MACQUARIE University

AUSTRALIAN INSTITUTE OF HEALTH INNOVATION

UNIVERSITY OF WOLLONGONG AUSTRALIA

Literature Review and Environmental Scan Report

AI Implementation in Hospitals: Legislation, Policy, Guidelines and Principles, and Evidence about Quality and Safety

Prepared for the Australian Commission on Safety and Quality in Health Care by:

Australian Institute of Health Innovation
Macquarie University

Prof. Farah Magrabi
Ms. Leonie Bates
Ms. Kalissa Brooke-Cowden
Ms. Tamasha Jayawardena
Ms. Amy Wang
Prof. Enrico Coiera

Australian Centre for Health Engagement, Evidence and Values
University of Wollongong

Dr Yves Saint James Aquino
Ms. Emma Frost
Prof. Stacy Carter
Dr Carolyn Adams
Ms. Alaa Almohanna

Version: 5 April 2024
Finalised: 29 May 2024

Table of contents

Figures

Tables

Boxes

Executive summary

To harness the enormous benefits of Artificial Intelligence (AI) in healthcare, we must implement and use it safely and responsibly. The purpose of this report is to provide a review of the recent literature and undertake an environmental scan to identify principles that enable the safe and responsible implementation of AI in healthcare. It presents evidence from the contemporary published literature about AI implemented in acute care as well as current, published legislation, policies, guidelines, and principles for AI implementation in healthcare. The findings will be considered by the Australian Commission on Safety and Quality in Health Care (ACSQHC) for future development of resources to assist healthcare organisations in evaluating and implementing AI.

Policy scan and principles for safe and responsible AI in healthcare

Chapters 2 and 3 report the findings from an environmental scan of international (USA, UK, New Zealand, Canada, Singapore), intergovernmental (WHO, OECD and EU) and national legislation and policy to gain insight about principles (e.g. guidelines, governing ideas, and strategies) for implementation of AI in acute care. The review covers both cross-sectoral legislation and policy that is relevant in healthcare, as well as healthcare-specific legislation and policy.

Key findings from the environmental scan of national and international legislation and policy are:

- Governance of AI in healthcare is not limited to new AI-specific laws, but also involves primary legislation and policy (e.g. privacy laws, human and consumer rights law, and data protection laws).
- Similar to Australia, national ethics frameworks are common in the reviewed countries and influence policy formulation. These frameworks are designed to support healthcare organisations in those jurisdictions by guiding the implementation of AI in their practice. The US Department of Health and Human Services drew on a national ethics framework to develop a playbook to guide health departments in embedding ethical principles in AI development, acquisition, and deployment (1). Internationally, governance approaches include establishing dedicated regulatory and oversight authorities (including healthcare-specific bodies), requiring risk-based or impact assessments, provisions to increase transparency or prohibit discrimination, regulatory sandboxing, as well as formal tools or checklists.
- Australia's National Ethics Framework is commonly used to frame Australian policy. The Australian Government has commenced development of a national risk-based approach to cross-sectoral AI regulation (2), based on four principles: i/ balanced and proportionate (achieved via risk-based assessment); ii/ collaborative and transparent (achieved via public engagement and expert involvement); iii/ consistent with international requirements; iv/ putting community first. This national approach will shape the future of AI governance and implementation in health services; in some jurisdictions, such as NSW, good progress has been made on developing state-based governance frameworks, including in health (see Section 3.3.1 page 49-50). The NSW Government's AI Ethics Principles are embedded in the NSW AI Assurance Framework, which applies to uses of AI in the NSW health system.
- Current developments in Australian governance and regulation of AI in healthcare include governance via existing cross-sectoral approaches (e.g. privacy and consumer law), regulation of software as a medical device, and specific health governance proposals from research and health organisations. The most significant developments in the healthcare sector are policy initiatives by the Australian Alliance for Artificial Intelligence in Healthcare (AAAiH) (73), The Royal Australian and New Zealand College of Radiologists (3), and the Australian Medical Association (4).

Legislative and policy environment

- The AAAiH National Policy Roadmap Process has recommended, by consensus, that Australia establish an independent National AI in Healthcare Council to oversee AI governance in health. This Council should be established urgently. Its work should be shaped by the National AI Ethics Principles and the recommendations made by consensus in the National Policy Roadmap process. One of the key issues to address is practical guidance on clarifying consent and transparency requirements. The Roadmap also recommended that the Council engage individual professional bodies to develop profession-specific codes of conduct, and oversee accreditation regarding minimum AI safety and quality standards of practice covering cybersecurity threats, patient data storage and use, and best practice for deployment, governance and maintenance of AI. Such accreditation could fall under the remit of the ACSQHC's accreditation scheme.
- AAAiH's recommendation for a risk-based safety framework also called for the improvement of national post-market safety monitoring so that cases of AI-related patient risk and harm are rapidly detected, reported and communicated.
- Both the AAAiH and the Medical Technology Association of Australia (MTAA) recommended development of a formal data governance framework as well as mechanisms to provide industry with ethical and consent-based access to clinical data to support AI development and leverage existing national biomedical data repositories.
- The Australian legislative and policy environment for AI is rapidly changing: upcoming developments include changes in cross-sectoral legislation (e.g. privacy law) and an intended national risk-based approach to AI legislation.
- Review of Australian guidance documents showed that detailed legal analysis of privacy requirements with respect to AI implementation in healthcare (see 3.3.4 Privacy and confidentiality), and detailed legal analysis of accountability and liability in AI use (see 3.3.7 Accountability and liability), may be warranted, as these are not as well resolved in Australia as in some other jurisdictions. This could potentially support legal reform.

Key issues for health organisations and clinicians

- Ensure high quality, local, practice-relevant evidence of AI system performance before implementation.
- Significant training and support for clinicians and other health workers is required during the implementation and integration of AI systems into existing clinical information systems or digital health solutions (e.g., electronic medical records, EMR). Training includes skill development to use the AI system, but also includes training in ethical and liability considerations, cybersecurity, and capacity to inform patients about the use of AI in their care (see Chapter 6, section 6.10 for details).
- Ensure AI implementation, and organisational policy, complies with existing legislation (e.g. data privacy, consumer law, and cybersecurity policy) and relevant AI ethics frameworks.
- AI governance should build on existing governance processes in healthcare organisations e.g. for patient safety, digital health and research ethics. This is necessary to ensure safe and responsible use of AI, as well as clarify lines of individual and organisation responsibility over AI-assisted clinical and administrative decision-making that comply with existing liability rules.
- Strengthen engagement with consumers, communities, and stakeholders in healthcare AI implementation to ensure trustworthiness, and to shape implementation and use of consumer- or patient-facing AI. An example of policy-orientated community engagement is illustrated by a national Australian citizens' jury convened to deliberate about AI implementation in healthcare. See Box 2 in Chapter 3, section 3.3.2 for the jury's recommendations.

- Implementation of AI in health services should ensure appropriate Aboriginal and Torres Strait Islander governance, by connecting AI governance processes in health systems to existing Aboriginal and Torres Strait Islander governance structures. Implementation should be in line with principles of Indigenous Data Sovereignty.
- Transparency and consent are key issues for implementation of AI in health services. Governance of transparency and consent should draw on existing expertise and governance systems in healthcare organisations, including clinical ethics committees, research ethics committees, digital health committees, consumer governance committees and risk management structures. In developing approaches to transparency and consent, health organisations should note that:
    - Fundamental requirements for consent in clinical contexts—that a person must have capacity, consent voluntarily and specifically, and have sufficient information about their condition, options, and material risks and benefits—remain unchanged by the use of AI.
    - There is limited guidance available regarding requirements for consent to the use of AI as an element of clinical care.
    - Across the policy documents reviewed, there is strong agreement that there should be transparency about the fact that AI is being used.
    - Also consider transparency regarding training data, data bias, AI system performance and evaluation methods.
    - Risk-based assessment could require greater transparency for higher-risk applications.
    - As noted above, consent and transparency are potential areas of focus for a National Council on AI in Health.
- Implement risk assessment frameworks to address the risk of bias, discrimination or unfairness in initial evaluation and ongoing monitoring of AI systems. See Appendix A for an example of an impact assessment tool.
- Ensure use of existing patient safety and quality systems for monitoring AI incidents and safety events (including hazards and near miss events) as well as post-market safety monitoring so that cases of AI-related patient risk and harm are rapidly detected, reported and managed.

Literature review and principles for safe and responsible AI in healthcare

Chapters 4 and 5 report findings from a scoping review of the literature to identify principles for safe and responsible implementation of AI at the health service level. The review covers 75 primary studies about AI systems deployed in acute care that were published in the peer-reviewed literature from 2021-2023 as well as nine studies reporting emerging safety problems associated with AI in healthcare.

For healthcare organisations, safe and responsible AI in builds on best-practice approaches for digital health. Key findings and principles for implementing AI systems at the health service level are as follows:

AI in acute care settings

Key finding 1: AI technologies are being applied in a wide variety of clinical areas, with studies identifying clear clinical use cases for their implementation. The most common clinical tasks supported by AI systems are diagnosis and procedures.

All the AI systems identified in the literature search were based on traditional machine learning (ML) techniques and most were *assistive* requiring clinicians to confirm or approve AI provided information or decisions. Up until December 2023, no studies had evaluated the implementation of AI in hospital operations or the clinical use of foundation models or generative AI in routine patient care.

Principle 1: Take a problem-driven approach to AI implementation, an AI system should address specific clinical needs. Confirm the specific clinical use case before implementation i.e. the types of patients and condition where the AI system is intended to improve care delivery and patient outcomes.

Approach to AI implementation

Key finding 2: The literature demonstrated multiple ways in which health services implemented AI systems such as to: i/ develop AI systems in-house; ii/ co-develop in partnership with technology companies; and iii/ purchase AI systems from commercial vendors (including AI systems subject to medical device regulation). Evidence of engagement with hospital ethics committees or clinical governance boards from a responsible use perspective was poorly reported in the studies reviewed.

Principle 2: Deployment of AI systems that have been developed externally or internally, is a highly complex process and should be undertaken in partnership with key stakeholders including healthcare professionals and patients. Consultation should occur with those who have specialist skills traversing clinical safety, governance, ethics, IT system architecture legal and procurement, and include the specific healthcare professionals as well as patient representatives and/or patient liaison officers.

Principle 3: When purchasing AI systems from commercial vendors, assess clinical applicability and feasibility of implementation in the care setting. Consider the system performance and whether the ML model will transport from its training and validation environment to the local clinical setting of interest. Consider feasibility of testing the AI using localised de-identified data sets or localised synthetic datasets to illicit utility and performance of the AI system in the local clinical area of interest, before conducting pilot implementation projects.

AI system performance

Key finding 3: AI system performance was usually assessed against a comparator (e.g. human or another device). Evaluation metrics such as sensitivity, specificity, positive predictive value, accuracy and F1 score were commonplace amongst the literature.

Principle 4: Ensure AI is fit for clinical purposes by assessing evidence for system performance against a comparator. Evaluate performance in the local context of interest using localised de-identified datasets or synthetic datasets, before conducting pilot implementation projects to measure AI system performance and answer any evidence gaps in prior assessments.

Key finding 4: Emerging evidence highlights the impact of distributional shift, stemming from disparities between the dataset on which AI systems are trained and deployment datasets. However, studies describing implementation lacked any reported quality assurance measures, such as post-deployment monitoring, auditing, or performance reviews.

Principle 5: Monitor AI system performance in-situ post deployment, by means of electronic dashboards or other performance monitoring/auditing methods to rapidly detect and mitigate the effects of distributional shift. This should be underpinned by technical support as well as processes around planned and unplanned system downtime.

Safety of AI in healthcare

Key finding 5: Emerging evidence underscores safety concerns associated with AI systems and their impact on patient care. Although literature reporting on AI-related adverse events has been limited, evidence from the US FDA's post-market safety monitoring emphasises the necessity of examining issues with AI systems beyond the known limitations of ML algorithms. Predominantly, issues with data acquisition were observed, while problems with use i.e. the misapplication of AI and its intended purposes were four times more likely to lead to patient harm that technical issues.

Principle 6: A whole-of-system approach to safe AI implementation is needed. Ensure that AI systems are effectively integrated into IT infrastructure as they are highly reliant on data and integration with the IT infrastructure and other clinical information systems. Data quality and requirements for any accompanying changes to the EMR and other supporting clinical information systems need to be assessed to ensure data provided to the AI system is fit for purpose and its output is accurately displayed to users.

Role of AI in clinical task, clinical workflow, usability, and safe use

Key finding 6: AI systems in the literature were predominantly assistive or providing autonomous information meaning users were required to confirm or approve AI provided information or decisions, and still had overall autonomy over the task at hand. However, problems with the use of AI were more likely to harm patients compared to algorithm issues in safety events reported to the US FDA's post-market safety monitoring.

Principle 7: Ensure that users are aware of the intended use of AI systems (see Box 3). Training around the intended use and safe use of AI should be developed in consultation with the AI developer, clinical governance, patient safety and clinical leaders. The training should be maintained and updated throughout the life cycle of the AI system.

Key finding 7: End user engagement to devise clinical workflows and training ahead of deployment were less well reported in the literature. When understanding interaction and adoption of AI systems into healthcare workflows, user experience data and user metrics uncovered facilitators and barriers.

Principle 8: Integrate AI systems with clinical workflow. Devise clinical workflows for AI systems in a real-world care setting to ensure AI is seamlessly integrated into practice. Evaluate early to ensure AI fits local

requirements and address any issues. A pilot implementation can be used to test and refine integration with clinical workflow and supporting systems.

Principle 9: Identify issues with system usability via user metrics and short, regular survey requests. Address these issues promptly by collaboration with the AI developer and clinicians using the system.

Clinical utility and effects on decision-making

Key finding 8: Decision change outcomes such as incorrect/correct decisions and the rate at which clinicians make decisions, their decision velocity, help to characterise effects of AI systems on clinical decision-making. Confidence, acceptability and trust in the AI system were important factors in decision change.

Principle 10: Limitations of the AI system abilities must be made clear to all staff engaging with the AI system. This can be fostered by collaboration with the AI developer and strong engagement with clinicians in both pre-deployment and post deployment phases. AI incidents and safety events (including hazards and near miss events) should be easy to report and escalate.

Principle 11: Before-and-after studies or historical cohort studies can be utilised to assess the clinical utility and safety of AI compared to a time period when AI was not implemented.

Effects on care delivery and patient outcomes

Key finding 9: Care process changes were not well described in the literature. However, clinical outcomes were ubiquitously reported as primary, secondary and exploratory outcomes, with many studies having a clinical outcome as the study primary endpoint.

Principle 12: Ensure AI systems are suitably embedded i.e. their use and clinical utility in a particular context is established using formative evaluation methods during implementation before conducting clinical trials to assess impact on care delivery and patient outcomes.

Conclusion

The adoption of AI technologies in Australian healthcare is still in its early stages. By safely and responsibly implementing the current generation of AI, Australian health services can prepare for the future. This involves building on existing governance processes, strengthening engagement with consumers, utilising the available data infrastructure, and establishing robust processes for evaluating the performance, clinical utility, and usefulness of AI assistance based on current best practices for implementing digital health systems. Preparation is crucial as healthcare AI systems evolve from making recommendations to autonomously performing clinical tasks. Moreover, Australia has the opportunity to provide guidance to other countries seeking to use modern AI systems to improve care delivery and patient outcomes effectively and safely.

1. Introduction

Globally, there has been an increase in the use of Artificial intelligence (AI) technologies in healthcare settings for a range of tasks requiring pattern recognition, reasoning or learning (5). While AI has been studied for more than 50 years, its current resurgence is largely driven by developments in machine learning (ML) and specifically deep learning. Recently, these deep learning methods have achieved unprecedented levels of performance in a variety of tasks such as language and image generation, using generative AI methods, including generative pretrained transformers (GPTs).

In healthcare, AI promises to transform care delivery as it has the potential to harness the vast amounts of genomic, biomarker, and phenotype data that are being generated across the health system and beyond (6, 7). AI is considered to have the potential to change work practices and ease pressures on the health system by automating clinical workflows. For example, by converting clinical tasks into algorithms that could lead to improved patient outcomes across the healthcare landscape.

Today, AI is being incorporated into a variety of clinical systems for detecting findings, suggesting diagnoses and recommending treatments in data-intensive specialties like radiology, pathology and ophthalmology (7). These AIs can aid human decision making, from systems that acquire and analyse data and provide options for decisions, to systems with the capability of making decisions entirely on their own (7, 8). With time, systems are expected to become increasingly autonomous, going beyond making recommendations to autonomously performing tasks such as controlling closed loop clinical machines like ventilators or insulin pumps, triaging patients or screening referrals (9, 10). With the public release of generative AI, their applications in assisting clinicians with many complex tasks like creating health records, writing referral letters, and generating summaries of the clinical evidence are rapidly emerging (11).

In Australia, there are limited examples of AI being used to deliver safe, high-quality healthcare. Few AI systems are in routine use and there is limited evidence of clinical benefits to date. To be successful in healthcare, AI must perform well in real-world clinical settings. Yet there are many complex challenges in the "last mile" of implementation that may make technically high performing algorithms perform poorly in real-world settings (12). This presents an opportunity for the Australian Commission on Safety and Quality in Health Care (the Commission) to develop resources to assist health services to evaluate and implement AI before the widespread uptake of these technologies. The Commission was established to contribute to improving health outcomes and experiences for all patients and consumers, and to improve the value and sustainability of the health system by leading and coordinating national improvements in the safety and quality of healthcare. To this end it has a mature program in digital health to optimise safety and quality in the implementation of digital clinical systems.

The purpose of this report is to provide a review of the recent literature and undertake an environmental scan to identify principles that enable the safe and responsible implementation of AI in healthcare. It presents evidence from the contemporary published literature about AI implemented in acute care as well as current, published legislation, policies, guidelines, and principles for AI implementation in healthcare. The findings will be considered by the Commission for future development of resources to assist health services in evaluating and implementing AI.

Chapter 2 presents a selective summary of the international legislative and policy environment. Chapter 3 examines Australian legislation and policy relevant for AI in healthcare. Chapter 4 presents a narrative review of the contemporary published literature about AI deployed in acute care, with a lens on safe implementation, clinical outcomes, workflow, and workforce impacts. Chapter 5 identifies and maps

emerging safety problems associated with AI in healthcare. The key findings from the policy review and principles for safe and responsible AI in healthcare are presented in Chapter 6. The key findings from the literature review and principles for safe and responsible implementation in health services are in Chapter 7, and the conclusion is in Chapter 8.

2. International legal and policy environment

2.1 Introduction

This chapter considers international jurisdictions and intergovernmental organisations that have developed law and policy to address the ethical and safety concerns arising from implementation of AI, with a focus on law and policy relevant to healthcare (13). According to the Organisation for Economic Cooperation and Development (OECD), by 2020 there were 50 countries developing or implementing national AI strategies (14). Research has shown a degree of variation across jurisdictions in approaches to AI for healthcare due to differences in existing regulatory norms and underpinning social values (13). We aimed to review a selection of international law and policy to gain insight into governance approaches that could underpin efforts towards quality and safety in the implementation of AI in acute care and would be of relevance in the Australian context.

To achieve this aim, we conducted a review of international legislation and policies, guided by the research question:

> *What are the principles (e.g. guidelines, governing ideas, and strategies) for implementation of AI in acute healthcare that are shared across international jurisdictions?*

We searched for legislation and policies from select countries (United Kingdom, United States, New Zealand) and intergovernmental organisations. We included documents that dealt directly with AI and were relevant to implementation processes (deployment, acquisition, regulation, review, distribution, and use) in the acute healthcare context.

2.2 Method

The first stage of our search strategy was to identify relevant documents cited in papers from the Australian Government's *Safe and Responsible AI* consultation (2, 15). In addition, we searched jurisdictional legislative databases, Google Advance, government websites of in-scope countries, websites of in-scope intergovernmental organisations, and secondary references from eligible documents. One researcher conducted the search between 8 January and 15 March 2024.

The search method and screening process was managed using a combination of:

- An Excel spreadsheet to record search results, capture details of potentially eligible documents, and remove duplicates.
- Covidence, an online systematic review software, to streamline screening of potentially eligible documents.
- Endnote, a reference management software, to store the full text of all documents included for review.

2.2.1 Eligibility criteria

See Table 1 for the inclusion and exclusion criteria.

Table 1: Eligibility criteria.

Inclusion criteria	Exclusion criteria
Legislation and policy, including guidelines and principles, of direct relevance to the use of AI in acute care.	Peer-reviewed literature, documents that fall outside the definition of policy and legislation, repealed

Inclusion criteria	Exclusion criteria
Legislation includes Acts, Regulations and other binding legislative instruments implemented at national, state or local levels of government Policy includes binding and non-binding statements by organisations that are intended to provide authoritative guidance	legislation, non-official documents (blog posts, news articles), policies or similar documents not about AI, Policies or similar documents not relevant to acute care, community responses to discussion papers
Jurisdictions: New Zealand, USA and UK Intergovernmental: WHO, UNESCO, EU and OECD	Exclude documents from countries not listed; exclude documents from intergovernmental organisations not dealing directly with healthcare or where documents are not directly relevant to the healthcare context
Type of organisation: For countries included, documents released by government agencies only	Documents released by non-government organisations (industry or professional bodies)
Published or available in English	Unavailable in English
Practice domain: Deal directly with AI and relevant to the acute healthcare context; acute healthcare defined as hospital care for short-term/acute conditions	AI applications that apply in other domains or industries that are not relevant for acute care (e.g. public health, billing), general standards (ISO) Not acute care, such as primary care or community care or public healthcare that have no overlap with hospital care
Focus on processes involved in implementation of AI: deployment, acquisition, regulation, review, distribution, use	Exclude guidance on the development of AI directed to developers
Full text is accessible	Full text not accessible
Date of publication: from 2018 to present	Earlier than 2018

2.2.2 Document sources

References cited in the Safe and Responsible White Paper and Interim Response

We extracted all the references from the Australian Government's Safe and Responsible White Paper (2), and the separate Interim Response (15) to the public consultation. The references were recorded in Excel, and included the following details: document title, country, and link to full text (if available). All the eligibility criteria were implemented. Documents from countries other than US, UK, New Zealand and Australia were included as long as they were within the practice domain (AI for the acute care context).

Legislative databases

Table 2 shows a list of legislative databases in the US, UK and New Zealand searched for documents using the same eligibility criteria. Search terms included "artificial intelligence", "machine learning", "automated decision making", "algorithm" and "intelligent system". Results were recorded in Excel, and included the following details: database, title of document, country/jurisdiction, link to full text, and date of publication.

Table 2: Legislative databases searched.

Country	Website
New Zealand	https://www.legislation.govt.nz/
US	https://uscode.house.gov/search/criteria.shtml https://www.ncsl.org/ https://www.federalregister.gov/
UK	https://www.legislation.gov.uk

Google Advanced Search

We used Google Advanced Search to search for policies from the US, UK and New Zealand. The following advanced search strategy was implemented for each jurisdiction: published in 2018 or later, .gov website domain, and country restriction. Search terms included "artificial intelligence", "machine learning", "automated decision making", "algorithm" and "intelligent system", "health", "policy", "framework". If the search yielded fewer than 50 documents, all results were recorded in Excel. If there were more than 50 documents, we included all documents before saturation or repetition was reached (typically after fewer than 100 documents). Results were recorded in Excel, and included the following details: database, title of document, country/jurisdiction, link to full text, and date of publication.

Targeted website search

Government agency websites in the US, UK and New Zealand, and were searched for existing policies on AI implementation in healthcare. We limited our search to websites of national or federal government agencies dedicated to:

- Healthcare services (e.g., Department of Health)
- Artificial intelligence (e.g., US https://ai.gov/)
- Medical device regulation (e.g., Medicines and Healthcare products Regulatory Agency, UK)

Intergovernmental policies

In addition, we searched websites of EU/European Parliament, WHO and OECD for intergovernmental policies and EU legislation. Each government agency and intergovernmental website was searched manually by using the website's search function and using search terms such as "artificial intelligence" or "automated decision making". We did not search the websites of individual EU member states.

Table 3 lists the websites systematically searched. All relevant documents were added to the Excel file containing results from the previous sources.

Table 3: List of government and intergovernmental websites systematically searched for AI policies.

Country or organisation	Organisation	Website
US	White House AI initiative	https://ai.gov/actions/
	Food and Drug Administration	https://www.fda.gov
	Department of Health and Human Services	https://www.hhs.gov/
UK	Department of Health and Social Care	https://www.gov.uk/government/organisations/department-of-health-and-social-care
	Office for Artificial Intelligence	https://www.gov.uk/government/organisations/office-for-artificial-intelligence
	Medicines and Healthcare products Regulatory Agency	https://www.gov.uk/government/organisations/medicines-and-healthcare-products-regulatory-agency
	National Health Services	https://www.nhs.uk/
NZ	Ministry of Health	https://www.health.govt.nz/

Country or organisation	Organisation	Website
	Medicines and Medical Devices Safety Authority	https://www.medsafe.govt.nz
EU	European Union/European Parliament	https://eur-lex.europa.eu/homepage.html https://www.europarl.europa.eu/thinktank/en/home
WHO	World Health Organization	https://iris.who.int/
OECD	Organisation for Economic Cooperation and Development	https://www.oecd.org/gov/regulatory-policy/

Cited references

During full-text screening for inclusion and data extraction, each eligible document's reference list was further screened for potentially relevant policy or legislation missed during the preceding methods of searching.

