simulated consultations, interviews, surveys, and peer feedback conducted by a research institute in the United Kingdom (MPS Foundation)[92] • 2 position papers from a European industry trade association (MedTech Europe) and a professional association (RANZCR)[84,86] • Report from a research institute (The Health Foundation)[103]

Theme	Academic sources	Grey literature sources
Data quality (n=9)	• Interviews with experts in autonomous systems technologies in healthcare[57] • Review article[64] • Literature review[70]	• 3 participatory reports from a patient organisation (European Patients' Forum)[6,89,90] • Report from a national professional or regulatory body (ICGP)[101] • Position paper from an industry trade association (MedTech Europe)[84] • Report from a national public body (NHS)[102]
Security (n=8)	• Interviews with experts in autonomous systems technologies in healthcare[57] • Interviews with medical doctors[74] • Review article[65] • International consensus process[7] • Expert consensus process[15]	• Scoping review and environmental/policy scan from a governmental agency (Australian Commission on Safety and Quality in Health Care)[96] • Citizen's jury involving members of the public conducted by an Irish patient organisation (IPPOSI)[91] • Report from a national public body (NHS)[102]
Human connection (n=3)	• Interviews with healthcare professionals[73] • Review article[21]	• Report from a national public body (NHS)[102]

Ahpra = Australian Health Practitioner Regulation Agency; ICGP = Irish College of General Practitioners; IPPOSI = Irish Platform for Patient Organisation, Industry and Science; MPS = Medical Protection Society; NHS = National Health Service; OECD = Organisation for Economic Co-operation and Development; RANZCR = Royal Australian and New Zealand College of Radiologists; WHO = World Health Organisation

Appendix 7. Mapping sources to themes

Source/theme	Transparency	Inclusivity and non-discrimination	Privacy	Human agency and oversight	Responsibility	Upholding people's rights	Safe care	Integration into care	Education, training, development and information provision	Data quality	Technical robustness and security	Human connection
Alelyani (2024)	x	x	x	x	x		x	x		x	x	
Armitage (2024)		x				x	x					
Elendu et al. (2023)	x	x	x		x	x						
Corfmat et al. (2025)		x	x	x	x	x			x			
Drabiak (2022)	x	x		x	x			x	x			
Elgin & Elgin (2024)	x		x		x				x			x
Jeyaraman et al. (2023)	x	x	x		x					x		
Jha et al. (2025)	x	x	x		x	x	x	x				
Kahraman et al. (2024)	x	x		x		x	x		x		x	
Katirai (2023)		x			x	x	x	x				
Kim et al. (2023)	x	x	x								x	
Lekadir et al. (2025)	x	x						x			x	
Maccaro et al. (2024)	x	x	x		x	x						
Marques et al. (2024)	x	x	x		x		x					
Mennella et al. (2024)	x	x	x	x	x	x	x	x				
Arbelaez Ossa et al. (2024)												
Prakash et al. (2022)												
Reddy et al. (2021)	x	x	x				x			x		
Rose & Shapiro (2024)												
Saenz et al. (2024)	x					x	x				x	
Savulescu et al. (2024)	x	x	x		x	x	x					x
Seroussi & Zablit (2024)	x	x				x		x				

Source/theme	Transparency	Inclusivity and non-discrimination	Privacy	Human agency and oversight	Responsibility	Upholding people's rights	Safe care	Integration into care	Education, training, development and information provision	Data quality	Technical robustness and security	Human connection
Sousa-Pinto et al. (2025)	x			x	x	x		x				
Tierney et al. (2025)	x	x		x								
Harishbhai Tilala et al. (2024)	x	x	x			x	x					
Upadhyay et al. (2023)	x							x				
Vandemeulebroucke (2024)												
Zuchowski et al. (2024)	x		x	x	x	x		x	x			
Australian Alliance for Artificial Intelligence in Healthcare (2021)			x	x			x		x			
Australian Commission on Safety and Quality in Health Care (2024)	x	x	x	x	x	x	x	x	x		x	
Australian Health Practitioner Regulation Agency (Ahpra) (2024)	x		x		x	x			x			
Australian Medical Association (AMA) (2023)	x	x	x		x	x	x					
Australian Medical Association (AMA) (2024)												
Care Quality Commission (CQC) (2020)	x			x			x	x				
College of Physicians and Surgeons of British Columbia (2024)	x	x	x	x								
College of Physicians and Surgeons of British Columbia (2025)				x	x							

Source/theme	Transparency	Inclusivity and non-discrimination	Privacy	Human agency and oversight	Responsibility	Upholding people's rights	Safe care	Integration into care	Education, training, development and information provision	Data quality	Technical robustness and security	Human connection
Council of Europe (2022)	x	x	x	x		x	x		x			
Council of Europe (2024)		x	x		x							
European Commission Joint Research Centre (2023)	x	x	x	x								
European Patients Forum (EPF) (2020)	x	x		x		x	x		x	x		
European Patients Forum (EPF) (2022)		x		x						x		
European Patients Forum (EPF) (2023)	x	x	x	x	x	x	x		x	x		
Irish College of General Practitioners (ICGP) (2025)	x	x	x		x			x		x		
Irish Platform for Patient Organisations, Science and Industry (IPPOSI) (2025)	x	x	x	x	x	x	x	x	x		x	
Medical Protection Society (MPS) Foundation (2025)		x		x				x	x			
MedTech Europe (2019)	x			x	x		x					
MedTech Europe (2019)				x				x	x	x		
National Health Service (NHS) (2019)	x	x	x	x				x		x	x	x
Organisation for Economic Co-operation and Development (OECD) (2021)	x	x		x	x	x	x	x	x			

Source/theme	Transparency	Inclusivity and non-discrimination	Privacy	Human agency and oversight	Responsibility	Upholding people's rights	Safe care	Integration into care	Education, training, development and information provision	Data quality	Technical robustness and security	Human connection
Royal Australian and New Zealand College of Radiologists (RANZCR) (2023)	x	x	x	x	x	x	x	x	x			
Royal Australian and New Zealand College of Radiologists (RANZCR) (2024a)												
Royal Australian and New Zealand College of Radiologists (RANZCR) (2024b)						x						
The Careworkers Charity (2024)					x							
The Health Foundation (2024)			x	x				x	x			
World Health Organisation (WHO) (2021)	x	x	x	x	x	x	x	x	x			

Published by the Health Information and Quality Authority.

For further information please contact:

Health Information and Quality Authority
George's Court
George's Lane
Smithfield
Dublin 7
D07 E98Y

Phone: +353 (0) 1 814 7400
Email: info@hiqa.ie
www.hiqa.ie

Health Information and Quality Authority
An tÚdarás Um Fhaisnéis agus Cáilíocht Sláinte