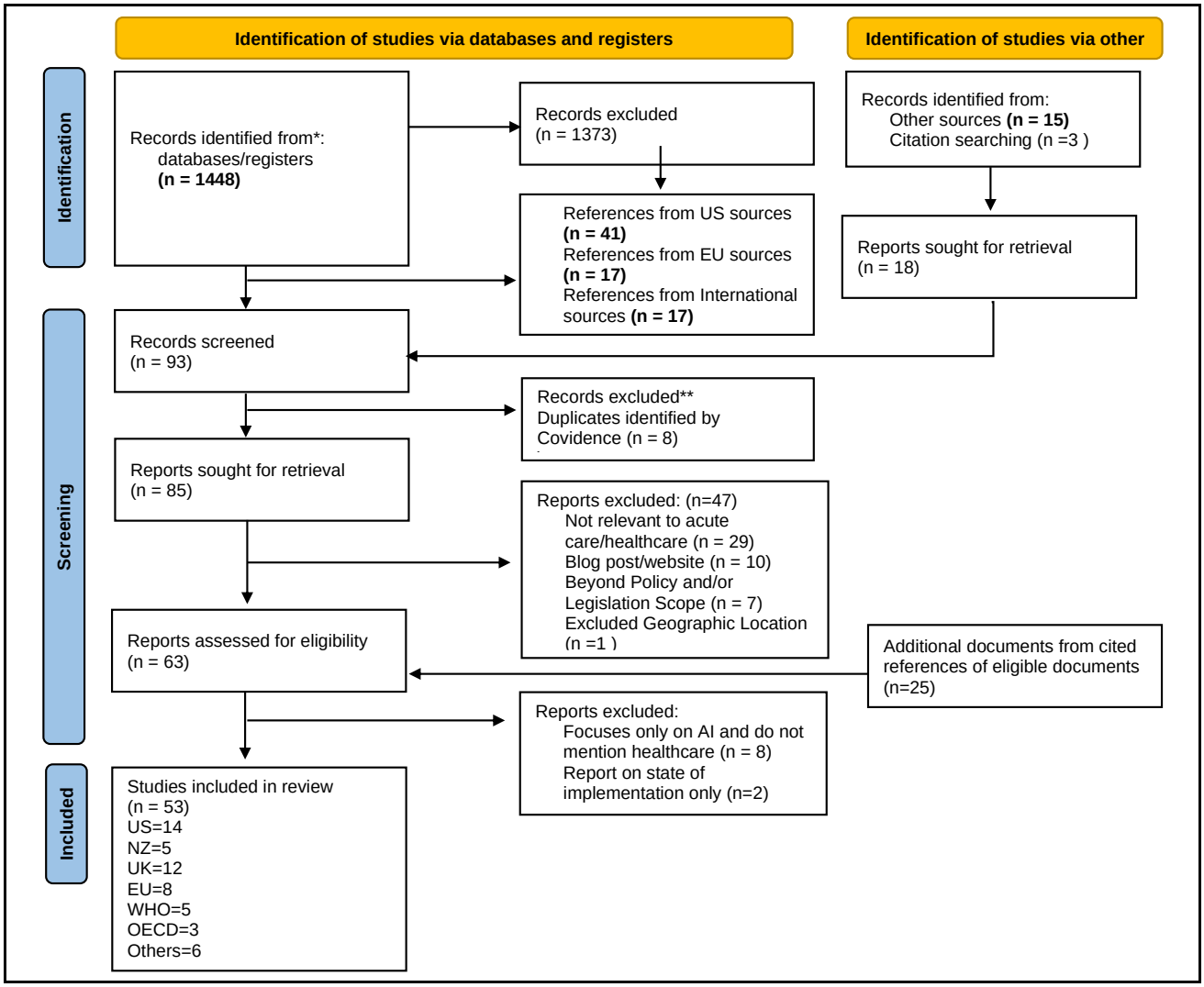
2.2.3 Screening

Initial searches yielded a total of 1,448 documents, with only 93 eligible for initial screening and retrieval. Data were managed in Excel, with details including the title and a link to the full text. After removal of duplicates (n=8), the full text of the remaining documents (n=85) was retrieved and screened based on title and executive summary (or introductory text if a summary was not available). Three researchers screened 10 randomly selected documents to establish a shared understanding of implementing the eligibility criteria. One researcher screened the remaining documents.

A total of 62 documents were eligible for full text screening. Three researchers screened five randomly selected documents in full, compared decisions and discussed to ensure clarity on the eligibility criteria. to be included in the review. One researcher screened the remaining documents.

During the full text screening, additional cited references (n=23) were identified for inclusion after screening against the same eligibility criteria. After this process, 48 documents were included in the review (See Figure 1).

Figure 1: Results of the search process.



2.2.4 Data extraction

We extracted data against pre-determined themes and also developed new themes inductively from the data in consultation with the ACSQHC. See Table 4 for the pre-determined themes and features extracted from the documents.

Table 4: List of features/themes for extraction.

Category	Features
• General features	Agency/organisation that authored the document implementing recommendations
	Agencies or organisations impacted by the document, or responsible for implementation
	Year of publication/release
	Jurisdiction (intergovernmental, national/country, state, district)
	Type of document • Legislation • Non-legislative regulatory instrument (e.g. TGA SaMD) • Policy

Category	Features
	• Guidelines • Principles • Organisational policy or position statement • Other
	Extent of relevance to acute healthcare • Document is exclusively about acute care • Document is about healthcare services in general, including acute care • Document is about all AI applications across industries and domains, including healthcare (sector-agnostic)
• Principles of implementation, such as governance, quality assurance, incident management etc.	Governance and Regulation of AI in acute care - structures, systems and principles
	Engagement with consumers, patients and citizens
	Equity, discrimination and human/patient rights
	Privacy and confidentiality
	Evaluation, monitoring and maintenance as an issue for governance
	Transparency
	Accountability and liability
	Consent considerations
	Worker training and support
	Cybersecurity
	Guidance specific to pathology tests and medical imaging
• Miscellaneous or not categorised	Key themes arising from data not included in the list

2.3 Findings

In aggregate, 53 documents were selected for inclusion in the review (see full list in Appendix B). We identified documents from the US (n=14), UK (n=12), New Zealand (n=5), Canada (n=2), and Singapore (n=4). We identified documents from intergovernmental organisations, namely EU (n=8), WHO (n=5) and OECD (n=3). The selected documents were read and analysed to extract information relevant to the research question (see Data Extraction Table 4 above).

The reviewed documents fell into different categories of enforcement, ranging from laws and regulations to non-legislative guidance documents. The non-legislative guidance documents include policies, strategy papers, position statements, discussion papers, frameworks, and guidelines.
In terms of scope and direct relevance to acute care, most of the documents were sector- and application-agnostic but are applicable to healthcare (e.g., data privacy laws). The rest of the documents are specifically designed for AI applications in healthcare, including acute care.

2.3.1 General principles for implementation of AI in healthcare

Legislative approaches to governance

The use of legislation (mandatory and binding instruments) to govern the implementation of AI across industries was observed in documents from the US, UK, Canada and EU (see Table 5 for a list of all pending or enacted legislation). US documents included state Acts or Bills prohibiting AI discrimination (16-18), a national Bill establishing the National Artificial Intelligence Initiative (19), an executive order (20), and a presidential memorandum (21). The EU introduced the Artificial Intelligence Act (22), which aims to

harmonise rules, prohibitions, requirements, obligations, and enforcement of human centric and trustworthy artificial intelligence in the EU. Canada's Bill C-27 (23) included enacting the *Artificial Intelligence and Data Act*, which regulates international and interprovincial trade and commerce in AI systems across sectors by requiring individuals to adopt measures that mitigate AI risks or harms (see Appendix A for the impact assessment levels).

Table 5: List of all AI-related* legislation mentioned in reviewed documents.

Jurisdiction	Title	Enacting statement
New Jersey, US	Bill S1402 (16)	An Act concerning discrimination and automated decision systems and supplementing P.L.1945, c.169 (C.10:5-1 et seq.).
District of Columbia, US	Bill 25-0114 (17)	To prohibit users of algorithmic decision-making from utilising algorithmic eligibility determinations in a discriminatory manner, to require corresponding notices to individuals whose personal information is used, and to provide for appropriate means of civil enforcement.
New York, US	A.3308/S.2277 (18)	Will enact the 'digital fairness act' focusing on the handling of personal information in the digital context and the use of automated decision systems to make core government and business decisions.
Federal, US	87 FR 47824 (24)	The proposed Rule, Non-discrimination in Health Programs and Activities, revises the implementing regulation for Section 1557 of the Affordable Care Act (25), and proposes robust provisions that will be more effective in protecting people from discrimination, with specific reference to clinical AI algorithms.
Federal, US	89 FR 1192 (26)	Final Rule on Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing
Federal, US	HR 6216 (19)	A bill to establish the National Artificial Intelligence Initiative, and for other purposes
EU	AI Act (22)	Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts
UK	HL Bill 11 (27)	The Artificial Intelligence (Regulation) Bill [HL] is a bill to make provision for the regulation of artificial intelligence; and for connected purposes
Canada	Bill C27 (23)	An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts

*We define AI-related legislation as laws (both pending and enacted) that mention AI in the title or enacting statement, or mention AI as use case in the main body of the document.

Below we describe the key themes and issues arising from the legislative documents, and these were risk-based assessment, regulatory oversight and authorities, impact assessment and AI system use policies, transparency, and discrimination.

Risk Based Assessment

The EU Parliament passed the AI Act (22) on 13 March 2024 with the intention of creating a uniform legal framework for developing, marketing, deploying, and using AI systems across the EU. It is, perhaps, the most comprehensive set of principles developed to date in the field. The Act takes a risk-based approach to regulating AI with four levels of risk identified: unacceptable risk; high risk; limited risk; and minimal

risk. Limited risk systems are lightly regulated to ensure transparency, that is, that users are aware that they are interacting with an AI. Minimal risk systems, for example, AI enabled video games and spam filters, are not covered.

The EU AI Act focusses most attention on high-risk systems that may have a significant harmful impact on the health, safety, and fundamental rights of people in the EU [AI Act Rec 46]. While most of the obligations in the Act fall on the *developers* of high-risk systems, some obligations are imposed on *users*, that is, those who *deploy and use* AI systems. AI systems are considered high-risk for a number of reasons under the Act that may impact in the critical care environment. For example, AI systems used to classify emergency calls, or to dispatch emergency services, and emergency healthcare triage systems, are explicitly categorised as high risk, 'since they make decisions in very critical situations for the life and health of persons' [AI Act Recital 58]. The Act will also apply where the AI system is itself, or is a safety component of, a medical device or an in vitro diagnostic medical device.

In relation to such high-risk AI systems, the Act requires users to put in place risk-management systems that consist of 'a continuous, iterative process that is planned and run throughout the entire lifecycle of a high-risk AI system. This process should be aimed at identifying and mitigating the relevant risks of AI systems on health, safety, and fundamental rights' [AI Act Recital 65]. The system should be regularly reviewed and updated to identify risks or adverse impacts and to ensure that mitigation measures are put in place to deal with known and foreseeable risks, including possible risks arising from the AI system and the environment within which it operates. Other requirements imposed on those who deploy and use high-risk AI systems include the need to keep accurate records; the need to ensure the AI literacy of staff dealing with the AI system; that AI systems are used in accordance with instructions; and the need for competent and trained human oversight of the system [AI Act Recs 66 & 91].

Canadian Bill C-27 (23), the *Digital Charter Implementation Act 2022*, is a Government Bill currently under consideration by the Canadian House of Commons. Part 3 of the Bill sets out the Artificial Intelligence and Data Act, which is intended 'to regulate international and interprovincial trade and commerce in artificial intelligence systems by requiring that certain persons adopt measures to mitigate risks of harm and biased output related to high-impact artificial intelligence systems' [Bill C-27 Summary]. The Government has indicated that a high-impact system will include AI applications in healthcare and emergency services, but excluding certain medical devices already covered by the *Food and Drugs Act*. [Letter from the Minister of Innovation, Science and Industry to the Standing Committee on Industry and Technology, House of Commons, Canadian Parliament, October 2023, 4] The 'person responsible' is defined to include a person who manages the operation of an AI system. The Bill requires a person who is responsible for a high-impact system to establish measures to identify, assess and mitigate the risks of harm or biased output that could result from the system [Bill C-27 clause 8]. The person must also monitor compliance with the mitigation measures. [Bill C-27 clause 9] The manager of the operation of the high-impact system must publish on a publicly available website a plain language description of the system, how it is used, the content it generates, and the mitigation measures established under clause 8 [Bill C-27 clause 11]. The Bill also provides for oversight by the relevant Minister and for certain criminal offences.

Regulatory and Oversight Authorities

The National Artificial Intelligence Initiative Act [HR 6216] (19) was introduced to the US federal House of Representatives in 2020 and is still under consideration. The Bill seeks to establish a National Artificial Intelligence Initiative Office and the National Artificial Intelligence Advisory Committee to advise the President and the Office on matters related to the initiative. The Bill also provides for the funding of research initiatives into AI systems and their future impact.

A Private Members Bill, the Artificial Intelligence (Regulation) Bill (27) was introduced into the House of Lords in the UK in November 2023. It proposes the establishment of an AI Authority with a range of responsibilities in relation to: coordinating the response of the UK Government to AI developments; assessing and monitoring risks to the economy arising from AI; providing education and awareness to businesses and individuals; and promoting interoperability with international regulatory frameworks. The Bill includes a set of regulatory principles set out in clause 2. As a Private Members Bill these proposals are unlikely to proceed. On 6 February 2024, the UK Government submitted a new policy proposal for AI regulation to the UK Parliament called *A Pro-innovation Approach to AI Regulation: Government Response* (2024). This is a more authoritative source for the approach of the UK Government and is discussed further below.

Impact Assessments and AI System Use Policies

The Digital Fairness Act A.3308/S.2277 (18) is a Bill introduced into the New York State Assembly in 2023. The Bill focuses on the handling of personal information in the digital context and the use of automated decision systems to make core government and business decisions. The Act requires public agencies that deploy or use automated decision systems to engage a neutral third party to conduct an impact assessment of existing and new systems and to publish that assessment for public comment. The assessment must include the costs and benefits of the system; the risks of harm and the measures the agency will employ to minimise the risks; and an evaluation of the development processes, design, and training data to identify the risk of inaccurate, unfair, biased, or discriminatory decisions impacting individuals.

Under the New York Bill, public agencies are also required to publish an Automated Decision System Use Policy that includes information on the capabilities of the system; the decisions that the system is used to make; any rules, processes, or guidelines the agency employs in relation to the use of the system; safeguards and security measures; training requirements; and audit and oversight mechanisms. The Act also provides a mechanism for seeking review of an automated decision by a human decision maker and seeking remedies where harm has been caused by the decision.

Transparency

Rule 89 FR 1192 has been developed by the Office of the National Coordinator for Health Information Technology (ONC), which is the 'principle federal entity charged with coordinating nationwide efforts to implement and use advanced health IT and to facilitate the electronic exchange of health information' [EHI, p 1193]. The Rule is intended to 'promote the responsible development and use of artificial intelligence through transparency and improves patient care through policies that advance standards-based interoperability and EHI exchange, which are central to the Department of Health and Human Services' efforts to enhance and protect the health and well-being of all Americans'. [p 1193] The ONC is responsible for implementing the Health IT Certification Program. Rule 89 'adopts new and revised standards and requirements for the certification of health IT under the Program.' [p 1193]. These new certification criteria apply to decision-support interventions. Rule 89 aims to more clearly distinguish between developing or modifying the technical functionality of a decision-support intervention and ongoing maintenance of that intervention. An example of a new technical functionality would be providing patients with a new capability to use an internet-based method—such as a free text box or check box—to request a restriction on certain uses and disclosures of their EHI [p 1418]. This is distinguished from ongoing maintenance required for Maintenance of Certification of the intervention, for example, providing a description of the process and frequency by which the intervention's validity is monitored over time [p 1268]; reviewing and updating, as necessary, risk management practices and summary information provided to the ONC via publicly accessible hyperlink [p 1254].

Discrimination

Some of the proposed legislation identified in Table 5 has a narrower focus, dealing specifically with the issue of discrimination as a result of the use of an AI system to make automated decisions. Proposed Rule 87 FR 47824 is an amendment to the implementing regulation for s 1557 of the federal US Affordable Care Act, which prohibits discrimination on the basis of race, colour, national origin, sex, age or disability under federal programs. The proposed rule includes provisions that prohibit discrimination through the use of clinical algorithms in decision making. This provision was prompted by the US Department of Health and Human Services becoming 'aware that clinical algorithms in state Crisis Standards of Care plans used during the COVID-19 pandemic may be screening out individuals with disability.' [Proposed Rule 87 FR 47824 clause 92.210] The amendments are intended to put covered entities on notice that they cannot use discriminatory clinical algorithms and may need to make reasonable adjustments in the use of the algorithms. Covered entities may be held liable for decisions made in reliance on clinical algorithms used for screening, risk prediction, diagnosis, prognosis, clinical decision-making, treatment planning, health care options, and allocation of resources. Decision makers are encouraged to consider such algorithms a tool that supplements their decision making, rather than a replacement of their clinical judgment.

Bill S1402 of the State of New Jersey specifically prohibits healthcare providers from discriminating through the use of an automated decision system. This is of interest because AI systems can discriminate where they are developed and trained using incomplete data sets. While this may appear to be an issue for developers, rather than users, Bill S1402 imposes liability for the resulting discrimination on the user of the system.

Bill 25-114 of the District of Columbia also deals with the prevention of discrimination but in the specific context of the use of algorithms. The Bill notes that algorithms have the potential to amplify discrimination based on characteristics such as race, gender, sexual orientation, disability, and age. The Bill aims to bring additional transparency and accountability to the use of algorithmic decisions by, for example, imposing notice requirements and auditing and reporting requirements. The Bill also prohibits adverse algorithmic decision-making based on protected characteristics. Again, liability for prohibited discrimination is imposed on the user of the algorithm, rather than the developer of the algorithm.

Existing non-AI laws and policies

Another general approach to governance of AI implementation in healthcare is to explicitly reference existing primary legislation and policies. Across countries, government agencies noted that existing data privacy laws are applicable to AI applications in healthcare. UK agencies, for example, cited the Data Ethics Framework, Data Protection Act of 2018, EU's General Data Protection Regulation, Human Rights Act of 1998 and Consumer Rights Act. Similarly, New Zealand agencies (28-30)cited the NZ Privacy Act 2020, Health Information Privacy Code, and the Te Tiriti o Waitangi (known in English as the Treaty of Waitangi).

Ethical frameworks

Appeal to ethical principles and frameworks is a common approach to the development of guidance documents. In UK's *Artificial Intelligence: How to Get it Right* (31), the National Health Service (NHS) recommended a governance framework that combines ethics and regulation, stating that regulations alone are not sufficient in addressing sensitive areas of healthcare, particularly when it comes to AI's impact on individuals, groups, systems or whole populations. In *Capturing the benefits of AI in healthcare for Aotearoa New Zealand* (28), the New Zealand Government developed 22 recommendations to assist healthcare AI policymaking based on 17 ethical principles. The European Commission's *Ethics Guidelines for Trustworthy AI* (32) included guidance on how to operationalise ethical principles, such as respect for

human autonomy, prevention of harm, fairness and explicability. These guidelines are cited in the EU AI Act as complementary to the legally binding requirements of existing EU laws and regulations. This finding suggested that in many jurisdictions, the working assumption was that ethics and regulation should work hand in hand when responding to the challenge of AI.

Centralised governance of AI in healthcare

The section above on legislative approaches noted that some countries have had bills introduced into their parliaments proposing dedicated national authorities to oversee AI. In a healthcare context, some countries have moved to establish a new body within their healthcare systems to centralise and coordinate the development and implementation of policies to govern AI. In the US, the Office of the Chief AI Officer within the Department of Health and Human Services facilitates effective collaboration on AI efforts across HHS agencies and offices (1). In the UK, the NHS established the AI Lab to determine guidance and regulations for developing and deploying AI systems in healthcare (31).

Regulatory sandboxes

Using Machine Learning in Diagnostic Services (33) reported findings from the UK's Care Quality Commission regulatory sandbox pilot. The Commission defined regulatory sandboxing as a practical approach to understand new types of health and social care service, establish criteria for good quality, and develop approaches to regulation. The document included information about how the Commission will carry out inspections and service ratings of providers that use ML in diagnostic services.

Procedural tools

Procedural tools mentioned in or recommended by the reviewed documents can be broadly classified as either risk assessment tools (echoing the risk-based assessment approach to legislative governance described above) or question-based checklists (see Table 6). Risk assessment tools are one of the most common procedural instruments to guide proper implementation of AI. Canada's *Directive on Automated Decision-Making* (34) included the Algorithmic Impact Assessment tool, which distinguishes four levels of risks. Results of the assessment are required to be completed prior to production of any AI system and updated on a scheduled basis, with assessment results released through the Government of Canada websites. New Zealand's *Emerging Health Technology: Introductory Guide* (30) included an Ethics and Algorithms Toolkit, a one-page checklist adopted from a similar toolkit developed by the Center for Government Excellence at Johns Hopkins University, Harvard DataSmart, and Data Community. The toolkit guides individuals building or acquiring algorithms through a series of questions to help understand and minimise the ethical risks. Unlike Canada's impact assessment that provided details on when and how to implement the toolkit, New Zealand's toolkit did not provide further details.

Question-based procedural checklists generally go beyond risk assessment. The European Commission (32) recommended the Trustworthy AI Assessment List, a tool consisting of guide questions under 7 categories and designed to guide AI practitioners (including implementers) to achieve trustworthy AI. The document described trustworthiness of AI using three components, namely lawful (complies with existing law and regulation), ethical (adheres to ethical principles) and robust (safe, secure and reliable). In Singapore, Infocomm Media Development Authority developed *AI Verify* (35) as an online, open-source toolkit to help end-user organisations to validate the performance of their AI systems against a set of 11 ethics principles grouped into five focus areas. The principles are assessed through a combination of technical tests and/or process checks.

Table 6: List of procedural tools.

Type of tool	Name	Source document	Scope
Risk assessment	Algorithmic Impact Assessment	Directive on Automated Decision-making (Canada) (34)	All sectors
	Ethics and Algorithms Toolkit	Emerging Health Technology: Introductory Guide (New Zealand) (30)	Health sector
	Risk Management Framework	EO 14110 Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (US) (20)	All sectors
Question-based checklists	Trustworthy AI Assessment List	Ethics Guidelines for Trustworthy AI (EU) (36)	All sectors
	Implementation and Self-Assessment Guide for Organisations	Companion to the Model AI Governance Framework (Singapore) (37)	All sectors
	Testing Framework	AI Verify: AI Governance Testing Framework and Toolkit (Singapore) (35)	All sectors
	Human-centred Implementation Protocol	Understanding Artificial Intelligence Ethics and Safety (UK) (38)	All sectors
	Assessment Template	A Buyer's Guide to AI in Health and Care (UK) (39)	Health sector
	Trustworthy AI Playbook	Trustworthy AI Playbook (US) (40)	Health sector

2.3.2 Key insights by country

A comparative overview of insights on recent approaches to implementation of AI in healthcare in United States, United Kingdom, New Zealand, Canada and Singapore are provided below.

United States

We reviewed 14 documents from the US, nine of which were legislation. A diverse collection of initiatives and legislative actions had been undertaken at various levels of government within the United States regarding AI and its applications. At the federal level, federal agencies such as the Food and Drug Administration (FDA), the Department of Health and Human Services (HHS), Congressional Research Service, the Government Accountability Office (GAO), and the Office of Management and Budget (41) issued guidelines, action plans, and regulations aimed at ensuring the responsible development and deployment of AI technologies. At the national level, President Joe Biden's Executive Order of October 30, 2023 (20) emphasised the safe, secure, and trustworthy development and use of AI. In particular, President Biden's Executive Order initiated a government-wide effort to guide responsible AI development and deployment through coordinated actions across various federal agencies, regulation of industry, and engagement with international partners.

Legislative efforts in the US generally focused on requiring impact assessments and reporting for AI or automated decision systems in critical areas such as healthcare. Legislation such as New York State's *Digital Fairness Act* (18) requires certain organisations to perform impact assessments of automated decision systems such as those using AI/ML or augmented critical decision processes. Similarly, the (federal) AI Accountability Framework (42) suggests the use of Privacy Impact Assessments (PIAs) by federal agencies to assess the privacy implications of an AI system's use of data. The presidential memorandum M-21-06 (21) encourages executive departments and agencies to prioritise performance-

based approaches applicable to all AI technologies that foster innovation while ensuring responsible AI development.

Four US documents (1, 24, 26) were specifically aimed at healthcare services, and all were developed or released by the Department of Health and Human Services (HHS). The AI Strategy (1) included a proposal to establish an AI Council and AI Community of Practice with HHS to support AI governance and implementation across HHS divisions, and in alignment with federal legislation and policy. The Trustworthy AI (43) Playbook (1) was developed by the HHS Office of the Chief AI Officer to help HHS divisions meet federal requirements on designing, developing, acquiring and using trustworthy AI. The playbook provided high-level information about the principles that underpin trustworthy AI, and detailed guidance for leadership teams within HHS planning to deploy AI.

US documents addressed the importance of mitigating bias and discrimination in AI systems used in healthcare. Guidance on ensuring protection against discrimination due to AI is included in legislation, such as the federal rule on *Nondiscrimination in Health Programs and Activities*,(24) and state legislation *Digital Fairness Act* (18) and *Stop Discrimination by Algorithms Act*(17).

United Kingdom

UK documents included 1 piece of pending legislation [*HL Bill 11* (27)] and 11 non-legislative guidance documents. Six of the documents were general purpose and covered AI applications across all industries, while the remaining five documents were specific to healthcare applications of AI.

Of direct relevance to healthcare services, including acute care, were documents released by the Care Quality Commission (33) and the National Health Service [NHS, (31, 39)], typifying governance approaches that rely on non-legislative instruments (e.g. policy papers, guidelines, codes of conduct). In addition, UK documents cited existing cross-sectoral legislation and policies, including the GDPR, Data Protection Act of 2018, Data Ethics Framework, Consumer Rights Act, and the Human Rights Act of 1998, among others.

The document *A pro-innovation approach to AI regulation: government response* (44) is the most recent, high-level statement of the UK Government's proposed regulatory approach relevant to any AI implementation. The approach included cross-sectoral principles (e.g. safety, transparency, fairness, and accountability), a context-specific framework, voluntary measures for developers, and a commitment to international leadership and collaboration. The *Bletchley Declaration* (45) an outcome of the AI Safety Summit 2023, demonstrated UK's commitment to work with the international community to ensure safe and responsible AI development. The declaration was signed by all 27 countries attending, including the UK, the US, Australia, China, as well as the European Union.

The UK's approach to governance of AI implementation and deployment in healthcare demonstrated a high level of coordination. Coordination was evidenced by the establishment of the NHS AI Lab to lead the development and implementation of the national strategy for AI in health and social care (31, 33, 39). Documents released by UK health agencies consistently identify relevant authorities that govern various stages of AI implementation (31). In *Artificial Intelligence: How to get it right* (31), the NHS provided a regulatory journey map that summarises relevant statutory bodies and stakeholders including the NHS, MHRA, General Medical Council, Information Commissioner's Office, National Data Guardian, and the Care Quality Commission.

Out of the 9 documents, the most relevant to acute care and healthcare providers was the *Buyer's Guide to AI in Health and Care* (39) published by the National Health Service. The guide consisted of 10

questions grouped under four categories (problem identification, product assessment, implementation considerations, and procurement and delivery) to assist healthcare organisations to undertake robust procurement exercises and make well-informed buying decisions about AI products.

New Zealand

There were 5 documents included from New Zealand(28-30, 46, 47), all non-legislative guidance. Three of the documents are specific to individuals and organisations working in the New Zealand health system: the Ministry of Health's *Emerging Health Technology: Introductory Guidance* (30), Department of the Prime Minister and Cabinet *Capturing the Benefits of AI in Healthcare for Aotearoa New Zealand* (28), and Health New Zealand - Te Whatu Ora's *Advice on the Use of Large Language Models and Generative AI in Healthcare* (46).

While there was no formal AI-specific legislation found in the search, the documents cited existing legislation and policies that apply to AI in healthcare. These include Pae Ora (Healthy Futures) Act 2022, Privacy Act 2020, Health Information Privacy Code 2020, and Therapeutic Products Act 2023, among others.

The reviewed documents shared a strong call for social licence prior to implementation. In *Emerging Health Technology: Introductory Guidance* (30), the Ministry of Health defined social licence as "an organisation's or project's legitimacy, credibility and trust in the eyes of the public or key stakeholders." The guidance further stated, "It is important to consider how the development of an algorithm conforms with the expectations from the Treaty of Waitangi and expectations from Māori regarding Data Sovereignty, and the opportunity that an algorithm can present for improving equity and outcomes for Māori."

In addition, New Zealand guidance documents placed strong emphasis on ensuring that AI applications lead to equitable outcomes. The *Emerging Health Technology: Introductory Guidance* (30) recommended several strategies to mitigate bias, including reducing data elements, population matching between training data and New Zealand local population, use of explainable algorithms, and implementation of data standards (e.g. Systematised Medical Nomenclature for Medicine–Clinical Terminology). Reducing the number of data elements/classes was a way to increase the algorithm's understandability. No further details were provided to guide which data elements need to be retained.

Specific regulatory tools mentioned in the documents include use of an Ethics and Algorithms Toolkit, Privacy Impact Assessments, commissioning an independent ethical review, and assurance frameworks through peer-review.

Other countries

References in Australia's *Safe and Responsible AI* discussion and consultation papers (2, 15) yielded additional documents from Canada and Singapore, as well as potentially relevant legislation and policies from Thailand and China. As there were no official English translations of the Thai and Chinese documents, these were excluded from the review.

Canada

Two documents from Canada cited by the *Safe and Responsible AI* papers were included, with one being a directive (34) and one being a bill (23) currently being considered in the House of Commons. Both documents offered regulatory guidance on general use of AI across industries, which may include healthcare applications.

The documents promoted both the use of impact assessment to reduce risk and the promotion of open governance. In the *Directive on Automated Decision- Making* (34), the Government of Canada recommended using an Algorithmic Impact Assessment framework, which consisted of four levels of impact based on potential risks, as well as administrative requirements depending on the level of impact.

Both documents showed the Government of Canada's commitment to transparency and open governance across stages of AI implementation. The *Directive* (34) for example, included recommendations such as providing end users information about use of automated-decision making systems, explanation of the AI systems, information about the systems' components (including source codes), and evidence of effectiveness and efficiency.

Singapore

Three documents from Singapore cited in the *Safe and Responsible AI* papers were included. All documents were non-legislative guidance documents,(35, 37, 48, 49) with only one was specific to implementation of AI in healthcare. The *Artificial Intelligence in Healthcare Guidelines*(49) released by the Ministry of Health set out responsibilities for developers and implementers, with recommendations on clinical governance, operational workflows, end-user communication, post-deployment monitoring and review protocols.

All documents relied on non-legislative approaches to governance, ranging from a self-assessment toolkit to self-regulation of implementation strategies. However, the documents cited a range of legislation in Singapore that extended to implementation of AI across sectors, including healthcare. This legislation included the Personal Data Protection Act of 2012, Health Products Act, Private Hospitals and Medical Clinics Act, Civil Law (Amendment) Bill 2020, and Professional Registration Acts, among others. The documents showed that agencies in Singapore shared a strong emphasis on operations management that covered evaluation, monitoring and maintenance as an issue for governance. For example, the Personal Data Protection Commission (PDPC) of Singapore published the second edition of the *Model AI Governance Framework* (48), which included a set of best practices to operationalise principles such as repeatability (consistency in performance), robustness (ability of a computer system to cope with errors during execution), traceability (process documentation), and auditability. A companion to the framework, called *Implementation and Self-Assessment Guide for Organisations* (37), was intended to guide all types of organisations that procure and deploy AI solutions for purposes including providing service to consumers or improving operational efficiency.

In addition, the documents showed government agencies were interested in providing actionable guidance on how to implement internal governance structures. The PDPC recommended adapting existing internal governance structures (e.g. enterprise risk management structures or ethics review boards), establishing new structures that specifically address AI implementation, and specifying key roles and responsibilities that can be allocated to personnel or departments having internal AI governance functions (48). In *Artificial Intelligence in Healthcare Guidelines* (49), The Ministry of Health recommended communication strategies to ensure transparency and improve clinical and public trust (see Case Study at the end of this chapter for an illustration of how transparency can work in practice). The ministry specified the responsibilities of implementers (i.e. licensed healthcare service providers), as well as the licensing requirements set out under a range of legislation (e.g. Private Hospitals and Medical Clinics Act). In response to the gaps in existing End User Licensing Agreements, the Ministry recommended that implementers should establish Service Level Agreements with developers to set clear and mutually agreed responsibilities that cover areas such as testing, monitoring, reviewing, and accessing algorithmic design.

2.3.3 Key insights from intergovernmental organisations

Insights on general principles of implementing AI in healthcare from EU, WHO and OECD are provided below.

European Union

The European Union (EU) documents included in this study were seven documents comprised of six non-legislative documents and one key piece of legislation relevant to the governance of AI across industries. EU documents provided an insight into EU's horizontal approach to regulation of AI, wherein legislation and policy apply to all AI across all sectors and applications. Several EU documents (50, 51) emphasised the need for a common framework for AI governance among member states that ensures AI development and deployment uphold principles and values enshrined in the Treaties and the Charter of Fundamental Rights of the European Union (50).

While there were no documents specific to healthcare, most documents highlighted specific AI implementation considerations in the healthcare context (32, 50, 52). These considerations include health information as sensitive data, impact of AI on quality of care, managing risk of bias and inequity arising from AI, and clarification of legal responsibility in cases of harm (50).

EU documents provided strategies to operationalise ethics principles and frameworks. In *Framework of ethical aspects of artificial intelligence, robotics and related technologies* (53), the European Parliament suggested common criteria and application process relating to the granting of a European certificate of ethical compliance, which evaluates ethical considerations across the AI lifecycle. Other strategies included establishing ethics committees to oversee discussions and recommendations about ethical AI implementation (32), developing an AI Assessment Checklist (32), and incorporating AI ethics into AI standards and regulation (53).

The World Health Organization (WHO)

Five documents published by the WHO were included. They provided comprehensive insights into the ethical, regulatory, and governance considerations relevant to AI in healthcare. WHO is a United Nations intergovernmental agency with 194 member states. WHO leads global efforts to promote health and health care and coordinate responses to health emergencies. WHO documents addressed aspects of AI implementation in healthcare including the importance of ensuring ethical AI practices (54), regulatory compliance (55), evidence-based validation and evaluation frameworks (56), and stakeholder engagement in the development and deployment of AI technologies in healthcare settings (57).

The reviewed documents highlighted WHO's collaborative approach to the governance and implementation of AI in healthcare. In *Regulatory considerations on artificial intelligence for health (55)*, WHO outlined key principles that governments and regulatory authorities can follow to develop new guidance or adapt existing guidance on AI in healthcare. A potential approach to governing AI on a global scale is "networked multilateralism" as proposed by the United Nations Secretary-General in 2019, which emphasised collaboration among a wide range of stakeholders (55).

As well as AI-specific regulation and legislation, other legislation at the national and international level will be relevant for manufacturers and developers to consider in the development and deployment of AI in health care. Notably, the European Union's General Data Protection Regulation (GDPR) was frequently cited in WHO documents as a key example (54, 55). The GDPR establishes various rights for individuals, including protection from automated decision-making. In *Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models* (54), WHO highlighted the relevance of GDPR and other data privacy laws to AI governance broadly. The relevance of such legislation is apparent in

situations where, for example, medical providers integrate "protected health information" into large language model (LLM) chatbots, potentially violating laws such as the Health Insurance Portability and Accountability Act (HIPAA) in the US.

The WHO global strategy on digital health (2020-2025) emphasised accelerating the development and adoption of responsible digital health solutions. Australia is a key player in this global initiative. Australian regulatory agencies such as the Therapeutic Goods Administration (TGA) are actively engaged in recognising and leveraging AI's capabilities across various healthcare domains, alongside other prominent players like the US Food and Drug Administration (FDA), Health Canada, and the European Commission.

The Organisation for Economic Co-operation and Development (OECD)

We included three non-legislative guidance documents from the OECD. The OECD is an international organisation that aims to provide evidence-based policies and frameworks. The OECD's 38 member countries engage with OECD experts, using their data and analysis to inform policy decisions.
Similar to EU and WHO, the OECD encouraged a principles-based framework for the implementation of AI across sectors. For example, the OECD AI principles were established as a foundational framework for promoting ethical and responsible management of trustworthy AI while ensuring respect for human rights and democratic values (48-50). The five OECD AI principles were: inclusive growth, sustainable development and well-being; human-centred values and fairness; transparency and explainability; robustness, security and safety; and accountability.

OECD emphasised international collaboration to establish regulatory frameworks tailored to AI deployment in healthcare settings. This collaborative effort underscores the need for standardised guidelines to ensure responsible AI implementation and mitigate associated risks effectively. The document *Collective action for responsible AI in health*(58) highlights the current global health system fragmentation as a barrier to responsible AI innovation and implementation in health. In addition, OECD emphasised the role of involvement and stakeholder engagement in the development and deployment of AI in healthcare.

The documents offer significant discussion of human-centric and equitable AI implementation. In *Framework for the Classification of AI systems* (59), OECD provided components of a human-centric approach to AI implementation, such as identifying individuals and groups affected by the AI implementation, guaranteeing benefits for people, and upholding human rights and democratic values.

2.4 Case studies on transparency in practice

Some policy documents acknowledged the difficulty in complying with transparency, confidentiality or consent requirements. The GDPR (50), for example, stated that the obligation to inform the data subject is waived when compliance is impossible, requires a disproportionate effort or impairs the achievement of the objective of the process (such as providing care). A typology of transparency (see Table 7) and two examples may help illustrate how transparency is implemented in practice. The recommended actions were extracted from guidance documents released by the UK Department for Science Innovation and Technology (44) and Singapore Ministry of Health (49).

Table 7: Typology of transparency and recommended actions.

Transparency of what	Transparency for what purpose	Recommended actions for deployers
1. Transparency to support consent	To provide meaningful information to support informed consent, such as limitations of AI, adverse events, and alternative (non-AI) solutions (49).	1.1. Consent for data sharing: Seek patient consent on collection, use and disclosure of health or personal data to be reused for re-training the AI system (49). 1.2. Consent for using AI: Consent required when using AI system should be no different from consent taken for other medical procedures performed by actual physicians (49).
2. Transparency of using AI in patient care	To improve user and service recipient awareness and understanding of AI; ensure organisational accountability; enable adequate regulation of safety (44, 55).	2.1. Inform clinicians and patients that they are interacting with an AI system. Information should include evidence of effectiveness and limitations of the system (49). 2.2. Provide sufficient information about alternative options to using the AI system (49).
3. Transparency in data use	To ensure compliance with privacy and data protection laws, such as EU's GDPR (50).	3.1. Inform patients about whether the AI system will collect their health data to be reused for re-training the AI system (49). 3.2. If applicable, inform patients about the benefits and risks of sharing their health data (49).
4. Transparency with respect to governance (including performance of AI system and organisational governance structures)	To build confidence and trust; ensure interoperability; enable independent audits (44).	4.1. Centralised documentation or process log to consolidate information necessary to assure end-to-end auditability (44). 4.2. Consult with existing bodies whose roles include implementing transparency practices. These include consumer advisory groups or hospital ethics committees.

Case study 1: language translation application

RadTranslate™ (www.radtranslate.com), a web-based AI application, was developed to provide standardised audible imaging examination instructions to patients in their preferred language using a simple user interface (60). The app is compatible with desktop, laptop, phone or tablet.
The researchers implemented several strategies that can be considered as upholding transparency. First, the design and implementation was guided by an evaluation framework that included consideration of "acceptability", defined as the "perception among stakeholders that an intervention is agreeable, palatable or satisfactory". However, the pilot study report does not make clear how this was operationalised. Second, the researchers consulted with end users (nurses, doctors, clinical operations managers, and technologists) about clinical workflow, obstacles to productivity, and acceptability of technology-based solutions, increasing transparency to clinicians regarding use of the technology in line with items 2.1 and 2.2 in Table 7. The article (60) reporting the results of the pilot study did not mention if steps were taken to inform patients whether their healthcare data will be reused to train AI, or whether the research team informed physicians and patients about existing governance structures to enable external audits.

Case study 2: Live transcription application

Example 2: An AI scribe was developed and tested at the Permanente Medical Group, a multidisciplinary clinical group in Northern California (61). This was an experimental use of the technology with a strong governance and oversight framework. The technology uses ML to produce real-time transcripts of clinician-patient encounters to rapidly convert speech collected from microphones on clinicians' smartphones into text. The app then uses natural language processing techniques to summarise key clinical content.

The team implementing the application employed several strategies for transparency. First, physicians were trained on the effective and safe use of the technology to ensure clinicians were fully aware of the technology they were using (in line with 2.1 in Table 7). Second, physicians were educated on how to inform patients using a standardised template to support informed consent (in line with item 1.2 in Table 7). Third, the team developed patient-facing educational handouts containing information that was verbally summarised by physicians to inform patients they are interacting with an AI system (see 2.1 in Table 7). Fourth, the team offered information about their approach to documenting the evaluation of the limitations, errors, and reasons some clinicians did not use the technology (in line with 4.1).

2.5 Chapter summary

This chapter reported findings from a systematic review of 53 international legislation and policy documents that govern the implementation of AI in healthcare. These documents were published by government agencies in the US, UK, New Zealand, Canada, Singapore; as well as intergovernmental organisations, namely the WHO, OECD and EU.

The literature provided insights into general principles for implementation of AI in healthcare. Legislative approaches included creation of AI-specific laws (see Table 5), as well as explicit reference to existing primary non-AI laws (e.g. data privacy laws). Non-legislative approaches primarily appealed to ethical guidelines, frameworks, or principles to facilitate the implementation of AI in healthcare. Reviewed documents provided examples of procedural tools to guide developers and implementers of AI across industries, including healthcare. These tools can be broadly classified as either risk assessment tools or question-based checklists.

The review provides insights into variations in governance norms and practices across jurisdictions. For example, creation of new AI-specific laws was more common in the US than in UK and New Zealand. In New Zealand, guidance documents share a strong call for social licence and public engagement prior to implementation of AI in any industry or sector. Intergovernmental organisations such as WHO and OECD emphasise collaborative approaches to AI governance.

3. Australian policy environment

3.1 Introduction

The Australian Government has implemented several initiatives in response to the opportunities and risk associated with AI. In 2019, the Government released the Artificial Intelligence Ethics Framework (62), a voluntary framework to guide businesses and governments to responsibly design, develop and implement AI. In June 2021, the Australian Government published its AI Action Plan (63), which outlines the Government's strategies to be a global leader in "developing and adopting trusted, secure and responsible AI."

A recent report on the state of AI governance in Australia shows a general lack of systemic governance to identify and address AI-related risks, with very few laws that are directed expressly towards AI systems (64). It is important to note, however, that the development, deployment, and use of AI are regulated by a range of existing technology- and sector-neutral laws of general application, governing areas such as product liability, medical devices regulation, intellectual property, negligence, privacy, and discrimination. While these laws may not have been developed with AI in mind and may not specifically refer to AI, they do apply to the development, deployment, and use of AI systems generally and in the health and acute care sectors where relevant. Privacy legislation, for example, applies to the handling of personal information, including health information, in the development and training of AI and where personal information is collected and used for assessment or processing by AI systems. In considering the development of AI specific legislation, it will be important to carefully consider this pre-existing regulatory framework and to identify gaps, if any, that have arisen as a result of unique characteristics of AI.

In this study, we undertook a systematic review of Australian legislations and policies governing healthcare AI adoption and deployment in acute care, guided by two research questions:

1. What are the key themes from organisations and jurisdictions that have produced guidance for AI implementation in acute healthcare in Australia?
2. What are the gaps in existing and current Australian legislation, policies, and guidelines that impact on the implementation of AI in acute care?

3.1.1. Defining legislation and policy documents

There are two sources of binding law in Australia: the common law, made by courts, and legislation, made by Parliament and others authorised by Acts of Parliament to do so. The common law may be relevant to the implementation of AI in healthcare where, for example, a patient is injured as a result of an error made either by the developer, the AI system or the organisation or individual using the AI system. In this case, the patient may sue for negligence (under the law of tort). Developments in the common law of application are, however, outside the scope of this project.

Legislation is made at the federal, state and territory level and is generally binding within the relevant jurisdiction. The term legislation refers to the laws created by Parliament set out in writing in Acts, Regulations, and other legislative instruments. Where laws are developed by others authorised by an Act of Parliament to do so, the laws are referred to as 'delegated legislation' or 'legislative instruments'.

Government policy is developed by the Executive Government, that is, Ministers and their departments. As a matter of administrative law policies are not binding in the same way that law is binding. Policies

must, however, be consistent with the law. Policies are intended to provide guidance rather than binding legal rules. A range of other organisations, such as industry and professional bodies, also have policies.

Other governance instruments include guidelines and principles. Guidelines and principles can be found in legislation or policy and so may be binding or non-binding, depending on whether they have legislative force or not. Even where they may not have the same stature as accepted government policy or legislation, they may still be of interest where they include relevant principles and guidelines.

Legislation, policy, guidelines and principles potentially relevant to AI in acute healthcare in Australia were within scope for this report.

3.2 Literature search method

The search strategy consisted of identifying relevant documents from: cited references of two key documents (2), Australian federal and state legislative databases, Google Advance, websites of government and non-government organisations involved in healthcare, and secondary references from eligible documents. One researcher conducted the search between 8 January to 15 March 2024.

The search method and screening process was managed using a combination of:

- An Excel spreadsheet to record search results, capture details of potentially eligible documents, and remove duplicates.
- Covidence, an online systematic review software, to streamline screening of potentially eligible documents.
- Endnote, a reference management software, to store the full text of all documents included for review.

3.2.1 Eligibility criteria

See Table 8 for the inclusion and exclusion criteria.

Table 8: Eligibility criteria.

Inclusion criteria	Exclusion criteria
Legislation and policy, including guidelines and principles, of direct relevance to the use of AI in acute care.	Peer-reviewed literature, documents that fall outside the definition of policy and legislation, non-official documents (blog posts, news articles), policies or similar documents not about AI, policies or similar documents not relevant to acute care, community responses to discussion papers
Jurisdictions: federal, state and territory levels; government and non-government agencies who developed policy, guidelines or legislation that is relevant to acute care.	Exclude documents from organisations not directly relevant to the healthcare context
Published or available in English	Unavailable in English
Practice domain: Deal directly with AI and relevant to the acute healthcare context; Acute healthcare defined as hospital care for short-term/acute conditions	AI applications in other domains or industries that are not relevant for acute care (e.g. public health, billing), general standards (ISO); Not acute care, such as primary care, community care, or public health care that have no overlap with hospital care

Inclusion criteria	Exclusion criteria
Focus on processes involved in implementation of AI: deployment, acquisition, regulation, review, distribution, use	Exclude guidance on the development of AI directed to developers
Full text is accessible	Full text not accessible
Date of publication: from 2018 to present	Earlier than 2018

3.2.2 Document sources

References cited in the Safe and Responsible White Paper and Interim Response

We extracted all the references from the Australian Government's Safe and Responsible AI White Paper (2) and the separate Interim Response (15) based on the public consultation. The references from Australian organisations were recorded in a separate list in Excel, and included the following details: document title, country, and link to full text (if available).

Legislative databases

Australian legislative databases were searched (see Table 9 for the list) for documents using the same eligibility criteria in Table 8. Search terms included "artificial intelligence", "machine learning", "automated decision making", "algorithm" and "intelligent system". Results were recorded in Excel, and included the following details: database, title of document, country/jurisdiction, link to full text, and date of publication.

Table 9: Australian legislative databases searched.

Jurisdiction	Website
Federal	https://www.legislation.gov.au/
Federal, states and territories	https://www.austlii.edu.au/
Australian Capital Territory	https://www.legislation.act.gov.au/
New South Wales	https://legislation.nsw.gov.au/
Northern Territory	https://legislation.nt.gov.au/
Queensland	https://www.legislation.qld.gov.au/
South Australia	https://www.legislation.wa.gov.au/
Tasmania	https://www.legislation.tas.gov.au/
Victoria	https://www.legislation.vic.gov.au/
Western Australia	https://www.legislation.wa.gov.au/

Google Advanced Search

We used Google Advanced Search to search for Australian non-legislative documents, including policies, principles, guidelines and frameworks. The following advanced search strategy was implemented: published in 2018 or later, Australia as regional restriction, and no website domain restriction. Search terms included "artificial intelligence", "machine learning", "automated decision making", "algorithm" and "intelligent system", "health", "policy", "framework". If the search yielded fewer than 50 documents, all results were recorded in Excel. If there were more than 50 documents, we included all documents before saturation or repetition is reached (typically after fewer than 100 documents). Results were recorded in Excel, and included the following details: database, title of document, country/jurisdiction, link to full text, and date of publication.

Selected website search

We searched Australian organisations' websites for additional policy documents. We limited our search to key Government agencies involved in healthcare, and non-government organisations that responded to the Safe and Responsible AI consultation.

Each website was searched manually by using the website's search function and using search terms such as "artificial intelligence" or "automated decision making". Table 10 lists the websites systematically searched. All relevant documents were added to the Excel file containing results from the previous sources.

Table 10: Australian organisations' websites searched.

Organisation	Type	Website
AIHW	Government agency	https://www.aihw.gov.au/
Australasian Sonographers Association	Non-government organisation	https://www.sonographers.org/
Australian Alliance for Artificial Intelligence in Healthcare	Non-government organisation	https://aihealthalliance.org/
Australian Digital Health Agency	Government agency	https://www.digitalhealth.gov.au/
Department of Health and Aged Care	Government agency	https://www.health.gov.au/
Digital Health CRC	Non-government organisation	https://digitalhealthcrc.com/
Medical Board of Australia	Government agency	https://www.medicalboard.gov.au/
Medical Software Industry Association	Non-government organisation	https://www.msia.com.au/
Medical Technology Association of Australia	Non-government organisation	https://www.mtaa.org.au/
National AI Centre	Government agency	https://www.csiro.au/en/work-with-us/industries/technology/national-ai-centre
NHMRC	Government agency	https://www.nhmrc.gov.au/
RANZCR	Non-government organisation	https://www.ranzcr.com/
Therapeutic Goods Administration	Government agency	https://www.tga.gov.au/
Women With Disabilities Australia	Non-government organisation	https://wwda.org.au/

Cited references

During full-text screening for inclusion and data extraction, each eligible document's reference list was further screened for potentially relevant policy or legislation missed during the preceding methods of searching.

3.2.3 Screening

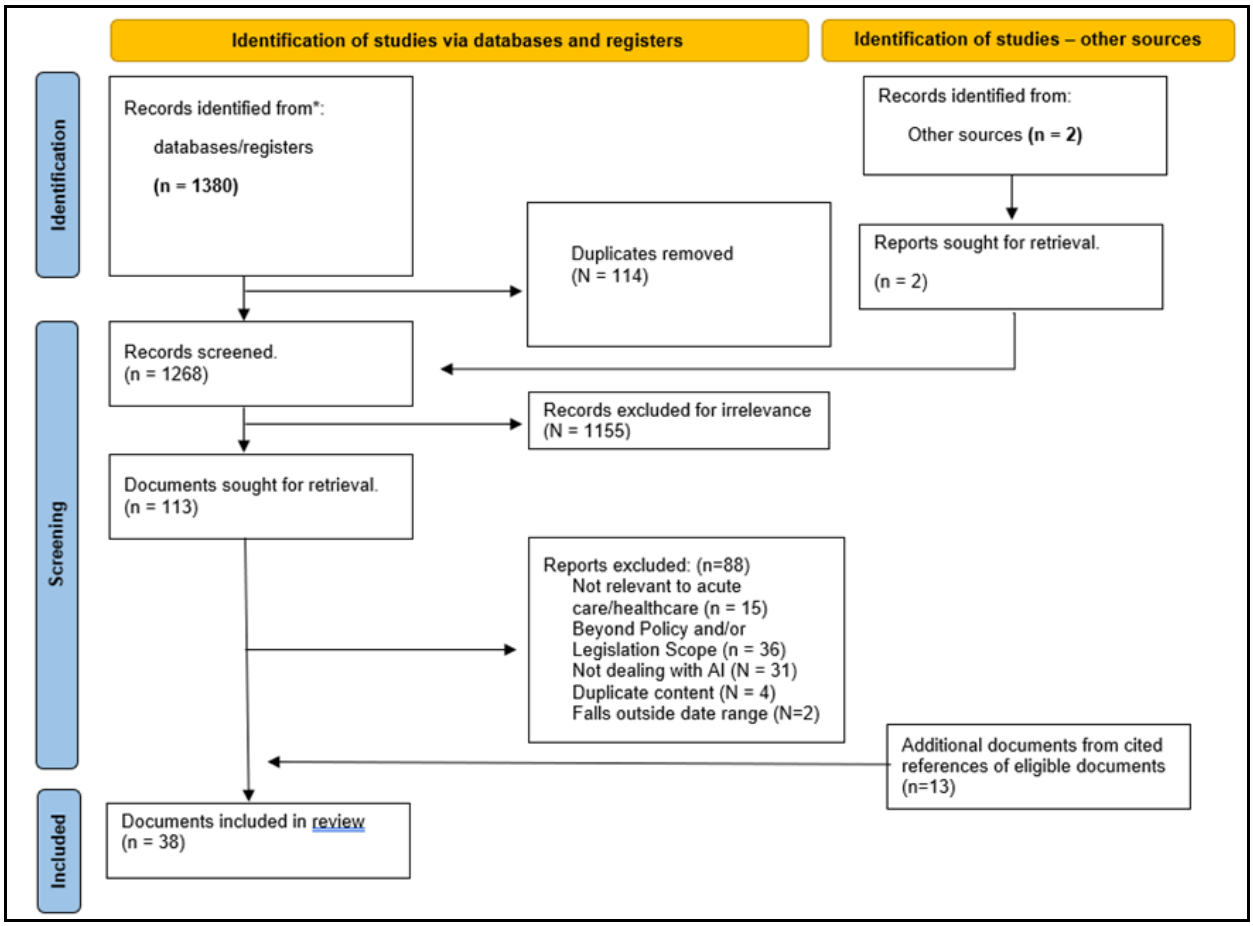
Initial searches yielded 1380 documents for initial screening. Two additional documents were found from other sources. Data were managed in Excel. After removal of duplicates (n=114), the remaining documents (n=1268) were screened based on title and executive summary (or introductory text if a summary was not available). Three researchers screened 50 randomly selected documents to establish a shared understanding of implementing the eligibility criteria. One researcher screened the remaining documents.

A total of 113 were eligible for full text screening. Three researchers screened 5 randomly selected documents in full, compared decisions and discussed to ensure clarity on the eligibility criteria. One researcher screened the remaining documents.

During the full text screening, additional eligible cited references (n=13) were identified and included in the review.

After this process, 38 documents were included in the review (See Figure 2 for the results).

Figure 2: Results of the search process (PRISMA).



3.2.4. Data extraction

We extracted data against pre-determined themes and also developed new themes inductively from the data in consultation with the Commission. See Table 4 in Chapter 2, section 2.2.4 for the pre-determined themes and features extracted from the documents.

3.3 Themes emerging from policy analyses

Our search for Australian policy documents ended on 1 March 2024. Several policy initiatives were underway at time of publication, including the current inquiry 'Artificial Intelligence (AI) in NSW', which was established by the NSW Parliament in June 2023 and is ongoing. These are not included as they had not yet been reported. Some Agencies have adjacent responsibilities, but had not released any specific policy or guidance about AI, so do not appear here. For example, the Office of the Australian Information Commissioner (OAIC) has regulatory responsibilities and powers under the Privacy Act 1988 and the Freedom of Information Act 1982 (FOI Act) and other legislation. The OAIC made a number of submissions during the study period (65), but did not release any primary policy or guidance, so are not included in this report.

Thirty-six documents were included in the review (See Appendix C for the full list). None of the included documents were legislative. Five documents (14%) provided direct guidance from the TGA on Software as a Medical Device (SaMD). The remaining thirty-one documents (86%) were policies, guidelines, or position statements. Fifteen of the documents (42%) reported on policies, positions or guidelines related to healthcare applications of AI specifically, and twenty-one of the documents (58%) described policies, positions or guidelines for general AI applications that were relevant to acute care settings.

In the following text, we discuss key issues arising across these documents. We first discuss structures, systems and principles of governance and regulation proposed in these non-legislative documents, then address 10 core areas of concern, comparing and contrasting approaches taken by different organisations in their policies ad guidelines.

3.3.1 Governance and Regulation of AI in acute care - structures, systems and principles

Based on the reviewed Australian policy documents, key themes in the governance and regulation of AI in healthcare focussed on cross-sectoral governance, regulation of software as medical device, and healthcare-specific approaches.

Cross-sectoral approaches to governance

The Safe and Responsible AI Discussion Paper, produced by the Australian Government Department of Industry, Science and Resources (DISR), acknowledges that Australia's regulatory landscape already regulates AI in various ways (2), via instruments such as the Privacy Act, Consumer Law, and through TGA regulations of Software as a Medical Device (SaMD). In addition, policies from State governments advised that the development and implementation of AI is guided by state-level instruments like the NSW Cyber Security Policy and the Victorian Data Security Frameworks (66, 67). Australia's approach to the regulation and governance of AI to date has not included the development of bespoke legislation for AI similar to the EU's AI Act (2).

The Australian Government, via their Safe and Responsible AI consultation process (2), have begun to develop a specific approach for Australia for regulating AI. The consultation process has supported the development of a national risk-based approach to regulation, where AI systems that are more likely to have negative consequences are subject to stronger controls and regulations (2). In their interim response to the consultation, the DISR commit to the development of an approach which adheres to the following principles:

- Balanced and proportionate - avoiding unnecessary burdens for lower-risk applications of AI.
- Collaborative and transparent - ensuring public and expert involvement.
- A trusted international partner - consistent with the Bletchley Declaration.
- Community First - putting people and communities at the centre.

Although the Department, via their interim response, acknowledges general support for a risk-based approach to regulation, it notes that operationalisation of a risk-based approach requires a strong risk assessment mechanism that is yet to be developed (2). Other policies have already begun implementing similar risk-based approaches, such as the TGA's SaMD regulation which contains exemptions for devices deemed to be lower risk (68), and the NSW Assurance Framework (currently under review), which is a risk-assessment tool that distinguishes between 'operational' and 'non-operational' AI based on whether the tool uses real-time data to make decisions that impact the community (67).

The policies and guidance included in this review contained strong recommendations that national frameworks for the regulation of AI should be developed and implemented in Australia. The Human Rights Commission recommended the introduction of legislation for human rights-based impact assessments before any government department uses AI-informed decision-making (69). The CSIRO's Data61 report on Australia's AI Ethics Framework recommends the development of cross-industry best practice guidelines to advise on gold-standard practices for the use of AI (70). Similarly, AATSE recommended that a single set of AI use standards should be developed to guide use of AI in the public sector, including in hospitals (25).

The Australian Government and NSW Government have both released similar sets of ethical principles for AI development and implementation with similar remits, as shown in Table 11 (71, 72). The intention of these principles is to provide a foundation for the development of more concrete mechanisms for ensuring that the use of AI is ethical and safe (71, 72). The Australian Government principles were endorsed by the Human Rights Commission who encouraged their use in human rights impact assessments (69), and were used by the Australian Government Architecture (AGA) (73) guidance on use of AI in the public sector (74). The NSW principles were developed as part of the NSW AI Strategy (R7) and underpin NSW's AI Assurance Framework (67). See Table 11 for both the NSW and Australian Government principles. A report by the Human Technology Institute notes that principle-based approaches to ethics on their own can be abstract and ineffective (64). In this report, the Human Technology Institute recommends that organisations should use such principles as baselines to explore more practical and applied governance strategies.

Table 11: Concordance of Australian Government and NSW Government AI Ethics Principles.

Australian Government principles (72)	NSW State principles (71)
"Human, societal and environmental wellbeing: AI systems should benefit individuals, society and the environment."	"Community benefit AI should deliver the best outcome for the citizen, and key insights into decision-making"
"Human-centred values: AI systems should respect human rights, diversity, and the autonomy of individuals."	
"Fairness: AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities or groups."	"Fairness Use of AI will include safeguards to manage data bias or data quality risks"
"Privacy protection and security: AI systems should respect and uphold privacy rights and data protection, and ensure the security of data."	"Privacy and security AI will include the highest levels of assurance. NSW citizens must have confidence that data used for AI projects is used safely and securely, and in a way that is consistent with privacy, data sharing and information access requirements."

Australian Government principles (72)	NSW State principles (71)
"Reliability and safety: AI systems should reliably operate in accordance with their intended purpose."	
"Transparency and explainability: There should be transparency and responsible disclosure so people can understand when they are being significantly impacted by AI, and can find out when an AI system is engaging with them."	"Transparency Review mechanisms will ensure citizens can question and challenge AI-based outcomes"
"Contestability: When an AI system significantly impacts a person, community, group or environment, there should be a timely process to allow people to challenge the use or outcomes of the AI system."	
"Accountability: People responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled."	"Accountability Decision-making remains the responsibility of organisations and individuals"

The NSW AI Assurance Framework (67), currently under review, has mandatory reporting requirements for state government uses of AI in NSW. A report released by the Centre for Automated Decision-Making and Society (ADM+S) and NSW Ombudsman notes that this framework has limited scope, and does not regulate, for example, local government uses of AI (75). However, it does apply to NSW health systems and services.

Regulation of Software as a Medical Device (SaMD)

The TGA's regulation of Software as a Medical Device (SaMD) was frequently referred to by clinical organisations in their policies and position statements. The TGA defines medical devices as those that diagnose, monitor, predict, prognose, treat, or alleviate disease, injury or disability; prevent disease, compensate for injury or disability; investigate, replace or modify the anatomy of a physiological or pathological process or state; control or support contraception; or are an accessory to a medical device (76). The TGA note that software is excluded where it does not meet those criteria (76, 77).

In their guidance specific to AI, the TGA note that devices that include Large Language Models (LLMs) will be subject to the same assessment and regulatory requirements as other medical devices, if they are deemed to be medical devices (78). The TGA also advise that clinical decision support (CDS) systems will be granted an exemption from regulatory requirements if they meet all three of the following requirements: (70) i/ the device does not directly process or analyse a medical image or signal; ii/ the device solely provides recommendations or support to health professionals; and iii/ the device does not replace the clinical judgement of any health professionals in making a clinical diagnosis or decision about patient treatment. Devices granted an exemption do not need to be approved by the TGA or registered in the Australian Register of Therapeutic Goods (ARTG), but manufacturers still need to notify the TGA that it is being supplied (76).

In their position statements on AI, both the Australian Medical Association (AMA) and the Australasian College of Dermatologists (ACD) emphasise that medical devices used in the clinic should be devices that are approved under the TGA's SaMD regulations (4, 79). The ACD specifically recommends that unregulated AI should not be used in clinical practice, including those that are excluded or exempt from TGA SaMD regulations or those that were approved before the TGA updated their risk assessment process in February 2021 (79). The Victorian Department of Health, in their guidance on clinical use of

generative AI, similarly bans clinical use of generative AI systems like ChatGPT for tasks like writing patient discharge notes, as they are unapproved and unregulated clinical uses of AI (80).

The Australian Alliance for Artificial Intelligence in Healthcare (AAAiH) note that not all clinical uses of AI, including clinical use of generative AI, falls within the TGAs SaMD remit, raising questions about how to approach their regulation (81). Both the AAAiH and AMA recognised that SaMD regulations were necessary but not sufficient to prevent harms related to implementing AI in healthcare services (4, 81).

Healthcare-specific governance approaches

The AAAiH roadmap recommended the implementation of a National AI in Healthcare Council with legislative responsibilities to ensure that AI is implemented in healthcare responsibly (81). The roadmap was developed through a broad national consultative process that included a national survey of 180 stakeholder organisations, and a policy development workshop with representatives of over 30 peak organisations in Australia. The AAAiH recommends that this Council would be led by Government and would contain cross-portfolio representation from agencies such as the TGA, NACCHOs, and the Australian Commission on Safety and Quality in Healthcare (see Box 1). The AAAiH recommend that one responsibility for the Council should be to oversee an accreditation process to ensure that hospitals and practices are prepared for cybersecurity threats and are storing and using patient data in accordance with best practice. The roadmap notes that this process could fall under the remit of the ACSQHC's accreditation scheme. The AAAiH recommends that this Council would also have responsibility for engaging with individual professional bodies to develop profession-specific codes of conduct.

Box 1: AAAiH Roadmap recommendations on safety, quality, ethics and security.

1. To better coordinate and harmonise the responsibilities and activities of those entities responsible for oversight of AI safety, effectiveness, and ethical and security risks, establish a National AI in Healthcare Council.
2. To ensure AI in healthcare is safe, effective and therefore does not harm patients, it needs to be developed and deployed within a robust risk-based safety framework.
3. For accreditation, healthcare organisations using AI should demonstrate that they meet minimum AI safety and quality practice standards.
4. Urgently communicate the need for caution in the clinical use of generative AI when it is currently untested or unregulated for clinical settings, including the preparation of clinical documentation.
5. Ensure the national AI ethical framework from the Department of Industry, Science and Resources supports the deployment of value- based clinical and consumer AI in routine practice.

The AMA recommends a similarly broad-reaching governance approach (4). The AMA Position Statement highlighted the importance of the government's role in regulating devices to ensure that they do not undermine the quality of care, and ensuring that investment in AI did not seek cost-effectiveness to the detriment of healthcare service quality (4). The AMA recommended that new regulations needed to be developed from a strong evidence base, to support patient outcomes, ensure decisions were made by humans, ensure that patients provide informed consent to procedures, and ensure that data is protected. In the statement, the AMA advised that existing legislation was not sufficient, and that new legislation should not impose additional burdens on the medical profession.

RANZCR developed broad principles (3), similar to those developed by the Australian and NSW State governments, to guide the ethical implementation of AI in healthcare. Those principles are:

- Safety – ensuring that AI systems are implemented with consideration of patient safety and, secondarily, workforce safety.

- Privacy and protection of data – ensuring that patient data is stored securely and in line with relevant laws.
- Minimisation of bias – ensuring that training data and outcome measures are transparently reported, including any populations that are under-represented in the training data.
- Transparency and interpretability – ensuring that results from AI systems can be understood and explained by healthcare professionals.
- Application of human values – ensuring that physicians apply humanitarian values when AI systems are used, and consider the personal preferences and values of their patients.
- Decision-making in diagnosis and treatment – ensuring that healthcare professionals continue to make decisions after a discussion with the patient, considering their presentation, history, treatment options, and preferences. AI can enhance decision-making capability, but not make final decisions.
- Teamwork – developing new multidisciplinary teams between health professionals, administrators, and developers to best leverage AI to deliver good care.
- Responsibility for decisions made – ensuring that, when using AI systems, responsibility for good outcomes is shared between the physician, the managers of the healthcare environment, and the developers.
- Governance – ensuring that any healthcare organisations using AI have transparent governance to ensure the practice is compliant with ethical principles, relevant standards and legal requirements.

They use these principles to structure their Standards of Practice for radiologists (82).

3.3.2 Engagement with consumers, patients and citizens

Table 12: Itemised recommendations on engagement with consumers, patients and citizens.

Organisation	Document	Type	Comments, recommendations and commitments
Department of Industry, Science and Resources	Australia's AI Ethics Principles (72)	Non-clinical	Under fairness principle – recommends engagement with consumers to ensure that AI systems are user-centric and accessible
	Safe and Responsible AI in Australia Consultation – Australian Government's Interim Response (2)	Non-clinical	Commits to working with citizens and putting people and communities at the centre of implementing and developing regulatory approaches to ensure AI is developed, designed and deployed to meet people's needs
NSW Government	Mandatory Ethical Principles for the use of AI (71)	Non-clinical	Recommends that "the community should be engaged on the objectives of AI projects and insights into data use and methodology should be made publicly available".
	Artificial Intelligence Assurance Framework (67)	Non-clinical	Assurance framework necessitates public consultation with communities who might be affected by the AI system.
Data61	Artificial Intelligence – Australia's Ethics Framework (70)	Non-clinical	Recommends investing in avenues for community feedback on AI to ensure development coincides with what Australians want
Australian Human Rights Commission	Human Rights and Technology: Final Report (69)	Non-clinical	Recommends that human rights impact assessments of AI systems should include public consultation
Australian Academy of Technological Sciences and Engineering	Submission to the Inquiry into artificial intelligence in New South Wales (25)	Non-clinical	Recommends efforts to ensure that the public know how to contest decisions that were made about them
Australian Medical Association	Artificial Intelligence in Healthcare – Position Statement (4)	Clinical	Recommends that all AI systems in healthcare should be co-designed, developed and tested with patients

Organisation	Document	Type	Comments, recommendations and commitments
Australian Alliance for Artificial Intelligence in Healthcare	A National Policy Roadmap for Artificial Intelligence in Healthcare (81)	Clinical	Highlights the importance of engaging with consumers, patients and citizens. Recommends developing digital literacy programs to help consumers engage in co-design projects for AI. Recommends engagement with NACCHOs and Aboriginal and Torres Strait Islander communities to ensure health data is collected and stored in a culturally safe and community-controlled way in line with principles of Indigenous Data Sovereignty. Highlights the role of clinicians in educating patients about the responsible use of AI, as part of a commitment to shared decision-making. Recommends professional codes of conduct and training emphasise the role of clinicians in educating patients.

Non-clinical

Citizen involvement in processes relating to the design and regulation of artificial intelligence were common in Australia's non-clinical and administrative guidance for AI development and implementation. At a federal level, Australia's AI Ethics Principles recommend engagement with consumers to ensure that AI systems are user-centric and accessible (72). Similarly, the Government's interim response to the consultation on Safe and Responsible AI highlights their commitment to working with publics and putting people and communities at the centre of regulatory approaches (2), with the aim of ensuring that AI is designed and implemented in ways that meet people's needs. Both documents recommend two-way channels of communication between citizens and decision-makers, where decision-makers are expected to take public feedback into account and use it to improve existing practices.

At the state level, only NSW documents mentioned citizen engagement. The NSW Mandatory Ethical Principles for the use of AI recommend that the community should be engaged in the 'objectives of AI projects' (71). The NSW AI Assurance Framework (67), in implementing the Mandatory Principles, necessitates consultation with the communities benefiting from, or impacted by, AI systems. If no community consultation is held, organisations must seek legal guidance or guidance from ethics committees about whether AI projects should proceed.

To implement public engagement at the state level, the NSW State Government launched 'Artificial Intelligence – Have your say' in December 2020 as part of their Artificial Intelligence Strategy (83) to seek community feedback on state government implementation of AI.

Clinical/health-specific

Community engagement was mentioned in two documents related to health-specific implementation of AI: the Australian Medical Association's (AMA's) position statement on AI in healthcare, and the Australian Alliance for Artificial Intelligence in Healthcare's (AAAiH's) roadmap. The AMA position statement recommends that all AI systems in healthcare should be co-designed, developed and tested with patients as a standard approach to applying AI in healthcare (4). The AAAiH Roadmap also highlights the importance of consumer engagement and recommends the development and implementation of digital literacy programs to enhance public capacity to participate in co-design projects (81).

Engagement with Aboriginal and Torres Strait Islander communities was only mentioned by one document. The AAAiH Policy Roadmap (81) recommends working with National Aboriginal Community

Controlled Health Organisations (NACCHOs) and affected Aboriginal or Torres Strait Islander communities to ensure health data used for AI projects is collected and stored in a culturally safe manner, noting the importance of Indigenous data sovereignty. Concrete steps and strategies for enacting patient engagement in AI projects in healthcare are still missing, and there is an opportunity to better standardise an approach to community engagement. An example of policy-orientated community engagement is illustrated by a citizens' jury (84) convened to deliberate on the question "Under which circumstances, if any, should artificial intelligence be used in Australian health systems to detect or diagnose disease?" (see Box 2).

Box 2: Australian national citizens' jury recommendations on healthcare AI (84).

A citizens' jury consisting of a diverse sample of Australians, recruited by random invitation, deliberated on the question: "Under which circumstances, if any, should artificial intelligence be used in Australian health systems to detect or diagnose disease?"
The jury made 15 recommendations:
1. We must have a charter for AI in the Australian health system and services. The charter must include (but not be limited to) the following:
- Underrepresented people, including Aboriginal and Torres Strait Islander people and [people from] minority populations
- Rural and remote [populations]
- Sustainability and environment
- Australian security and sovereignty
- Ethics and human rights
2. There must be an independent decision-making body to manage the charter. We recommend representation from across all stakeholder groups. We recommend the board chair is independent of the health system and investors to avoid bias.
3. Our recommendation in the application of AI in health care is that it must be continually evaluated to ensure the benefits to patients and health care professionals outweigh the harms
4. Our recommendation is that access to AI in health care must be the universal right of all Australians
5. There must be a guideline for patient rights. It is important to have guidelines that are inclusive of and non-discriminatory [with respect] to: individual values/beliefs, choice, accessibility, respecting underrepresented peoples, and being culturally appropriate.
6. We recommend that health care workers must be trained in AI systems that are to be implemented to their practice environment before clinical use.
7. We recommend that professional bodies must have clear directions regarding the use and intended outcomes of AI in the domains for which they are responsible.
8. We recommend that monitoring, auditing, and reporting be made mandatory to the appropriate governing body [or] bodies. Such processes should include but are not limited to unfavourable outcomes, performance, misuse and any benefits to the patients, clinicians, and health care systems.
9. Upon submission to the regulator, an AI system must provide information on its intended purpose and efficacy, its training dataset, flaws and limitations of use.
10. For AI systems to be approved in Australia, they must perform equal to or better than current standard health care practice.
11. It is important that AI training datasets must strive to be adequately representative and inclusive to capture Australia's multiculturalism and diversity.
12. Encourage and consider having AI software in health be free and open-source software to ensure transparency, public ownership, financial integrity, collaboration, security, privacy and trust.
13. We recommend that research used to underpin the use of AI in health care must be peer-assessed in an unbiased, independent, and robust manner. Australian data, with a sample representative of the population, should be used, but overseas data can be used when justified.

14. Research assessing the performance of AI screening tools should reflect real world clinical practice and follow standardised procedures in trial design. Data analysis and reporting should be transparent, and conclusions should reflect system performance.
15. We recommend that there is a comprehensive and fully funded community education program. This will ensure that the community is brought along with developments in and the application of AI in health. This should be located within a broader program of general digital health literacy that recognises particular community needs such as age, gender, ethnicity etc.

3.3.3 Equity, discrimination and human or patient rights

Table 13: Itemised recommendations on equity, discrimination and human or patient rights.

Organisation	Document	Type	Comments, recommendations and commitments
Data61	Artificial Intelligence – Australia's Ethics Framework (70)	Non-clinical	Recommends that AI should do no harm, and that the principle of 'non-instrumentalism' should be considered – whether the technology treats human beings as more than a 'cog in service of a goal'. Recommends that people should be informed when AI is being used in ways that impact them, and that there should be efficient processes to challenge algorithm outputs.
Department of Industry, Science and Resources	Australia's AI Ethics Principles (72)	Non-clinical	Principle of fairness highlights that AI systems should be inclusive, and that AI systems can perpetuate societal injustices if they are not designed and implemented with inclusiveness in mind. Recommends that AI should be developed to serve people's best interests, benefit all humans, align with human values, and serve humans. Principle of contestability highlights people's rights to contest decisions that affect them. Recommends that AI should not prevent citizens from being able to contest decisions and recommends that processes for contestability should be accessible for vulnerable groups.
NSW Government	Ethical Policy Statement (85)	Non-clinical	Commits to developing government uses of AI that are focussed on improving 'community outcomes', including lifting education standards, keeping children safe, and improving the healthcare system.
	Mandatory Ethical Principles for the use of AI (71)	Non-clinical	Recommends that AI should have clear community benefits and deliver the best outcome for the citizen. Recommends safeguards to manage data bias – projects should demonstrate focus on diversity and inclusion, datasets should be representative and appropriate for the problem to be solved.
	Artificial Intelligence Assurance Framework (67)	Non-clinical	Recommends that AI should not be used to make unilateral decisions that impact people or their human rights. Contains assessment items about the risk of discrimination from unintended bias, project operationalisations of fairness, preventing bias, and monitoring the system outputs to ensure outcomes are fair.
Australian Academy of Technological Sciences and Engineering	Submission to the inquiry into artificial intelligence in New South Wales (25)	Non-clinical	Recommends public outreach so that people know how to contest AI-assisted decisions. Recommends public investment into healthcare AI systems and public funding to cover costs of procedures for patients, to prevent AI systems becoming inaccessible to public patients.
Australian Human Rights Commission	Human Rights and Technology: Final Report (69)	Non-clinical	Recommends that citizens should have the right to be notified when a decision is made about them is materially influenced

Organisation	Document	Type	Comments, recommendations and commitments
			by AI. The notification should contain information about how to contest the decision.
Commonwealth Ombudsman	Automated Decision-making – Better Practice Guide (41)	Non-clinical	Highlights consumers' right to contest decisions that affect them, even when those decisions are made by AI
Australian Government Architecture	Artificial Intelligence Policy (Position) (86)	Non-clinical	Recommends that risk assessments should address whether the system is performing in a way that achieves equitable outcomes.
	Adoption of Artificial Intelligence in the Public Sector (87)	Non-clinical	Recommends that agencies should ensure their development and use of AI delivers public benefit. Recommends that people should know when AI is being used and should have the right to give feedback and contest decisions that affect them. Recommends consideration of anti-discrimination laws and how discrimination will be prevented in the use of AI. Advises that bias can disproportionately impact disadvantaged groups. Recommends considering whether a process is in place to ensure outcomes are fair when using generative AI systems.
Australasian College of Dermatologists	Position Statement: Use of Artificial Intelligence in Dermatology in Australia (79)	Clinical	Mentions that AI systems should enhance outcomes for Aboriginal and Torres Strait Islander patients
Australian Alliance for Artificial Intelligence in Healthcare	A National Policy Roadmap for Artificial Intelligence in Healthcare (81)	Clinical	Recommends engagement with NACCHOs and Aboriginal and Torres Strait Islander communities to ensure health data is collected and stored in a culturally safe and community-controlled way.
Australian Medical Association	Artificial Intelligence in Healthcare - Position Statement (4)	Clinical	Recommends that AI should never lead to greater inequities and that tools should ensure equity irrespective of race, age, gender, socioeconomic status and physical ability. Recommends that AI diagnoses should be accompanied by appropriate treatment – AI should not provide access to diagnosis without treatment
Royal Australian and New Zealand College of Radiologists	Ethical Principles for AI in Medicine (3)	Clinical	Recommends that data used to train AI systems should be representative of the intended population on which it is used.

Non-clinical

Equity, discrimination, and human or patient rights was one of the most commonly mentioned themes in the documents. Overall, the non-clinical documents focussed on ensuring that AI was used in ways that prevented harm and ensured public benefit. Focus was on ensuring that AI projects have clear community benefits (71, 72, 85, 87) and align with human rights and human values (67, 72). The NSW Government, in AI Assurance Framework, specifically recommends that AI should not be used to make unilateral decisions that impact people and their human rights (67), and Data61 encourages consideration of the principle of 'non-instrumentalism' – ensuring that technologies serve human values rather than treating people as 'cog(s) in service of a goal' (70).

Contestability was described by the Australian Government, Commonwealth Ombudsman, NSW Government and the Australian Human Rights Commission as a right held by citizens or consumers(41, 69, 72, 87). The Australian government defines contestability as such: "When an AI system significantly

impacts a person, community, group or environment, there should be a timely process to allow people to challenge the use or outcomes of the AI system." (72)

Australian organisations highlighted that citizens should be able to understand and contest any decisions made about them, including those made by (or assisted by) AI systems. Beyond just providing channels for people to contest decisions, policies and guidance from the DISR, AGA, Data61 and the Australian Human Rights Commission highlighted that citizens' right to contestability necessitated that organisations notify people when AI-informed decisions are made about them (69, 70, 87) and actively work to ensure that information about how to contest the decision is available, accessible, and efficient (25, 69, 70, 72). Australia's AI Ethics Principles specifically identify that avenues for contestability should be made accessible to vulnerable groups (72).

Equity and fairness were frequently mentioned as important considerations when developing and implementing AI projects. It was often noted that biased AI systems would have a disproportionately negative impact on disadvantaged groups(2, 70, 72, 86). Guidelines on implementation of AI in the public sector from Australian Government Architecture recommend ensuring that projects consider Australian anti-discrimination laws and avoid implementing systems that discriminate against certain groups (87). The Human Technology Institute report advised that Anti-Discrimination laws prevented organisations from using AI systems that directly or indirectly discriminate against people with protected attributes (64). Beyond consideration of existing laws, the Australian Government Architecture and the NSW Government recommended risk assessment frameworks or processes to ensure that outcomes from AI projects were equitable(67, 71, 86). The NSW AI Assurance Framework contains self-assessment items focussing on project operationalisations of fairness, preventing bias, and monitoring AI system outputs to ensure that outcomes are fair (67).

Clinical

The clinical documents typically addressed patient rights through the lens of patient advocacy and patient-centred care. The AMA position statement highlighted the importance of patient-controlled care, recommending that patients should retain the right to make their own informed decisions about their care and have control over their medical records and how their data is used and disclosed (4).

Most statements focussed on how hospitals, clinics, and individual physicians could ensure that patients were sufficiently informed about the use of AI in their care. The AHPRA Medical Radiation Practice Board guidance recommended clinician capacity-building to ensure that clinicians were able to inform patients about how AI would be used in their care (88). RANZCR guidance recommended that hospitals and clinics have information about their use of AI/ML that is available and accessible to patients (82).

Equity and fairness were addressed in the clinical documents with a focus on responsible design and implementation of AI systems to prevent unfair or biased outcomes. Both the AMA and RANZCR position statements recommend that the data used to train AI systems should be diverse, inclusive, and relevant to the intended population for which the AI tool will be used (3, 4). Both the Australasian College of Dermatologists (ACD) and the AAAiH roadmap specifically recommend that action should be taken to ensure that Aboriginal and Torres Strait Islander communities benefit from the introduction of AI (79, 81). The AMA Position Statement highlighted that the introduction of AI should never lead to greater health inequities (4).

A statement from the Australian Academy of Technological Sciences and Engineering (AATSE) advocated for public investment into AI innovations in healthcare to improve health equity. The statement argued that public hospitals should have access to new tools, and that public funding should

cover the full cost of using these tools, to ensure that any health benefits arising from the introduction of AI are accessible to public patients (25).

3.3.4 Privacy and confidentiality

Table 14: Itemised recommendations on privacy and confidentiality.

Organisation	Document	Type	Comments, recommendations and commitments
Data61	Artificial Intelligence – Australia's Ethics Framework (70)	Non-clinical	Recommends that AI systems should ensure private data is protected and kept confidential. Recommends that those developing and implementing AI should stay up to date on developments in AI technology and how these developments may create new privacy threats.
AATSE	Submission to the inquiry into artificial intelligence in New South Wales (25)	Non-clinical	Recognises that the use of AI creates privacy concerns, but does not provide any recommendations to address this
AGA	Artificial Intelligence Policy (Position) (86)	Non-clinical	Recommends that risk assessments of AI should consider people's right to privacy
	Adoption of Artificial Intelligence in the Public Sector (87)	Non-clinical	Recommends that public servants should consider privacy laws when using or implementing AI
	Interim Guidance on government use of public generative AI systems – November 2023 (74)	Non-clinical	Recommends that public servants do not input any private information into generative AI systems while completing their duties
Department of Industry, Science and Resources	Safe and Responsible AI in Australia Consultation – Australian Government's Interim Response (2)	Non-clinical	Acknowledges work on privacy law reforms to address some privacy issues created by the use of AI
	Australia's AI Ethics Principles (72)	Non-clinical	Recommends aiming to ensure respect for privacy and data protection, ensuring proper data governance and management
NSW Government	Artificial Intelligence Ethics Policy | Key Considerations (89)	Non-clinical	Recommends that agencies abide by existing privacy laws
	Ethical Policy Statement (85)	Non-clinical	States that AI will not be used by the government when it poses a risk to data or privacy
	Artificial Intelligence Assurance Framework (67)	Non-clinical	Risk assessment framework identifies levels of risk and how much control is required for each risk level, based on the sensitivity of the data being used by the AI system.
WA State Government	Interim Guidance for WA Public Sector Agencies on Adoption of Artificial Intelligence (90)	Non-clinical	Recommends that employees are aware of privacy laws when using generative AI. Recommends that only information classified as unofficial should be used in open generative AI technologies.
QLD State Government	Use of Generative AI for Government – Information Sheet (91)	Non-clinical	Recommends that employees should have appropriate permissions to use information as input into generative AI systems. Recommends that commercial AI systems should only be used with publicly available information – internal drafts should not be used as inputs into generative AI systems.
Commonwealth Ombudsman	Automated Decision-Making – Better Practice Guide (41)	Non-clinical	Addresses how existing privacy law should be used to implement automated decision-making systems that uphold privacy principles, including by having open and transparent processes, only collecting data where it is reasonably necessary, giving notice to individuals about how their information will be handled, only disclosing information for its authorised purpose, and ensuring that information is handled securely.
Human Technology Institute	The State of AI Governance in Australia (64)	Non-clinical	Addresses how companies implementing AI should ensure that privacy laws are upheld. Recommends that companies should have ongoing governance of AI systems, including

Organisation	Document	Type	Comments, recommendations and commitments
			monitoring of systems that are continuing to learn over time. Recommends that organisations are mindful of laws against using people's information for purposes other than what was disclosed upon collection of that information, and that organisations consider how this could prevent certain data being legally used for training or testing AI systems.
OVIC	Artificial Intelligence – Understanding Privacy Obligations (66)	Non-clinical	Addresses how organisations can develop AI that complies with existing privacy laws. Advises that data collection for AI development should be lawful and not unreasonably intrusive. Recommends that sensitive information should not be collected without direct consent from the individual but acknowledges that some exceptions are covered under IPP10.
AAAiH	A National Policy Roadmap for Artificial Intelligence in Healthcare (81)	Clinical	Recommends that health organisations should have minimum accreditation standards for data storage and management of patient data. Recommends development of mechanisms to enable consent-based industry access to healthcare data for the development of AI systems.
MTAA	Digital Health: Breaking Barriers to Deliver Better Patient Outcomes (92)	Clinical	Recommends that Australia develop a national framework for data governance to enable organisations developing digital health tools to access clinical data safely. Highlights the Canadian Institute for Health Information's approach to data governance as an exemplar approach.
RANZCR	Ethical Principles for AI in Medicine (3)	Clinical	Recommends that healthcare data should not be transferred from the clinical environment. Advises that safeguards for data sharing should be commensurate with the risk of re-identification.
	Standards of Practice for Clinical Radiology – Chapter 9: Artificial Intelligence (82)	Clinical	States that hospitals or Practices should pseudonymise data used to train or test AI. When sharing the data, the Practice should ensure it is as confidential as reasonably achievable. Information should be shared in secure channels and data should be disposed of at the conclusion of the data sharing agreement.
AMA	Artificial Intelligence in Healthcare – Position Statement (4)	Clinical	Recommends that AI should protect patient privacy, predominantly through patient control over their own health records. Patients should retain control over how their data is used and disclosed, and no disclosure should occur without patient consent.
AHPRA Medical Radiation Board	Artificial Intelligence: Guidance for Clinical Imaging and Therapeutic Radiography Professionals, a summary by the Society of Radiographers AI working group (88)	Clinical	Recommends that physicians should be aware of how the use of AI can have implications on patient privacy
ACD	Position Statement: Use of Artificial Intelligence in Dermatology in Australia (79)	Clinical	Recommends that physicians do not compromise patient privacy. Does not provide further guidance.

Non-clinical

Recommendations pertaining to privacy and confidentiality in the non-clinical documents typically referred to relevant existing privacy laws. Several of the organisations recognised that AI could create privacy issues, suggested that those in charge of implementing or using AI be aware of existing privacy laws, and recommended that any AI systems should be designed and implemented to ensure that privacy is protected(25, 41, 70, 87, 90). The Safe and Responsible AI interim response pointed to other work on privacy law reforms as addressing important privacy issues associated with AI (2).

A minority of the documents provided more specific recommendations for ensuring AI systems upheld principles of privacy. The report on AI Governance from the Human Technology Institute states that organisations should have ongoing governance of their use of personal information in AI systems, including where systems continue to learn over time (64). The Commonwealth Ombudsman recommends establishing a positive organisational culture where transparency is valued (41). The Human Technology Institute (64), Commonwealth Ombudsman (41), and OVIC (66) each advise that organisations should not collect personal information for the development of AI systems unless reasonably necessary and must ensure that this information is collected reasonably and fairly. Both the Human Technology Institute (64) and Commonwealth Ombudsman (41) recommend being mindful about laws against using information for purposes other than the intended purpose for collecting the information, which could prevent organisations from reusing datasets for the purposes of AI development.

A subset of documents which addressed public sector use of generative AI provided specific guidance for ensuring privacy was protected. Guidance for Western Australian (WA) public sector agencies recommended that government employees are aware of existing privacy laws (90), and guidance from WA, QLD, NSW and AGA advise against using private, sensitive, or official information as inputs for generative AI (86, 90, 91, 93).

Clinical

There was substantial variation in the approaches recommended by the clinical organisations. The Medical Technology Association of Australia (MTAA) recommended a national governance framework to ensure privacy is protected, whilst providing avenues for data access to industry (92). The AAAiH policy (81) roadmap had a risk-based approach, recommending a minimum accreditation standards for patient data storage, and mechanisms to allow for consent-based industry access to data. Both recommendations were made with the intention of balancing the protection of sensitive data with the potential benefits of utilising the data for developing innovative technologies.

The RANZCR and AMA statements made more specific recommendations related to the storage and transfer of patient data in hospital and clinics. RANZCR's position statement recommended that patient data not be transferred from the clinical environment without patient consent or approval from an ethics committee, unless required by law (3). AMA's position statement continued its focus on patient control, recommending that patients should retain full jurisdiction over how their data is used and disclosed (4).

The AHPRA and ACD position statements also mentioned privacy, but only insofar as recommending that physicians be aware of the potential impacts of AI on the privacy and security requirements of patient data (88), and ensure that physicians do not compromise patient privacy (79).

Policy implications of privacy law in Australia

In Australia, privacy legislation exists at the federal, state and territory level. In general terms the federal Privacy Act 1988 (Cth) applies to federal government agencies and the private sector. State and territory privacy legislation generally applies to state and territory public sector agencies. However, in three jurisdictions — NSW, Victoria, and the ACT — there is also separate health privacy legislation that applies across the private and public sectors. This creates a complex web of regulation for those handling health information in these jurisdictions, particularly in the private sector, including those handling health information in the development, deployment, and use of AI systems in the acute care environment. This kind of regulatory complexity creates a level of overlap and uncertainty for industry and consumers in relation to the collection, use and disclosure of personal health information. While the acute care sector is already dealing with this regulatory complexity in those jurisdictions with overlapping

health privacy regulation, such regulatory complexity can lead to a cautious approach to privacy issues, including in relation to the deployment and use of new technology by agencies and organisations.

3.3.5 Evaluation, monitoring and maintenance as an issue for governance

Table 15: Itemised recommendations on evaluation, monitoring and maintenance as an issue for governance.

Organisation	Document	Type	Comments, recommendations and commitments
Data61	Artificial Intelligence – Australia's Ethics Framework (70)	Non-clinical	Recommends that AI should generate net benefits, and should be regularly monitored to ensure the system still adheres to ethical principles
AATSE	Submission to the inquiry into artificial intelligence in New South Wales (25)	Non-clinical	Recommends regular reporting for AI systems used in administrative decision-making, to ensure that publicly available research can audit the systems and ensure that they are working correctly.
AGA	Artificial Intelligence Policy (Position) (86)	Non-clinical	Recommends that risk assessments are reviewed and updated as technologies are further developed and matured with monitoring and reporting systems. Recommends prioritising benefits to society, human rights, and impartial treatment in government use of AI.
	Adoption of Artificial Intelligence in the Public Sector (87)	Non-clinical	Recommends that assurances be put in place so that humans can maintain visibility of unintended consequences, including during ongoing maintenance. Recommends ongoing monitoring to ensure that AI performance is not degrading over time.
Department of Industry, Science and Resources	Safe and Responsible AI in Australia Consultation – Australian Government's Interim Response (2)	Non-clinical	Identifies that many responses to the original consultation supported initiatives like internal and external testing of AI systems before the release of a system, as well as mechanisms for ongoing monitoring.
	Australia's AI Ethics Principles (72)	Non-clinical	Recommends that any inferences drawn from AI should be monitored in an ongoing manner, and systems should be tested regularly to ensure that they still meet the purpose for which they were implemented.
NSW Government	Artificial Intelligence Assurance Framework (67)	Non-clinical	The AI Assurance Framework is a self-assessment framework used in NSW for ongoing monitoring of AI projects.
Commonwealth Ombudsman	Automated Decision-Making – Better Practice Guide (41)	Non-clinical	Recommends that systems should have comprehensive audit trails so that they can be reviewed. Citizen complaints data should be part of the ongoing review of AI systems.
Human Rights Commission	Human Rights and Technology: Final Report (69)	Non-clinical	Recommends an independent audit of all government use of AI to ensure that human rights are upheld.
AAAiH	A National Policy Roadmap for Artificial Intelligence in Healthcare (81)	Clinical	Recommends AI be developed and deployed within a robust risk-based safety framework: • Pre-market, vendors must provide regulators with rigorous evidence that their algorithms perform well in real-world settings. • Post-market, improve the effectiveness of national post-market safety monitoring so that cases of AI-related patient risk and harm are rapidly detected, reported and communicated. Recommends, for accreditation, healthcare organisations using AI should demonstrate that they meet minimum AI safety and quality practice standards.
RANZCR	Ethical Principles for AI in Medicine (3)	Clinical	Recommends that, where algorithms are developed offshore, they should be tested and evaluated on local data before being implemented in Australia. RANZCR

Organisation	Document	Type	Comments, recommendations and commitments
			emphasise that this is especially important for Indigenous populations.
	Standards of Practice for Clinical Radiology – Chapter 9: Artificial Intelligence (82)	Clinical	Recommends that algorithms should be tested for their performance before being implemented, and that this performance should be a consideration when the Practice is considering whether to implement the tool. ML or AI used in a Practice must have been tested, reached saturation of learning, and met requirements for regulatory approval. "the ML or AI can continue to learn but this must be done in a parallel version of the tool, which cannot be used in patient care until its performance has been tested" The practice must ensure that the developer used 'appropriately independent' data sets for training, validation and testing phases. The AI tool must not have capacity for ongoing learning when implemented at the Practice - all changes need regulatory approval. Recommends an initial performance audit within a set period of time after deploying. Recommends ongoing audits, performed by independent parties. Performance should be reported annually against standards and considering any patient feedback. Audits should also assess whether the radiology team understands the use of AI and ML systems and tools. Adverse events need to be reported to the relevant regulatory agency (TGA in Australia) and to the manufacturer (9.5.3)
AMA	Artificial Intelligence in Healthcare – Position Statement (4)	Clinical	Recommends that AI should be subject to regular review and audit for quality assurance, safety, and clinical enhancement. Audits should be transparent and accountable. Recommends auditing and updating of algorithms to ensure they are based on most current data available - to improve equity & diversity Recommends that AI should ensure that it is utilised only where it improved health outcomes for patients (3.8) as supported by best practice evidence.
ACD	Position Statement: Use of Artificial Intelligence in Dermatology in Australia (79)	Clinical	Recommends real-world evaluations to show evidence of performance. To be implemented, performance should be equivalent to, or better than, physicians

Non-clinical

Most of the non-clinical documents mentioned that AI systems should undergo evaluation, monitoring and maintenance. For initial evaluations, the organisations recommended prioritising evaluations of tools' benefits to society, adherence to human rights, and adherence to ethical principles (61, 54, 32). For ongoing monitoring and maintenance, the organisations highlighted the importance of maintaining human 'visibility' of any errors or unintended consequences coming from AI systems (87). The organisations recommended ongoing auditing to ensure that AI systems were still fit for the purpose for which they were implemented, were delivering community benefits, and were not degrading in performance over time (2, 70-72). The AATSE submission explicitly recommended that performance data for administrative AI systems should be reported publicly, so that publicly available research could independently audit the systems and ensure they are working correctly (25), and the Human Rights

Commission's report recommended an independent audit of all Government use of AI to ensure human rights were being upheld (69).

NSW's Artificial Intelligence Assurance Framework (67) is an exemplar of an evaluative self-assessment tool for ongoing maintenance of AI systems, to ensure they are delivering benefits and adhering to ethical standards.

Clinical

The clinical documents were foremost focused on ensuring that the implementation of AI did not compromise patient safety or experience. The ACD recommended that AI should be evaluated to ensure that performance is at least equivalent to physicians (79), and AMA and RANZCR both recommended that tools should not be used if they do not have the independent real-world evidence to demonstrate their safety and equivalent or superior performance to human physicians (3, 4, 82). The organisations made recommendations to ensure that the data on which the algorithms were trained were appropriate for the tasks for which they were being implemented. The AMA position statement recommended that algorithms should receive post-market updates to ensure they are based on the most current and relevant data to improve performance (4). RANZCR recommended that, where algorithms are developed offshore, they should be tested and evaluated on local data before being implemented, with a particular focus on whether they meet the needs of Indigenous populations (3). To facilitate these evaluation procedures, the AAAiH roadmap recommends the creation of a national council to implement a risk-based safety framework and enact minimum standards of practice for healthcare organisations in relation to use of AI (81), and the AMA made a similar recommendation that practices develop their own risk management frameworks to ensure patient safety. RANZCR (82) recommend that any adverse events occurring in Australia associated with AI systems should be reported to the TGA.

Two clinical documents made recommendations related to the potential for AI systems to continue to be trained on new data after implementation. The AAAiH roadmap (81) recommended that evaluation processes include post-market monitoring to account for AI products that are continuously trained and change over time. In contrast, RANZCR (82) recommend that any AI systems implemented in clinical care must be locked: tools can continue to learn in a 'parallel version' which should undergo regulatory approval again before it is used.

3.3.6 Transparency

Table 16: Itemised recommendations on transparency.

Organisation/ agency	Document	Type	Comments, recommendations and commitments
Data61	Artificial Intelligence – Australia's Ethics Framework (70)	Non-clinical	Acknowledged that technical explanations produced by AI systems were often not useful for end users or were not possible to provide for black box systems. Recommends that AI systems be transparent enough so that they can be evaluated and audited.
AGA	Artificial Intelligence Policy (Position) (86)	Non-clinical	Recommends allowing for human review of automated decisions, ensuring that they are transparent and explainable to a degree that allows for auditing.
	Adoption of Artificial Intelligence in the Public Sector (87)	Non-clinical	Recommends that agencies should consider implementing systems with clear rationales for decision-making processes.
	Interim Guidance on government use of public generative AI systems – November 2023 (74)	Non-clinical	Recommends disclosure of when generative AI has been used to inform activities or generate information.

Organisation/ agency	Document	Type	Comments, recommendations and commitments
Department of Industry, Science and Resources	Safe and Responsible AI in Australia Consultation – Australian Government's Interim Response (2)	Non-clinical	Recognises that opaque systems can make it difficult for users to identify errors and assure quality.
NSW Government	Mandatory Ethical Principles for the use of AI (71)	Non-clinical	States that AI systems should have a review mechanism to answer citizens' questions about the use of data or the AI informed outcomes. Review mechanism should also allow citizens to challenge outcomes from AI systems.
	Generative AI: basic guidance (93)	Non-clinical	Recommends that workers acknowledge and attribute where any content has been developed with the assistance of generative AI
QLD State Government	Use of Generative AI for Government – Information Sheet (91)	Non-clinical	Recommends that public service workers identify when content has been created using generative AI.
Commonwealth Ombudsman	Automated Decision-Making – Better Practice Guide (41)	Non-clinical	Recommends that systems should be transparent enough to allow for appropriate auditing.
OVIC	Artificial Intelligence – Understanding Privacy Obligations (66)	Non-clinical	Mentions the importance of having transparent administrative automated decision-making systems to allow for citizens to contest decisions made about them. Mentions that full explainability is not always possible or desirable, but that meaningful and accessible explanations are important. Recommends that organisations and agencies are transparent about their use of AI.
Human Rights Commission	Human Rights and Technology: Final Report (69)	Non-clinical	Recommends that decisions made by AI should be auditable so that consumers can exercise their right to an explanation. People have rights to technical and understandable reasons for a decision. Reasons generated by AI systems should be understandable for someone with relevant expertise. Recommends that a relevant body should provide guidance to organisations and agencies on how to generate reasons that meet these criteria.
ADM+S	Automated Decision-Making in NSW (75)	Non-clinical	Recommends the implementation of a public register for administrative decision-making technologies. Recommends a graded approach to transparency for administrative automated decision-making systems. "For example, in the ADM context, a register might require more or less disclosure depending on whether the system is used for data capture, predictive analysis, decision support, decision-making, or enforcement." Recommends transparency as an approach for knowledge-sharing between government departments.
AAAiH	A National Policy Roadmap for Artificial Intelligence in Healthcare (81)	Clinical	Recommends that developers should be transparent about the populations used for training and testing AI systems.
RANZCR	Ethical Principles for AI in Medicine (3)	Clinical	Recommends that healthcare practitioners should be able to interpret the clinical appropriateness of a result reached and weigh up the potential for bias. Systems should be transparent enough to allow for this. Recommends that the location where data was collected and tested should be stated for transparency reasons.
	Standards of Practice for Clinical Radiology – Chapter 9: Artificial Intelligence (82)	Clinical	Recommends that developers should clearly and transparently state the training and testing datasets, as well as information about past performance of the AI tool, so that hospitals or Practices can effectively evaluate the tool and decide whether it is appropriate to be implemented.

Organisation/ agency	Document	Type	Comments, recommendations and commitments
AMA	Artificial Intelligence in Healthcare – Position Statement (4)	Clinical	Recommends that AI should have transparent evaluation processes that clearly indicate how well the tool performs.

Non-clinical

Transparency and explainability were very frequently mentioned in the non-clinical and administrative documents. The Safe and Responsible AI Interim response recognised that opaque or unexplainable AI systems can make it difficult to identify errors and ensure quality (2). The Data61 report acknowledged that technical explanations produced by AI systems were often not useful for users, or were not possible to produce in instances where black box algorithms were being used (70). The AGA, Human Rights Commission, Data61, the Commonwealth Ombudsman, OVIC, and the NSW Government each recommended that AI systems should be transparent enough to allow for citizens to contest decisions made about them and for systems to be effectively and independently evaluated (41, 66, 69, 70, 86). The Human Rights Commission report identified that systems should be able to produce 'reasons' for a decision, and that those reasons should be understandable to a person with relevant expertise (69). The policies provided no precise examples of the information that should be transparent, or the reasons that should be generated by AI systems, although it was acknowledged that this information would be different depending on the application and its risk level (70).

ADM+S recommended a risk-based approach to transparency requirements for automated decision-making systems, with higher-risk uses of systems subject to more disclosure requirements (75). The report highlights that better transparency about governance of AI systems could be beneficial for knowledge sharing between government departments and recommends a national register for recording information about uses of automated decision-making systems.

Several policy documents recommended transparency about when AI was being used. The AGA, QLD Government, and NSW Government, in their policies about government use of generative AI, each recommended that agencies should be transparent about when generative AI has been used to generate content (86, 91, 93). OVIC recommends disclosure about when AI is being used in general (66), and the NSW Assurance Framework contains items to assess whether agencies have informed citizens of instances when they are interacting with AI (67).

Clinical

The AMA and RANZCR policies as well as the AAAiH roadmap recommended that there should be clear and transparent reporting of the datasets used to train and test AI systems, any populations that may be over- or underrepresented in those datasets, the performance of AI systems, and the evaluation or assessment processes undertaken to determine the performance (4, 81, 82). The intention behind these recommendations was to ensure that practices are aware, when deciding whether to implement AI systems, of potential issues with transferability of systems to new healthcare settings, and of the potential for algorithmic bias when populations are underrepresented in training data. RANZCR explicitly recommend that systems should be transparent enough that healthcare practitioners should be able to interpret the clinical appropriateness of a result reached and weigh up the potential for bias (3).

3.3.7 Accountability and liability

Table 17: Itemised recommendations on accountability.

Organisation/ agency	Document	Type	Comments, recommendations and commitments
Data61	Artificial Intelligence – Australia's Ethics Framework (70)	Non-clinical	Recommends that in high-risk situations, decisions should be made by humans, but that automated decision-making systems could be appropriate to assist human decision-makers.
AGA	Adoption of Artificial Intelligence in the Public Sector (87)	Non-clinical	Recommends that agencies should develop risk management frameworks to ensure that there are clear human intervention points in automated decision-making systems, to ensure decisions have human input.
	Interim Guidance on government use of public generative AI systems – November 2023 (86)	Non-clinical	Advises that public service workers are responsible for decisions, and that generative AI should not be used to make decisions.
Department of Industry, Science and Resources	Safe and Responsible AI in Australia Consultation – Australian Government's Interim Response (2)	Non-clinical	Acknowledges that responses to the Safe and Responsible AI paper supported human-in-the-loop processes for high-risk uses of AI.
NSW Government	Artificial Intelligence Assurance Framework (67)	Non-clinical	States that AI projects should have four separate roles held by four separate individuals: one responsible for the outcomes of the project, one responsible for the use of AI insights, one responsible for the technical performance of the AI system, and one responsible for data governance. The people in these roles should be identified in the self-assessment process, and should be appropriately trained for their roles.
	Mandatory Ethical Principles for the use of AI (71)	Non-clinical	Recommends that decision-making should remain the responsibility of organisations, agencies, and individuals, even when decisions are informed by AI. Recommends that AI functions should be overseen by people with relevant expertise, with assurances put in place for the outcomes of AI projects.
WA State Government	Interim Guidance for WA Public Sector Agencies on Adoption of Artificial Intelligence (90)	Non-clinical	Recommends that existing decision-makers remain responsible for machine-assisted decisions
QLD State Government	Use of Generative AI for Government – Information Sheet (91)	Non-clinical	Advises that public service workers are responsible for the generative AI content that they use.
Human Rights Commission	Human Rights and Technology: Final Report (69)	Non-clinical	Recommends retaining human responsibility for decisions made by AI.
ADM+S	Automated decision-making in NSW (75)	Non-clinical	Recommends identification of a responsible officer for government projects involving AI, similar to US proposals to create 'Chief AI Officers' to oversee government use of AI.
RANZCR	Ethical Principles for AI in Medicine (3)	Clinical	Recommends that responsibility for the delivery of healthcare should remain with the healthcare professional, but responsibility for the ethical use of an AI system and its outcomes should rest with the healthcare professional in combination with hospital management and the manufacturer of the AI tool. Recommends hospitals or Practices should have accountable governance to oversee implementation and monitoring. Recommends that autonomous AI systems (operating without direct oversight) "must be used under very carefully tested and monitored circumstances. AI systems must not be used without human oversight where results could impact the patient, and the use of these tools must be carefully

Organisation/ agency	Document	Type	Comments, recommendations and commitments
			considered in the light of the clinical context and potential patient risk"
	Standards of Practice for Clinical Radiology – Chapter 9: Artificial Intelligence (82)	Clinical	Advises that AI can be used to inform value judgements, but never to make value judgements on its own. Recommends that responsibility to decide whether or not to use an AI/ML tool should rest with the physician. Recommends the employment of a CRIO (Chief Radiologist Information Officer) trained in appropriate skills to engage with other staff affected by the deployment of AI/ML. The CRIO should oversee software upgrades and any contingency planning.
AMA	Artificial Intelligence in Healthcare – Position Statement (4)	Clinical	Recommends that AI should never replace physicians' judgements. If physicians and AI disagree, there should never be an expectation for the AI decisions to be used. Recommends that the decision to use or not to use AI should rest with the physician and not the hospital or Practice.
ACD	Position Statement: Use of Artificial Intelligence in Dermatology in Australia (79)	Clinical	Recommends that AI only be used in augmentative roles, never replacing physician judgement.

Non-clinical

Policies typically recommended that humans retain accountability for decisions, even when they are informed by AI. The Human Rights Commission and the NSW Mandatory Principles both recommend that decision-making should remain the responsibility of organisations, agencies, or individuals, and that accountability for AI-informed decisions should not be any different to decisions made without the influence of AI systems (69, 71). The Data61 report agreed that in lower risk situations, AI-assisted decisions were appropriate, but in high-risk situations, decisions should be made by humans (70). A policy from Australian Government Architecture stated that workers should not use generative AI to make any decisions (74). Policies from Western Australia, Queensland and NSW state governments on use of generative AI stated that workers are responsible for any AI-generated content that they produce, share or use while performing their duties (90, 91). Overall, the policies strongly advocate for responsibility for decisions to be retained by individuals and organisations or agencies, even in circumstances where AI is being used to assist those decisions.

Some policies contained recommendations for how to ensure that organisations or agencies retain accountability for decisions when AI is being used. The Australian Government's interim response to the Safe and Responsible AI paper supports the development of human-in-the-loop processes in high-risk situations (2). The AGA recommend that Government agencies develop risk management frameworks which identify human intervention points in decision-making to ensure that decisions are not being made wholly by AI systems (87).

The Australian Government's AI Ethics Principles recommend that the person or people responsible for keeping an AI system safe should be identified (72). The NSW policies similarly recommend that accountable roles for managing the safety of AI systems should be identified (67, 71). The report released by ADM+S made a similar recommendation for the identification of a responsible officer (75). The NSW Assurance framework (67) necessitates identification of four separate roles held by four separate individuals: one role responsible for the use of AI insights, one responsible for the outcomes of the AI project, one responsible for the technical performance of the AI system, and one for data governance. The policy recommends that those in these roles should be appropriately trained and should be identified during the risk assessment process.

Clinical

The clinical documents recommended that physicians should retain authority and control over decisions and diagnoses, even when they are assisted by AI. The ACD recommended that AI only be used in augmentative roles, never replacing physician judgement (79). Similarly, the AMA recommended that AI should never replace physician judgement, and that in instances where physicians disagree with decisions made by AI, there should not be an expectation for the AI decision to be followed (4). RANZCR allowed for the use of AI to 'inform' value judgements, but never to solely make those value judgements (82). RANZCR make allowances for autonomous AI systems (those making decisions without being directly overseen by a professional) under 'very carefully tested and monitored circumstances' in situations where 'results could [not] impact the patient'(3).

Both RANZCR and the AMA recommended that the decision to use or not use an AI tool should rest with the physician, and not with the hospital or practice (4, 82).

RANZCR, in their Standards of Practice for Radiologists, recommend that the implementation and use of AI/ML should be overseen by a Chief Radiologist Information Officer (CRIO), who oversees software upgrades and contingency planning, and who supports other clinical staff who are interacting with AI systems (82). RANZCR advocate for shared responsibility for AI-informed decisions, with responsibility for healthcare outcomes resting with the clinician, but with accountability for the ethical use of the AI tool resting with the clinician, the hospital or practice, and the developer of the AI tool (3, 82).

3.3.8 Consent

Table 18: Itemised recommendations on consent.

Organisation/ agency	Document	Type	Comments, recommendations and commitments
Data61	Artificial Intelligence – Australia's Ethics Framework (70)	Non-clinical	Identifies that the Privacy Act has requirements for consent for use of data. Mentions that Australian law does not have a 'right to be forgotten' as the EU does, despite the requirement in the Privacy Act for consent to be 'current and specific'.
NSW Government	Mandatory Ethical Principles for the use of AI (71)	Non-clinical	Recommends that AI projects should 'clearly demonstrate … agreement for the consent for data use, with sufficient information provided n how the data will be used to ensure informed consent'.
OVIC	Artificial Intelligence – Understanding Privacy Obligations (66)	Non-clinical	Recommends that information should be collected from individuals directly where possible to ensure that the individual knows it is being collected and consents to its collection. States that consent is usually necessary for the collection and use of sensitive data, unless covered by an exemption under IPP10. Recommends seeking consent where possible, as it is beneficial for promoting public trust. Acknowledges that seeking consent is not always possible for AI projects because it creates complications. For example, allowing individuals to withdraw their consent will not always be possible once data has been used to train a model.
AAAiH	A National Policy Roadmap for Artificial Intelligence in Healthcare (81)	Clinical	Recommends the development of a process for consent-based industry access to healthcare data for the development of AI systems. Does not provide further guidance on whose consent would be needed or under what circumstances.

Organisation/ agency	Document	Type	Comments, recommendations and commitments
RANZCR	Ethical Principles for AI in Medicine (3)	Clinical	Recommends consideration of the rights of the patient, including the right to revoke consent. Recommends that consideration be given to Indigenous data sovereignty.
	Standards of Practice for Clinical Radiology – Chapter 9: Artificial Intelligence (82)	Clinical	States that the Practice or hospital should seek the consent of the patient or approval of an appropriate ethics board to waive consent procedures before sharing data for AI projects.
AMA	Artificial Intelligence in Healthcare – Position Statement (4)	Clinical	Mentions that patients should have a right to consent to any procedure, and that patients have a right to know where and how AI is being used. Recommends that patients should need to consent for their data to be disclosed, even if it is de-identified.

Non-clinical

The Data61 report identified that the Australian Privacy Act has certain requirements for consent for use of data (70). These include that the individual is adequately informed before giving consent, that the individual gives consent voluntarily, that consent is current and specific, and that the individual has the capacity to understand and communicate their consent. The report states that these laws should be considered when developing and implementing AI but does not go into further detail about how this should be approached. The NSW Mandatory Principles recommend that AI projects 'clearly demonstrate … consent for data use' (71). The NSW principles do not provide guidance related to any consent exemptions that may be relevant under the current Privacy Act.

OVIC's AI privacy guidelines provide more detail about how consent could be actioned in AI projects but does not provide any proscriptions beyond what is already legislated in the Privacy Act (66). The guidelines recommend that information used to train or test AI should be collected directly from individuals wherever possible for transparency, particularly when data is sensitive, but that the Privacy Act allows exemptions under certain circumstances so that explicit consent does not always need to be sought. The OVIC guidelines advocate for consent as beneficial for increasing public trust in projects, but acknowledges that explicit consent is not always feasible, especially when an individual revoking their consent would mean that data would need to be removed from AI training datasets.

Clinical

The AAAiH roadmap recommends development of processes for consent-based industry access to healthcare data for the purposes of AI development but does not provide further detail on how consent should be managed or whether it would be necessary under all circumstances (81). RANZCR and the AMA, in contrast, both state that practices should seek explicit consent from patients for use or sharing of their data for the development of AI systems (4, 82). AMA affirm that even when data is de-identified, patient consent should still be sought (4). RANZCR state that patients' right to revoke consent should be observed, and that consideration should be given to Indigenous data sovereignty and governance (3).

3.3.9 Worker training and support

Table 19: Itemised recommendations on worker training and support.

Organisation / agency	Document	Type	Comments, recommendations and commitments
AATSE	Submission to the inquiry into artificial intelligence in New South Wales (25)	Non-clinical	Recommend mandatory training for public servants on the ethical use of AI, including training for educators and physicians.

Organisation / agency	Document	Type	Comments, recommendations and commitments
AGA	Adoption of Artificial Intelligence in the Public Sector (87)	Non-clinical	Recommends that technical and policy staff should have appropriate training to create and maintain the technology application and understand the impact of it on the overall process it supports. Recommends that senior executives responsible for automated decisions should have the necessary skills to consider ethics in the decision-making process.
	Interim Guidance on government use of public generative AI systems – November 2023 (74)	Non-clinical	Recommends training for APS staff before using generative AI.
NSW Government	Generative AI: basic guidance (94)	Non-clinical	Recommends that workers are aware of concepts like prompt sensitivity and hallucination before using generative AI.
	Artificial Intelligence Strategy (83)	Non-clinical	Reports on government efforts to incorporate training about AI into the 'skills framework for the information age' so that it is added to ICT training curriculums.
QLD State Government	Use of Generative AI for Government – Information Sheet (91)	Non-clinical	Recommends that workers familiarise themselves with the data used to train generative AI, so that they can make an informed decision about whether it is appropriate to use.
Commonwealth Ombudsman	Automated Decision-Making – Better Practice Guide (41)	Non-clinical	Recommends training for all staff for administering automated decision-making systems in administrative roles.
Human Technology Institute	The State of AI Governance in Australia (64)	Non-clinical	Recommend that organisations invest in strategic expertise in AI across the organisation, including why AI systems are being used, how they operate, and how they can be used safely.
RANZCR	Ethical Principles for AI in Medicine (3)	Clinical	Recommends that hospitals or Practices developing AI should ensure that AI is used and developed by appropriately trained people. Recommends the development of multidisciplinary teams between clinicians, developers and administrative staff for overseeing the implementation and development of AI.
	Standards of Practice for Clinical Radiology – Chapter 9: Artificial Intelligence (82)	Clinical	States that all clinicians using AI systems need to be trained in its use before incorporating into their workflow –Practice or hospital should provide appropriate training for users of AI systems. Radiology team should be trained to understand risks and shortcomings of AI, and how to understand the output and how it relates to a particular patient.
AMA	Artificial Intelligence in Healthcare – Position Statement (4)	Clinical	Recommends that implementation of AI should be done in consultation with the medical profession and community. Changes to education and training should be made to account for the potential for unforeseen consequences.
AHPRA Medical Radiation Board	Artificial Intelligence: Guidance for Clinical Imaging and Therapeutic Radiography Professionals, a summary by the Society of Radiographers AI working group (88)	Clinical	Strong focus on capacity-building for health workers throughout the document. Recommends that health workers should be aware of how to explain he role of AI to patients and provide care that meets patients' needs. Recommends that curriculums are updated so that health workers are exposed to AI applications during their training.
ACD	Position Statement: Use of Artificial Intelligence in Dermatology in Australia (79)	Clinical	Recommends that dermatologists develop knowledge and skills for how to use and monitor AI whilst upholding other ethical principles for AI use.
Victorian Department of Health	Health Service Use of Unregulated Artificial Intelligence (AI) - Health Service Advisory (80)	Clinical	Recommends providing advisory material to staff on the risks of using AI for clinical purposes.

Organisation / agency	Document	Type	Comments, recommendations and commitments
RACMA	Position Statements: Digital Health (95)	Clinical	Advocates for "The role of qualified medical leaders in development, leadership and governance of digital health systems … [and the] judicious use of Artificial Intelligence as an adjunct to clinical service delivery under the supervision of qualified health professionals."
TGA	Medical device cyber security information for users (96)	Clinical	Recommend that physicians should be upskilled in safe and secure use of medical devices.

Non-clinical

Many of the policies recommended that staff using AI systems should receive training in how to administer the tools safely and effectively. The Commonwealth Ombudsman recommends that training should be provided for staff administering automated decision-making systems in administrative roles (41). They recommend that organisations and agencies should carefully assess staff training requirements when implementing automated decision-making systems and implement processes so that all relevant staff understand how to "adequately explain a decision made by an automated system or identify an appropriate escalation path for a customer seeking information" (p.23). AATSE similarly recommend mandatory training for public servants on the ethical use of AI, which they extend to roles such as educators and physicians (25). AGA recommends that technical and policy staff should have the appropriate training to create and maintain the AI system and understand the impact of the system on the process it supports (87). They further recommend that senior executives accountable for any AI-informed decisions should have "the necessary skills to consider ethics in the decision-making process" (87). The Human Technology Institute recommend that corporate leaders invest in strategic expertise in relation to AI across the organisation to develop an understanding of why AI systems are being used, how they operate, and how they can be used safely (64).

The NSW AI Strategy mentions that the state government have incorporated training about AI into the 'Skills Framework for the Information Age' so that ICT professionals receive relevant training about AI (83).

Four of the generative AI policies contain recommendations about worker training and support. The AGA, NSW and Queensland governments recommend training for public servants before using any generative AI applications, to ensure that staff are aware of the data used to train generative AI systems and understand concepts like prompt sensitivity and the capacity for generative AI applications to hallucinate (86, 91, 93). The Victorian Department of Health provide healthcare worker-specific guidance for training in generative AI, recommending that advisory material be provided to all staff on the risks of using generative AI applications for clinical purposes (80).

Clinical

The clinical policies contain recommendations on clinician capacity-building and training to ensure that clinical AI systems are implemented safely. The AMA (4) identify that the use of AI in healthcare delivery will "create unforeseen consequences for the safety and quality of care for the patient as well as for the healthcare workforce and medical profession" (section 2.21). They therefore recommend changes to education, training, supervision, and examination to address the potential for unforeseen consequences. The AHPRA Medical Radiation Board's position statement is predominantly focussed on clinician capacity-building, recommending updates to curriculum so that practitioners are exposed to AI applications during their training, and ensuring that clinicians are informed enough about how AI is being used to effectively inform their patients and deliver patient-centred care (88).

The RACMA position statement on digital health acknowledges that AI systems could improve healthcare service delivery if implemented judiciously (95). The statement commits to ensuring that digital health is reflected in the medical teaching and training institutions' curriculum and advocating for judicious use of AI supervised by appropriately qualified health professionals. The ACD recommends dermatologists develop knowledge and skills to uphold the other ethical principles outlined in the position statement when implementing AI in their practice (79). The TGA state that clinicians should be upskilled in safe and secure use of medical devices, including those that use AI, so that they can safely use the tools and communicate risks to patients (96).

RANZCR's Standards of Practice for radiologists necessitates that all clinicians using AI systems should be trained in their use before the tool is incorporated into their workflow (82). They state that the practice or hospital should provide training for clinicians so that the radiology workforce understands the risks and shortcomings of AI, and understands AI outputs and how they relate to a specific patient. In their position statement (3), RANZCR recommend the development of multidisciplinary teams with AI developers, clinicians and medical administrators to share knowledge and skills and work to each other's strengths when implementing AI.

3.3.10 Cybersecurity

Table 20: Itemised recommendations on cybersecurity

Organisation/ agency	Document	Type	Comments, recommendations and commitments
Department of Industry, Science and Resources	Australia's AI Ethics Principles (72)	Non-clinical	Recommends ensuring that security mechanisms are in place for AI projects, including the identification of potential vulnerabilities in systems and resilience measures for adversarial attacks.
NSW Government	Mandatory Ethical Principles for the use of AI (71)	Non-clinical	Under Privacy and Security principle, advises that "NSW citizens must have confidence that data used for AI projects is used safely and securely, and in a way that is consistent with privacy, data sharing and information access requirements. Any project outcome will be undermined by lack of public trust if there is any risk of a data breach or that personal data could be compromised."
	Generative AI: basic guidance (93)	Non-clinical	Recommends that workers do not open any links generated by generative AI systems.
Commonwealth Ombudsman	Automated Decision-Making – Better Practice Guide (41)	Non-clinical	Recommends that agencies refer to the Digital Service Standards for advice on data security
Human Technology Institute	The State of AI Governance in Australia (64)	Non-clinical	Makes recommendations to ensure that projects meet cybersecurity requirements under the Privacy Act. Recommendations include ensuring organisations are destroying or de-identifying information that is no longer needed and reporting any data breaches to OAIC.
OVIC	Artificial Intelligence – Understanding Privacy Obligations (66)	Non-clinical	Refers to Victorian Data Security Frameworks which recommend certain actions based on the risk of the data being held. Recommends that organisations and agencies take measures to protect the data that they hold.
MTAA	Digital Health: Breaking Barriers to Deliver Better Patient Outcomes (92)	Clinical	Recommends implementing robust security measures and updating 'legacy' technologies because a system is only as strong as its weakest link. Recommends that the TGA reconsider their approach to device 'upgrades' to prevent patches for security

Organisation/ agency	Document	Type	Comments, recommendations and commitments
			protection being regarded as 'performance upgrades' and therefore requiring substantial red tape.
RANZCR	Ethical Principles for AI in Medicine (3)	Clinical	Recommends that data be stored securely and in line with relevant laws.
	Standards of Practice for Clinical Radiology – Chapter 9: Artificial Intelligence (82)	Clinical	States that practices should demonstrate appropriate security measures to protect patient information and implement a user registry to track access to patient information.
TGA	Medical device cyber security information for users (96)	Clinical	States that health professionals have a responsibility to report any cyber security issues directly to the TGA. They need to understand cyber security enough to adequately inform patients about the risks, otherwise patients cannot provide informed consent. Patients should feel that they can ask their physicians questions about device security.

Non-clinical

Cybersecurity was mentioned in a minority of the policies. Both the NSW Mandatory Principles and Australia's AI Ethics Principles mention security (71, 72). Both address privacy and security together, recommending that agencies and organisations ensure that private data is stored safely and securely to avoid data breaches which compromise public trust (71, 72). The Australian Government's principles recommend the identification of potential security vulnerabilities and the implementation of security measures to account for potential abuse risks of AI systems (72).

The NSW Government, in their AI Assurance Framework (67), require that projects adhere to the NSW Cyber Security Policy. Similarly, OVIC (66) refer to the Victorian Data Security Frameworks which provide guidance for project cybersecurity measures based on the risk level of the data held. The Commonwealth Ombudsman recommend that agencies refer to the Digital Service Standards for advice on data security (41). The Human Technology Institute provide guidance for ensuring that AI projects meet cybersecurity requirements under the Privacy Act, including ensuring that organisations are destroying or de-identifying information that is no longer needed, and ensuring that the OAIC is notified of any data breaches (64).

Clinical

Some of the clinical policies contained guidance or recommendations for cybersecurity measures. RANZCR, in their position statement, recommended that data be stored securely in-line with relevant laws (3). In their Standards of Practice for radiologists, they state that Practices or hospitals should demonstrate appropriate security measures to protect patient information and implement a user registry to track access to patient information (82). The TGA state that health professionals have a responsibility to report any cybersecurity incidents associated with the use of medical devices to the TGA (96). In addition, they advise that physicians should understand the cybersecurity risks of using software-based medical devices enough to explain risks to patients, and that patients should be confident asking their physicians questions about device cybersecurity (77).

The MTAA recommend that TGA reconsider their approach to device 'upgrades' to make it easier for medical device companies to update software-based medical devices with security patches (92). They recommend that security updates should not be subject to the same approval processes as performance updates, to make it easier for companies to quickly identify and ameliorate any security issues.

3.3.11 Guidance specific to pathology tests and medical imaging

Three of the documents provided specific guidance for medical imaging: the AHPRA Medical Radiation Practice Board's Guidance for Clinical Imaging and Therapeutic Radiology Professionals (88), the Australian College of Dermatologists' Position Statement on Use of Artificial Intelligence in Dermatology in Australia (79), and Chapter 9 of RANZCR's Standards of Practice for Clinical Radiology (82). Recommendations from these organisations have been addressed in the subsections above. None of the documents included in the review contained specific guidance for pathology laboratories.

3.4 Case study: the NSW Government approach to governing AI

The NSW approach to managing the risks of AI includes a series of Mandatory Ethical Principles (71), from which items were developed for a mandatory self-assessment framework for agencies implementing AI systems. The AI Assurance Framework (67), currently under review, helps agencies identify the risk level of their AI projects. For higher-risk projects, self-assessments must be submitted to the AI Review Body, which in turn makes recommendations for how to pursue the project. These recommendations are non-binding but must be documented. Responsible Officers, who must be identified in the self-assessment process, remain accountable for project outcomes.

The Mandatory Principles and the AI Assurance Framework are implemented as part of a broader NSW strategy to build capacity for the responsible introduction of AI (83). The NSW Artificial Intelligence Strategy involves broader-reaching initiatives such as capacity-building in the ICT workforce, and avenues for citizens to provide feedback on Government use of AI.

3.5 Chapter summary

This chapter presents the results of a review of Australian AI polices and legislation relevant to the acute care context. Thirty-six documents were included in the final review. Whilst Australia has no legislation directly relevant to AI, several pieces of legislation regulate the development and use of AI. The Australian Government's consultation on Safe and Responsible AI has indicated that there is support for further risk-based regulatory frameworks to ensure AI is implemented responsibly. This chapter presents recommendations from the included policies across 9 themes: engagement with consumers, patients and citizens; equity, discrimination and human or patient rights; privacy and confidentiality; evaluation, monitoring and maintenance; transparency; accountability; consent; worker training and support; and cybersecurity.

4. AI in acute care: effects on care delivery and patient outcomes

4.1 Introduction

The application of AI or ML models to improve patient care across medical disciplines has become an apex interest to decision makers and providers of healthcare in many high-income countries (97); and increasingly so in low and middle income countries (98).

During the era of deep learning (99), convolutional neural network architecture enabled by big healthcare data conditions (100) ignited research into the performance and utility of ML systems, notably in the field of medical imaging (101) but also in a range of biological signalling data settings (102-105) and in supporting clinical decision making (106). However, there are justified barriers to the implementation of ML into the healthcare settings, and the limited use of ML-enabled clinical decision support (CDS) juxtaposes with its perceived capacity to realising the quadruple aims for all healthcare: improve population health, improve patient's experience of care, enhance caregiver experience and reduce the rising cost of care (107).

A scoping review was undertaken to scan the international landscape of deployed AI in acute care settings, with a lens on safe implementation, clinical outcomes, workflow, and workforce impacts.

4.2 Search strategy and study selection

The full literature search strategy and PRISMA flowchart are available as Appendix documents (D and E) but briefly: a one reviewer search of seven databases (Medline, Web of Science, CINAHL, PubMed, PsycINFO, Cochrane and Embase) yielded 3917 articles. Citation searching of two pieces of literature (97, 108) by a second reviewer contributed to a further 50 pieces of literature. After de-duplication by Covidence and screening by one reviewer, 148 articles were assessed for eligibility by two reviewers. Seventy-five articles met the eligibility criteria and are summarised in Appendix F.

The literature search was limited to three years (2021-2023) and limited to peer-reviewed, primary research published in the English language.

4.3 Data extraction, summarising and reporting findings

Multiple characteristics of the literature were extracted and managed via an Excel Workbook. The following sections provide details of these characteristics.

4.3.1 Descriptive characteristics of studies reporting AI implementation in acute care settings

For each included study, we extracted the first author, year of publication, the country in which the research took place as specified by the author and its corresponding WHO region (109).

AI developer: Where the owner or developer of the ML system was explicitly stated in the study, it was categorised as:

1. Commercial (i.e. private company)
2. Academia (university research institute)
3. In-house (hospital)
4. Collaboration (more than one from these categories)
5. Unknown

Health Authority (HA) registration status: Similarly, if explicitly stated in the study that the ML system had HA approval or CE mark, it was categorised as 'Registered', else 'unknown' or 'pending' if an application had been lodged.

ML type: When the study described the approach used for training of the ML system, it was categorised as:

1. Supervised: when ground truth was provided (110) for example by domain expert labelling of CT images,
2. Unsupervised: where ground truth was not provided (110), or
3. Reinforcement learning: when the literature described this category of ML which learns a policy that maximises the cumulative reward over time by trial and error (110).
4. Mixed/multiple models: more than one from these categories.

Deployment: The extent to which the ML system was deployed is described as either 'Deployed' for soft or hard launch, and "pilot stage" if the study described it as a pilot, pre-implementation or pre-deployment.

Study design and the number of centres involved: These were categorised as:

a) Interventional (as in randomised), observational or health economic research, and
b) single or multi-site.

These descriptive characteristics of the literature are presented in

Table 21.

4.3.2 Clinical characteristics of studies reporting AI implementation in acute care settings

To consolidate the ML system's role in the clinical setting, a second set of variables were extracted from the literature and are presented in Table 21.

Medical Specialty: The key healthcare professional engaged in the use of the ML system was extracted and defined as per the Medical Board of Australia's list of specialties (111).

Disease area: Categorisation arose iteratively and was consolidated after extraction of all the literature.

Clinical task: Clinical tasks supported by the ML system were categorised into (8):

1. Diagnosis: assisting with the detection, identification or assessment of disease, or risk factors.
2. Triage: assisted with prioritising cases for clinician review, by flagging or notifying cases with suspected positive findings of time-sensitive conditions, such as stroke.
3. Procedure: assisted users performing diagnostic or interventional procedures (an action intended to achieve a result in the delivery of healthcare such as determining, measuring or diagnosing a condition or a parameter).
4. Treatment: provided recommendations for therapy.
5. Monitoring: assisting clinicians to monitor patient trajectory over time.

ML system level of autonomy: The extent by which the ML system performs a task independent of a clinician was examined using a previously published three-level classification based on how clinical tasks are divided between the clinician and AI (8):

1. Assistive: The ML system and the clinician contributions to the task overlap, but the clinician provides the decision on the task. Such overlap or duplication occurs when clinicians need to confirm or approve ML system provided information or decisions.
2. Autonomous information: In these systems, there is a separation between ML system and clinician contributions to a task, with the ML system contributing information that clinicians can then use to make a decision e.g. an imaging system that provides a coloured imaging display to help a clinician differentiate human tissues.
3. Autonomous decision: Here the ML system makes the decision for a clinical task, which can then be enacted by clinicians or the ML system e.g. a ML system screens CT images for intracranial haemorrhage and automatically notifies, flags, and prioritises clinician review of images identified as showing evidence of a bleed.

Stage of human information processing: The way in which humans process information was broken down into four distinct stages (10). In an effort to assess human-machine interaction, these four stages of human information processing have been given an 'automated by a machine' analogy (10). For each extracted study, the stage/s of information processing that the ML system performs were categorised as (8):

1. Information acquisition: The ML system automates data acquisition and presentation for interpretation by clinicians. Data are preserved in raw form, but the device may aid presentation by sorting, or enhancing data.
2. Information analysis: The ML system automates data interpretation, producing new information from raw data. Importantly, interpretation contributes new information that supports decision making, without providing the decision. For example, the quantification of QRS duration from electrocardiograms (ECG) provides new information from ECG tracings that may inform diagnosis without being a diagnosis.

3. Decision selection: The ML system automates decision making, providing an outcome for the clinical task. For example, prompting and thereby drawing attention to malignant lesions on screening mammograms indicates a device decision about the presence of breast cancer.
4. Action implementation: The ML system automates implementation of the selected decision where action is required. For example, an implantable cardioverter-defibrillator, having decided defibrillation is required, acts by automatically delivering treatment.

Information Value Chain outcome measures: Reported effects of the ML intervention were categorised as per the outcome measures in the established theory, design and evaluation framework called the Information Value Chain (112). When applied to healthcare information systems, the Value Chain leverages a five-step chain (beginning with a user interacting with an information system, receiving new information, decision change, care process altered, and outcome changed) with decision science to help yield the specific benefits of a given technology and why expected benefits may not transpire.

1. Interaction outcomes: when the research described user experience outcomes such as System Usability Scale (SUS) scores, or adoption metrics of the technology by relevant clinicians.
2. Information retrieval outcomes: when the ML system provides new information, the study quantifies the new information received.
3. Decision change: the study quantifies decision changes as a result of the use of the ML such as incorrect or correct decisions or decision velocity.
4. Care process altered: the study reports outcomes related to care process change (i.e. referrals made, treatment changed)
5. Outcome change: when the research described clinical outcomes such as number of detected polyps, patient reported outcomes such as EQ-5D or there were safety outcomes.

4.3.3 Exemplar case studies

After full literature review, three studies were selected as exemplars in study design and execution of deploying ML systems in an acute setting. One study was chosen for its broad deployment across 21 hospitals (113), another that included radiologists (114), the highest impacted specialist according to the literature search; and a third case study of ML deployed in a Cancer Pathology setting (115) – the disease area with the highest number of deployed AI according to the literature search.

4.4 Results

4.4.1 Key characteristics of literature

Appendix F provides a summary table of the 75 articles that reported a total of 76 ML systems deployed in acute healthcare settings.

Kanbar et al. (116) described two deployed models in two separate case studies, in the one piece of literature. The literature spans 20 different countries, with the Americas and West Pacific Regions accounting for 78% of the identified literature. The countries that dominated these regions were the USA (n=30/34) and China (n=12/25) respectively.

The type of AI/ML system was not always well described in the literature, with 40% of the research studies using ambiguous terminology- phrases such as 'Deep Learning' or 'Machine Learning'- with no further specifics, or referring the reader to previously published literature for full details. Supervised ML models were the most often described ML model (49%), with ground truth labelling of the training dataset by clinical experts.

Studies were largely observational (n=61), with a single vs multi-site vs unknown number of sites split of 66%, 30% and 3% respectively; and covered a broad range of study designs and methods: retrospective, prospective, quantitative, and qualitative. Two health economics research articles were identified. Of the 13 interventional studies, nine were multi-centre studies.

4.4.2 Device health authority approval or CE mark

The health authority (HA) approval status or CE mark of the specific ML system was mentioned in several studies; but with 78% of the literature not stating ML system HA approval, CE mark, or exemption reasons, this regulatory aspect went largely unmentioned in the literature.

Two articles spoke to pending application or being exempt due to the device being investigational (117, 118). Four studies utilised devices with a CE Mark: Navoy sepsis model was described as a CE mark SaMD (119), e-Stroke Suite with a CE mark (120), a case-based reasoning algorithm for antibiotic prescribing CDS with a CE mark (121)and a health economics study that looked at CE and/or FDA cleared devices that aided detection of vessel occlusion in acute stroke (122). Aside from this health economics study, a further ten use cases of FDA approved ML algorithms and devices were present in this literature. One had breakthrough authorisation – assisting users to acquire point of care cardiac ultrasound for COVID-19 patients (123), seven were for stroke indications, one for chest x-ray image analysis (124) and another ultrasound device for cardio-respiratory evaluation in COVID-19 patients (125).

Table 21: Characteristics of studies reporting AI implementation in acute care settings.

Characteristic	N= (76)*	%
Year of publication		
2023	10	13
2022	39	51
2021	27	36
WHO region (109)		
Americas	34	45
Western Pacific	25	33
Europe	12	16
Eastern Mediterranean	3	4
South-East Asia	2	3
African Region	0	0
Country		
USA	30	39
China	12	15
Taiwan	5	6
Rest of World^	31	41
AI developer		
Commercial	29	38
Collaboration	20	26
In-House (hospital)	14	18
Academia	5	6
Unknown	8	11
Type of machine learning (ML)		
Supervised ML	37	49
Unsupervised ML	2	3
Reinforcement learning	1	1
Mixed/multiple models	5	6
Ambiguous terminology or not described	31	41
Health Authority (HA) registration or CE mark status		
Unknown	59	78
HA registered	11	15
CE mark	4	5
Exempt or pending	2	3
Deployment status		
Deployed (soft or hard launch)	54	71
Pilot stage	22	29
Study design		
Observational	61	80
Interventional	13	17
Health economics research	2	3
Study sites		
Single Site	44	58
Multi-centre	27	35
Unknown	5	6

*Kanbar et al. 2022 – two deployed models in two separate case studies described in the one piece of literature.
^ 17 countries with ≥ 3 publications

4.4.3 Medical specialists and ML use

Described in Table 22, of the 18 different medical specialties identified in the literature, the most prevalent medical specialist was the Radiologist (n=25/76) and by association Radiology related allied health roles. For Radiologists and their allied health professionals, much of the ML deployed assisted with the analysis of Computer Tomography (CT) images for multiple disease areas: cancer of the colon or lung, COVID-19, renal, respiratory, stroke and orthopaedics: rib fractures and cervical spine fractures. Other imaging modalities included one study describing thyroid nodule detection via ultrasound (126), one study of cardiac MRI (127), one study of hand x-ray images (118) and two studies of chest x-ray images (124, 128). Radiologists were also the healthcare professional most exposed to health authority registered or CE mark devices (n=10/16).

Infectious disease physicians in both the adult and paediatric settings (129) were the next most frequent specialist implementing ML, predominately for sepsis (n=4) and COVID-19 (n=4) but also pneumonia (130) and the use of a ML-enabled CDS to assist with antibiotic prescribing (121).

4.4.4 Disease areas summary

Through iterative categorisation, the literature yielded 24 different diseases areas in which ML systems were deployed.

Oncology (solid tumours) represented the most frequent disease area in the literature (n=17/76) and encompassed several facets of cancer care from screening: colorectal (n=4) and gastrointestinal (131), lung (132) and skin melanoma (133), to grading (n= 3), localising (n=2), radiation treatment planning (n=4) and forecasting acute emergency admission of patients undergoing cancer treatment (134). Cancer studies accounted for six of the 13 studies that were interventional design. Of those six interventional studies, five were a type of randomised control study design in either colorectal cancer screening or gastric cancer screening. As such, these were endoscopy led procedures that had ML systems embedded to assist with identifying polyps or adenomas.

Stroke was the second most common disease area in the literature (n=11) with ML systems deployed almost exclusively for diagnosis of stroke (n=10). The single other study was a deep-learning algorithm used to process ('de-noise') CT perfusion images, thereby enhancing the effectiveness and safety of thrombolytic therapy given in acute cerebral infarction (135). All 11 studies used ML image analysis of head CTs, and seven of the ML systems were classified as having a high level of autonomy (autonomous decision). This is because analysis of the CTs was automatically triggered upon image acquisition, with any suspected cases automatically prioritised for physician review via a secure messaging platform or via the existing image database system (PACS and RIS).

Eight stroke studies described the ML system as either having a CE mark and/or FDA approved, with the remaining 3 having no explicit statement. The majority of stroke studies were of observational study design (n=9), with the remaining two being one each of Health Economics (HE) research (122) and interventional design. The interventional study was a stepped-wedge randomisation design where all sites eventually received the intervention (114). The HE research highlighted some cost-saving potential and QALY gains for using ML to aid detection of vessel occlusion – this is described in more detail in section 4.12.

Respiratory related conditions were the third most common health area in which ML was deployed (n=8). The disease indications in this group were diverse: pulmonary nodule detection during emergency department CT scans (117), pneumonia (130), acute respiratory distress syndrome ACDS (136), chest x ray (124, 128), point of care lung ultrasound (43), extubation assessment (137) and pulmonary embolism

(138). Three of the solutions were clinical decision solutions (130, 136, 137), with the remaining five studies being image analysis for diagnosis tasks. The health data modalities were diverse and included electronic health record (EHR) data inclusive of pathology reports, radiology reports, progress notes and vital signs. Seven out of eight studies were observational, with one study that implemented a stepped-wedge cluster controlled interventional trial (130).

4.5 Clinical tasks to which AI has been applied.

4.5.1 Diagnosis

Diagnosis tasks were the most frequent clinical task in which AI was deployed in this literature search (n=40), straddling 15 of the 24 identified disease areas. In the literature, diagnosis was often aided by ML image analysis, but also image registration and image reconstruction. Aside from imaging data, there was also diagnosis via ML analysis of other health data modalities, namely physiological parameter data and EHR data.

Diagnosis through ML assisted image analysis was the most prevalent. These were CT imaging (n=14), x-ray (n=3)(118, 124, 128), echocardiogram (n=1) (139), ultrasound (n=1) (125) and other medical images: skin lesions (n=3)(133, 140, 141), fundus eye image (n=1) (142) and whole slide image analysis (n=2) (115, 143).

There were two instances of ML being utilised for image registration – use of fixed image and moving image - which was ultrasound video and stills to provide clinical grading of thyroid nodules (126) and in point of care lung ultrasound for acute respiratory distress syndrome (ARDS) (43). Image reconstruction by ML enhancement of acquired images was found in one study (139). Other physiological parameters were also used for diagnostic purposes – ECGs for acute myocardial infarction detection (144), SA node changes to predict sepsis in NICU patients (129) and EEG patterns for sleep studies (145).

Nine studies had ML systems that leveraged EHR data for diagnosis – some inclusive of pathology reports and other physiological parameters such as oxygen saturation. These were predominantly for sepsis diagnosis (n=4) and COVID-19 (n=2) (73, 146), but also included epilepsy to forecast surgical intervention candidacy (116), assessing specific adverse outcomes for patients undergoing hip surgery (147) and a pneumonia CDS (130). Only two studies described their ML system as a CDS (130, 148).

Diagnosis used a wide range of ML systems where specified: natural language processing, deep neural networks with classifiers (such as random forest, logistic regression) or TRIER algorithms used to extract waveform shapes for convolutional neural network (CNN) training)(145).

4.5.2 Triage

Seven studies described the use of ML for triaging tasks, five used EHR data with two specifying the ML was a trained Natural Language Program (NLP) (116) and Open-NLP (149). Two triaged for COVID-19 severity (150, 151), one triaged chest pain (152), one triaged pulmonary embolism (138) and two for patient acuity (149, 153). One ML system was used to identify potential paediatric clinical trial participants (116).

4.5.3 Procedure

Thirteen studies described ML systems utilised during medical procedures. These were all imaging-based procedures with three using analysis of CTs, two using ultrasound and six with colonoscopy videos and stills. These were in just three clinical areas: cancer (n=10) of which seven were screening for colorectal

cancer, one for gastric cancer (154) and two for CT auto-segmentation for organ at risk planning in breast cancer (155) and prostate cancer (156), cardiovascular disease (n=2) and orthopaedics (n=1). In the orthopaedics study, the ML was used by anaesthesiologists to enhance image quality and thereby find the optimal injection point to provide effective nerve block during scapular fracture surgery (157).

4.5.4 Treatment

There were five studies describing ML aided treatment tasks in a diverse set of diseases: an EHR embedded algorithm to provide a recommendation on the number of pack red blood cells to transfuse to a patient (158), a deep-learning based CT analysis system to evaluate effectiveness and safety of thrombolytic therapy for cerebral infarcts (135), a ML system to predict radiation dose distribution when generating treatment plans for breast radiotherapy (159), a case-based reasoning algorithm for antimicrobial prescribing decision support (121) and an ML prediction model that advises on extubation readiness (137). Three of these five studies used EHR data inclusive of physiological parameters (121, 137, 158) whilst the remaining two used CT imaging analysis to guide treatment decisions.

4.5.5 Monitoring

Eleven articles described ML in the context of patient monitoring. Six of the monitoring systems could be classified as monitoring for acute deterioration: clinical deterioration (113, 160, 161), adult and neonatal mortality (162, 163) respectively, and risk of oncology patients presenting at emergency departments (134). A suicide prediction model was silently deployed in Epic eMR to calculate real-time suicide risk (164).

Two studies described use cases of ML monitoring chronic conditions: ML-enhanced renal CT imaging was used in a chronic kidney disease context (165) and another was for monitoring ulcerative colitis (166).

Aside from these, two articles described the monitoring algorithm in context of a CDS: one AI enabled CDS was used for six-hourly venous thromboembolism (VTE) monitoring (167), the other was to promote lung protective ventilation from possible ARDS (136). This was an NLP-enabled CDS synchronous alert tool that was associated with existing computerised ventilator protocols and targeted patients with possible ARDS not receiving Lung Protective Ventilation.

4.6 Role of AI

4.6.1 ML system autonomy

Over two thirds of the literature described ML systems that had limited autonomy (assistive or autonomous information). An example of an autonomous information ML system was the use of a deep-learning computer-aided polyp detection system for colorectal cancer screening, with a real-time visual prompt of a green box indicator and an audible sound on detection of a suspicious lesion (168). An example of assistive ML system was auto-segmentation of breast cancer CTs to delineate organs at risk (OaR), which the radiologist could either correct or accept (155).

Eighteen articles described ML systems with autonomous decision-making, a higher level of autonomy. Fifty percent of these were characteristically in areas that had time critical decision making: stroke (n=8) and acute myocardial infarction (n=1).

The remaining 50% comprised of other use cases: COVID-19 related triaging and screening (73, 125, 146) to assist with appropriate levels of required care and reduce COVID-19 exposure to medical personnel; a Quality Assurance (QA) program with AI assisted image analysis of Lung CTs combined with NLP assisted

analysis of CT reports to identify pulmonary nodules, followed by prioritisation of the radiologists' workflow (117), and EyeWisdom that detected and graded diabetic retinopathy severity from fundus images, generating a one page report (142).

Table 22: Characteristics of AI implemented in acute care settings.

Clinical characteristics	N=76*	%
Medical specialty (111)		
Radiology	25	33
Infectious diseases	9	12
Emergency medicine	7	9
Gastroenterology and hepatology	6	8
All other specialties^	29	38
Disease area		
Cancer-solid tumours.	17	22
Stroke	11	14
Respiratory	8	11
Sepsis	5	7
COVID-19	5	7
All others[$]	30	39
Clinical task supported (8)		
Diagnosis	40	53
Procedure	13	17
Monitoring	11	14
Triage	7	9
Treatment	5	7
Level of autonomy		
Assistive	30	39
Autonomous information	28	37
Autonomous decision	18	24
Stage of human information processing (10)		
Information acquisition	7	9
Information analysis	12	16
Decision selection	57	75
Action implementation	0	0
Information Value Chain outcome measurements (112)		
Interaction	26	34
Information received	43	57
Decision changed	34	45
Care process altered	22	29
Outcome changed	47	62
Value and quality assurance measures		
Cost-effectiveness	4	5
AI/ML system performance metrics reported/not reported	46/30	60/40
AI/ML system post-deployment auditing reported/not reported	2/74	3/97

^ 14 other medical specialties.
*Kanbar et al. 2022 – two deployed models in two separate case studies described in the one piece of literature.
[$] 19 other disease areas with count of ≥4.

4.6.2 Human information processing stages

The autonomous decision ML systems found in the literature were exclusively automating the decision selection stage of the human information process (n=18). Of the 28 ML systems that were providing autonomous information, they were automating the decision selection (18/28), information analysis (4/28) and information acquisition stages of the human information process (6/28). ML systems that were assistive were predominantly automating decision selection (21/30).

4.7 AI system performance

AI model evaluation metrics such as sensitivity, specificity, positive predictive value, accuracy and F1 score were commonplace amongst the literature. Section 4.11 describes how some of these metrics have knock-on effects with clinical decision making.

There were 49 studies that assessed the ML system performance against a comparator. This was accomplished via several types of study design found across the literature:

- *Before and after studies / historical cohort studies:* 11 studies compared the AI model performance against an epoch where there was no AI in place.
- *Prospective data collection studies:* AI vs ground truth by expert consensus (n=5), AI vs lateral flow in COVID19 triage (73), AI-CDS for COVID-19 deterioration: silent deployment vs visible (151), AI vs human vs human corrected AI (n=3) (133, 155, 164) and human correcting AI (156).
- *Retrospective data studies:* (n=2) historic CTs reviewed by AI vs reviewed by physician (132, 169).
- *Randomised control studies, cross over studies or wash out phase studies:* (n=25) there were mixed methods here wherein physicians used the AI or did not in a particular task. Mostly this was achieved via randomisation (n=10), or by repeating the procedure, first with AI and then not with AI (n=5). Two studies were tandem procedure i.e. first performed without AI and then immediately repeated with AI (170, 171), whilst three studies had a wash-out period method, whereby the same task on the same sample was performed first with/without AI and then again 4 to 12 weeks later (115, 143, 172).

Twenty-seven studies did not compare the AI model performance against anything. These tended to be pieces of literature that were case base reports sharing implementation learnings, or user experience research from interview or survey data. In Adams et al. study (173), all patients were exposed to the AI, but the outcomes of interest was to explore the impact timely physician interaction had on clinical outcomes by comparing patients that had physician interaction within 3 hours of the TREWS sepsis alert, compared to those who responded later than 3 hours after the alert.

4.8 Clinical workflow integration

Evidence of pre-implementation, deployment, and post-implementation activities to assist with clinical workflow integration were located across the literature.

4.8.1 Training dataset alignment

Twenty-four studies (approximately one-third) gave some description of training the algorithm on a localised dataset. These were often for algorithms that were developed in-house, by academia or collaboratively between the two (n=18), such as Thynet DL algorithm for diagnosis of thyroid malignancy trained on more than 18,000 images from two local hospitals (126) and Hinson et al.'s work-flow integrated an ML COVID-19 triaging system – trained on more than 21,000 emergency department visit data extracted from the five hospitals where it was implemented (151). There were five cases where

commercially sourced algorithms were further trained and evaluated on local data sets including the KATE triage model evaluated on 800 local medical records (149) and ENDOANGEL (154).

4.8.2 Engagement with hospital ethics committees or clinical governance boards

Whilst many of the studies stated research ethics or institutional review board endorsement was sought to conduct the research and was granted, there were no specific reference found in any of the literature that demonstrated engagement of clinical governance or ethics as part of the workflow integration strategy or responsible use check.

4.8.3 Integration with existing IT infrastructure

Forty-one studies mentioned integration with existing IT infrastructure including electronic medical records and imaging databases like Picture Archiving and Communication System (PACS) and Radiology Information System (RIS). Thirty-one studies did not mention IT integration in their report, and four studies showed that their ML system was not fully integrated into existing IT infrastructure. These were Glissen-Brown et al.'s CADe system installed on a separate computer system (170); the COVID-19 triaging system as described by Garzon-Chavez et al, where CT images were uploaded to the Huawei Cloud AI for analysis and reporting back (150); Zhang et al. deploying the prototype DL software on an independent workstation for rib fracture detection accuracy and reading efficiency (169); and Kermani et al.'s description of a web-based Case Based Reasoning prediction system for neonatal survival (163).

4.8.4 End user engagement

Eleven studies described end user engagement during the development or deployment of the ML system, with 66 studies not reporting on this aspect of workflow integration. The end user was always a healthcare professional and were generally contributing either domain knowledge expertise or being part of the change implementation process as a local champion.

Most of the 11 studies that had end user engagement described, were implementing ML systems developed in-house (hospital) (n=7/11), one was a commercially sourced ML system (20), whilst the developer was not specified in other three studies. Table 23 describes the end user engagement undertaken in these studies.

4.8.5 End user training

Twenty-three studies described some form of end user training, either as a condition for participating in the research or ahead of the full deployment or pilot phase deployment of the ML system.

4.8.6 Other implementation steps

Nine studies had notable information about the implementation of the ML system. Jordan et al. (153) examined the way emergency department triage nurses understood, contextualised, and incorporated the KATE CDS system into their own understanding and practice of triaging. The researchers perceived an initial negative attitude towards the CDS system at implementation due to the quick rollout and perceived lack of explanation. They emphasised, "*communication of clearly defined benefits for improving both nursing practice and patient outcomes within the known context is also important for widespread adoption*".

Two studies highlighted efforts to 'build product to market fit' (114, 129), by offering support to customise and implement the HeRO system based on local clinical protocols for predicting neonatal morality, or

assist with modifying scanning protocols to enable upload to a cloud-based AI, Viz.AI, respectively. Three studies described piloting (for up to one year) prior to wider deployment (113, 116) and another described a small scale real clinical scenario feasibility test to validate the diagnostic efficiency of the DeepCT system for detecting intracranial haemorrhage (174). Dean et al. (130) and Adams et al. (173) refer readers to sister publications that describes implementation in further detail.

4.8.7 Post deployment quality assurance

Three studies described post-deployment ML monitoring, auditing or any other type of performance review. Boussina et al. (175) described a plan for their DL model for early prediction of sepsis to be monitored weekly for performance parameters and to identify potential model drift. Dean et al. (130) deployed ePNa: a NLP enabled-clinical decision support system using real-time and historic EHR data extraction for pneumonia diagnosis, risk stratification and antibiotic therapy. As part of the post implementation phase, they described having "*ongoing technical support…and study authors conducted audit and feedback at regular intervals*".

Martinez et al. (113) described the post deployment work done for the Advance Alert Monitor program (AAM) for in-hospital clinical deterioration. "*…local oversight and performance improvement plans were put in place*", "*Quality tracking dashboards supported short- and medium-term monitoring…Each local facility had an AAM long-term oversight structure*", "*Weekly and monthly performance dashboards were reported to support ongoing performance improvement and evaluation*." The implementation of AAM is described in detail in section 4.13 Exemplar Case study 1.

Table 23: End user engagement described in the literature.

Study author	ML system	End user engagement	Level of involvement
Ou et al. 2022 (143)	Count and classify atypical urothelial cells from whole-slide images (WSI) for urine cytology.	Cytopathologist	Evaluation of model by review of inference results and providing feedback, propagating rounds of iterative improvements to the model before satisfactory performance was achieved ahead of workflow deployment.
Martinez et al. 2022 (113)	Advance Alert Monitor (AAM), to improve early detection and intervention for in-hospital deterioration.	Virtual quality nurse consultant, Rapid response team nurse, hospitalists, ICU physicians and nurses, social services and palliative care leaders.	Increased the surveillance to hourly from every six hours, increasing the sensitivity from 25% to 50%, eliminated the frontline need for 24/7 vigilance, and minimised alert fatigue by allowing for interpretation and strategic silencing of alarms based on clinical realities defined and refined by active engagement with frontline feedback. Local site champions with defined roles and responsibilities were involved in the workflow and other aspects of deployment (communication, safety culture).
Kanbar et al. 2022 (116)	ACTES: an NLP based automated clinical trial eligibility screener for real-time identification of patients for research studies in a paediatric emergency department.	Clinical Trial Research Coordinators (CRC) and physicians	*"AI solutions were designed and integrated with feedback from end users. The epilepsy and ACTES corpora were created by manual annotation of patient notes by providers. Throughout the algorithm design and implementation process, providers were included in the build and ultimate integration. First, the biomedical informatics team shadowed providers for workflow observation. Second, the biomedical informatics team attended clinical meetings that included faculty, staff, and clinical research coordinators for a minimum of 10 hours to get feedback and ensure the design was appropriate. Third, mock-up designs were shared at a minimum of 3 meetings to discuss the process of using and interacting with the AI solution in the form of a CDS tool. In cases where the CDS tool could provide an alert, the providers were consulted on their preferred alert method (e.g. email or text message alerts)."*
	EPILEPSY ID: generates surgical candidacy score for each patient using NLP	Epileptologists	
Hinson et al. 2022 (151)	EHR embedded clinical decision support (CDS) system that leverages ML to estimate short-term risk (scoring 0-10) for clinical deterioration in patients with or under investigation for COVID-19.	Emergency clinicians	*"A system to generate patient-level risk estimates and deliver EHR embedded CDS to emergency clinicians in real-time was developed with software engineers and end-users under a human-centered design framework".* *"…model outcomes were translated to one of ten COVID-19 Deterioration Risk Levels using risk thresholding; thresholds were determined by consensus between technical and clinical team members using graphical plots, calibration curves, and outcome frequency tables."* *"CDS content and appearance was developed iteratively, guided by direct feedback from prospective end-users".*

Study author	ML system	End user engagement	Level of involvement
Choudhury et al. 2022 (158)	AI-based Blood Utilization Calculator, delivers data-driven personalised recommendations for the number of packed red blood cells to transfuse for a given patient.	Any clinician who used the BUC.	Post deployment: study deployed to understand low user engagement with the BUC.
Li et al. 2022 (147)	An ML-based application assisting anaesthesiologists in assessing specific adverse outcomes for patients required to undergo hip repair surgery.	Broadly defined as "clinical expert"	Feature selection: clinical expert opinions were used to select 22 preoperative variables from the Local Hospital Information System dataset from which to calculate risk of adverse events of interest.
Dean et al. 2022 (130)	ePNa: a CDSS to guide pneumonia diagnosis, risk stratification, microbiological studies, site of care and antibiotic therapy.	Broadly defined as "clinician"	*"…active clinician engagement in tool development and deployment"*
Hwang et al. 2022 (145)	A CDSS that automatically score sleep studies from EEG patterns and other physiological data collected during sleep studies (Polysomnography).	Sleep technicians	*"To design a CDSS within this framework, our development process included three phases: (70) interviews with polysomnographic technicians to identify why users might desire explanations from the CDSS when adopting AI-based sleep scoring systems, (2) user observations of how polysomnographic technicians score sleep stages from EEG recordings to determine the information that could help them, and (3) an iterative design process to construct a user-friendly CDSS interface that addresses the formulation of explanations in the system. After development, the polysomnographic technicians performed quantitative and qualitative evaluations of the system."*
Martinez-Gutierrez et al. 2023 (114)	Cloud-based AI-algorithm (Viz.AI) trained to detect Large Vessel Occlusion, Acute Ischaemic Stroke.	CT technologist Multi-disciplinary team	After the information technology security review had been completed, team members from the vendor then began working with CT technologists to modify CT acquisition protocols such that images would be sent at the time of acquisition to the cloud-based AI server. Weekly team meetings occurred between representatives from the vendor and a team consisting of each campus's stroke coordinators and lead members from neurology, radiology, emergency medicine, neuro-intervention, and nursing to monitor progress.

Study author	ML system	End user engagement	Level of involvement
Wang et al. 2023 (152)	AI based triage system: detect ST-elevation myocardial infarction (STEMI) on electrocardiography (ECG), and a computerised risk score provide a clinical risk score (ASAP) to prioritise patients for ECG examination.	Three board-certified cardiologists	12-lead ECG training dataset labelling with their consensus serving as the ground truth. To evaluate the model's performance before its deploying, an additional 4007 twelve-lead ECGs from patients in the ED were tested against the consensus (ground truth) of 3 board-certified cardiologists, and these 4007 twelve-lead ECGs were the internal test cohort.

4.9 Usability of AI

4.9.1 User interaction with AI

As described in section 4.3.2, the Information Value Chain begins with system interaction, and so the literature was initially searched for any measures of user interaction with the ML system. Twenty-three studies yielded insights into interaction. Broadly, these were in the form of user experience measures (n=19) and/or user adoption metrics (n=13). For example, Choudhury et al. described 119 clinicians out of 273 having utilised the AI-based Blood Utilization Calculator (BUC) embedded in the EHR, and through collecting scores to validated questions from the extended unified theory of acceptance and use of technology UTAUT-2, agreed that by engaging with the BUC system, it could improve patient outcomes and did not put them or their patients at risk (158).

4.9.2 Usability assessment

Surveys and questionnaires were employed in 15 studies, commonly capturing feedback from the end user particularly via the validated System Usability Scale, but also Likert scale-type questions, satisfaction scores and open-ended questions. Two studies that focused on delineating OaR for radiation therapy planning put questions towards physician-perceived quality of automated contours (155, 156). A subjective analysis was performed by four radiation oncologists in the Kneepkens et al. study (159), performing a blind comparison of three plans and judging them for clinical acceptability and ranking based on preference. Kermani et al. asked physicians to rate they confidence and acceptability in the outputs generated by the Case Based Reasoning prediction system for neonatal survival (mortality risk score) and length of stay (163).

In-person interviews were conducted in four studies (137, 145, 153, 161). Interviewee numbers were typically smaller than the number of respondents to surveys, and topics were varied. Hwang et al. explored via 10 participant interviews, topics such as trust, impact on workload, helpful aspects and unhelpful aspects of a deployed CDS for sleep stage tasks (145). Jordan et al. interviewed 13 emergency triage nurses to explore the cultural and technological elements of the implemented CDS KATE (153).

In addition to broader questionnaires, Sarti et al. (137) interviewed 15 respiratory therapists experienced with using the Extubation Advisor Tool (EA) and used thematic analysis to deduce facilitators and barriers to EA's implementation. Schwartz et al. (161) conducted 17 interviews with clinicians who had used CONCERN – a CDS system that can predict in-hospital deterioration. The human-computer trust conceptual framework (176) was used to explore clinicians trust whilst interacting with CONCERN.

4.9.3 Use metrics

Adoption metrics were described in varied detail in 13 studies. As an example: Dean et al. (130) described user metrics for ePNa, a real time ML-CDS system for pneumonia diagnosis deployed across 16 community hospitals. Overall, ePNa was used by the ED clinician in 67% of eligible patients with pneumonia after deployment. Use was 69% in the 6 larger hospitals but 36% in the 10 smaller rural hospitals. Seyam et al. showed 3017 out of 4450 patient CT scans used the AIDOC system (177).

Rabinovich et al. (128) described metrics and user satisfaction in context with the four determinant factors in the Technology Acceptance Model (178). Actual system use determined via interface access metrics showed that it was accessed in 15% of x-ray studies (n=1186), with an average of 8 accesses per day over a 5-month period. Perceived ease of use was measured through a validated survey using the System Usability Scale (SUS).

4.10 Effects of AI on clinical decision-making

When interacting with a ML system, the information provided by the ML system to the human user may precipitate a decision change, the third step in the Information Value Chain. Study designs that compare human alone vs. human assisted with AI are compelling arenas in which to look for superior or non-inferior decision making, and inferior decision making through phenomena such as automation bias (an over reliance on the AI). This literature search yielded 21 studies that had this study design.

The aspects of decision change that are of interest here are the accuracy of the information that persuades the end user, and whether it accelerates that decision making time. For example, when Rabinovich et al. surveyed radiologist residents who had used TRx (an AI-based system for automated detection of chest x-ray findings), they perceived a poor performance for lung opacities due to many false positives (128). When Byun et al. evaluated the integration of an auto contouring system for delineating OaR for breast radiotherapy, manual contouring took a mean time of 37 minutes to complete, compared to 6 minutes when the radiologist was correcting the AI's contouring attempt (155).

All literature was examined for evidence of incorrect decisions by the ML algorithm – principally false negative and false positive rates, and quantification of decision velocity changes. More than half of all the literature reviewed (62%) described either of these two aspects of decision change. Indeed, decision change outcomes and clinical outcomes were the most reported aspects of the information value chain found in this literature.

4.10.1 False-positive and false-negative rates

Just under half of all the literature reviewed (45%) described decision accuracy parameters in their research. As an example, Cerminara et al. (38) conducted a prospective observational study in a skin cancer screening setting, comparing 2D Convolutional Neural Network (CNN) enabled Total Body Photography vs 3D CNN TBP vs dermatologist alone vs dermatologist assisted by either AI. They found that overall, dermatologists performance deteriorated after AI collaboration (AUC-ROC without AI 0.91 vs. with AI 0.88), with the false positive rate increasing when the dermatologist utilised AI in their decision that was subsequently compared to the histology ground truth of 75 excised skin lesions.

Eng et al. (118) detected possible automation bias at one of 6 sites that participated in their randomised control trial comparing accuracy of skeletal age assessments of hand x-ray examinations performed with or without AI. Radiologists assisted by the AI had a higher diagnostic error than radiologist who did not (control), although this result was not quite significant (mean age difference, 10.9 months [with-AI] vs 9.4 months [control]; p= 0.06).

In a demonstration of non-inferiority and use of AI in less experienced physicians, Alessandro et al. (131) conducted a multi-centre randomised controlled trial of 10 non-expert endoscopists performing a variety of colonoscopies with and without real-time deep-learning computer aided detection (CAD) (25). This was to elucidate whether physician experience has influence on the surrogate outcome parameter of adenoma detection rate that has become the mainstay quality measure in AI assisted colonoscopies. They did demonstrate non-inferiority in the ADR detection rate for the CADe assisted physicians, and when adjusting for age, gender and indication, demonstrated superiority. However, the non-neoplastic polyp resected rate (the 'unnecessary' polypectomies) were 12.1% and 11.8% in the CADe and control group respectively–indicating a false positive rate influence.

Confidence, acceptability and trust are facets of this human-machine decision-making process and were well captured in Schwartz et al.'s study (68). Using the 'Human-computer trust conceptual framework" (84), seventeen clinicians were interviewed at two hospitals with the objective of elucidating the factors that influence trust in a workflow integrated CDSS for in-hospital deterioration (CONCERN). Whilst the participants had been using CONCERN for 1-6 months, perceived understandability and perceived technical competence of the CDSS influenced clinicians trust in it, as deduced by thematic analysis of the interview data by two coders that generated the highest Cohen κ coefficient of agreement. Perceived understandability is "the sense that the human supervisor or observer can form a mental model and predict future system behaviour", and clinicians described their want to evaluate the factor contributing to the CONCERN score. A quote was provided by the research team to encapsulate that want: *"..what do we think is contributing to that or even reviewing...what went into that. And just be like, do we trust this? Do we not?"*. Perceived technical competence is defined as "the system is perceived to perform tasks accurately and correctly based on the information that is input", and as an example, a physician said during interview "*the more accurate it is, in my opinion...the more trust I have in the tool*".

4.11 Effects of AI on care delivery and patient outcomes

As a consequence of interacting with an ML system, the user may alter from their usual course of care process (i.e. make a referral to another healthcare provider or switch to a different therapy). Lastly, the human-machine interaction could alter a clinical, patient reported or safety outcome. Both stages can have significant impact on resources, process, standard of care and quality of care given to patients. The literature was searched for care process change and outcomes change.

4.11.1 Care process change

This aspect of the information value chain was not well described in the literature, with 28% of the literature commenting or measuring a change in the process of care.

Some positive examples of changes in care process were the observed 14% increase in the appropriate prescribing of anticoagulant drugs for hospital acquired VTE as a result of the implementation of an AI-enabled CDS for VTE risk (167), and greater rates of appropriate antibiotic selection as a result of implementing a case based reasoning CDS to support antimicrobial prescribing decisions (121).

Knighton et al. (136) evaluated service outcomes - defined as the impact the intervention has on the clinician and related work processes- after deployment of an NLP-enabled alert tool that identifies possible ARDS. This was achieved quantitatively by measuring Lung-protective-ventilation guideline nonadherence (where possible ARDS was detected yet the patient treatment was nonadherent to guidelines) and measuring how many of the recommendations generated by the alert tool were followed, so as to quantify over or under use of resources and services as a consequence of the deployed tool. Over the study time frame, 2876 individual alert messages sent contained 3281 individual recommendations (some alert code types included more than 1 recommendation) grouped into two general categories: recommendations promoting adoption of the computerised ventilator protocols (34% of alert recommendations) or (2) specific ventilation recommendations to those already using the computerised ventilator protocol (66% of alert recommendations) (70). Overall, 48% of the recommendations were followed within the defined adherence timeframe.

Other studies reported care process change through user experience or qualitative data collection. 27/35 users of the PTIM score (a ML EHR-embedded calculator for prediction of mortality) reported that it assisted in determining the course of the treatment plan and surgical intervention timing (162).

4.11.2 Outcome change

Clinical, safety or patient reported outcome measures were present in approximately two-thirds of the literature. Across all disease areas, clinical outcomes were reported the most frequently (40/47), safety outcomes in 12/47 studies, and only one study described a patient reported outcome (139). Chen et al. (139) reported patient QoL measured by SF-36 scale as a secondary outcome measure for patients with Acute Left Heart Failure that received either a standard echocardiography or a CNN-echo.

Examples of non-clinical outcomes reported in studies are model evaluation metrics (precision, sensitivity, specificity with no exploration on clinical outcomes) measuring how well ML algorithms enhance image quality, against indexes such as structural similarity index measure and figure of merit index (179) or signal-to-noise ratios (127, 165) and implementation type-outcomes: commonly this was by comparing how much time it took to perform a task with and without AI.

The diversity of disease states where ML has been deployed is evident; not least by the several rounds of iterative grouping it took until they were consolidated to twenty-four areas. Due to this diversity, clinical outcomes could not be explored across the literature in its entirety. Instead, the outcomes of the three most common disease areas in the literature: cancer, stroke and respiratory disease, were all explored; all of which had a broad range of clinical outcome measures as described in Figure 3 (A, B and C).

Outcomes in cancer studies

The seventeen cancer studies were a mix of solid tumour cancers: bladder (143), breast (155, 159), colorectal (115, 131, 168, 170, 171, 179, 180), gastric (154), prostate (156), lung (132), skin (133) and thyroid (126). Hong et al. did not look at a specific cancer group but rather looked at ML systems trained to predict high risk cancer patients for acute care episodes (134), whilst Wong et al. evaluated ML for auto-segmentation for radiotherapy in multiple solid tumour groups (181).

The bladder, colorectal, gastric, lung, thyroid and skin cancer studies were leveraging ML at the screening stage. The advantages of detecting cancer at an early stage are the drivers behind screening efforts; and the literature reflected clinical outcomes appropriate to demonstrating the value of ML systems in reducing the suspicious lesion miss-rate. This was especially so in the colorectal cancer studies (n=7). Some of the studies reported adenoma miss rates being reduced by AI compared to high-definition white light (HDWL) or standard colonoscopy, but there are incidences of non-inferiority and benefits of AI with only certain types of adenomas and polyps, as described in Table 24. All five of these studies were interventional in design.

Whilst the bladder, lung, thyroid and skin cancer studies were singular, they all demonstrated some encouraging results. This may be a reflection of AI-enabled screening being a relatively mature use case of AI implementation and being on the cusp of demonstrating non-inferiority to human alone screening.

There were three studies that explored ML in the treatment planning realm of cancer care. The prostate study leveraged a deep learning auto-segmentation algorithm for both OaR delineation and target volumes tasks in short-course radiation therapy, one breast cancer study similarly looked at ML-assisted OaR delineation for radiotherapy planning (155), and the other breast study compared two ML systems that predicted radiation dose distribution to then generate treatment plans for breast radiotherapy (159).

Outcomes in stroke studies

For the 11 stroke studies, diagnostic accuracy study endpoints were again common, but half of the studies (n=6) described disease area relevant clinical outcomes such as 30-day mortality, the number of thrombectomies or the number of IV thrombolytics given.

For four of the studies, clinical outcomes were the study primary endpoints:

- Chien et al. (174) demonstrated a statistically significant reduction in length of stay for DeepCT diagnosed intracranial haemorrhage (560.67 ± 604.93 min with DeepCT vs. 780.83 ± 710.27 min without DeepCT; p=0.0232). Martinez-Gutierrez et al. (114) observed no change.
- Thrombolysis rates increased with the implementation of e-Stroke Suit (11% to 18%)) as well as thrombectomy rates (2.8% to 4.8%) (120)
- NIH Stroke Scale (NIHSS) scores were superior following thrombectomy with Deep CNN that processed CT perfusion images compared to those who underwent thrombectomy without AI (14%, P<0.05) (135)
- 30-day mortality decreased in the Aidoc-AI group compared to pre AI group (pre-AI 27.7% vs post-AI 17.5%, odds ratio=0.48, p=0.004) (182)

Whilst not directly linked to patient outcomes, three studies reported shortening various timeframes in the identification and treatment of stroke:

- CT angiogram to treatment team notification time reduced when Viz.Ai was deployed (7 min vs 26 min; p<0.001) (183)
- Door to Arterial puncture or Door to Groin times were faster with Viz.Ai: (141 vs 185 min; p=0.027) (183) with Martinez-Gutierrez et al. reporting an 11.2 minute time saving with Viz.Ai (114) and e-Stroke Suite (42 vs 44 mins) (120),
- CTA to Arterial puncture time was faster with Viz.Ai (101 vs 164 min; p=0.009) (183) and e-Stroke Suite (145 vs 174mins)(120).

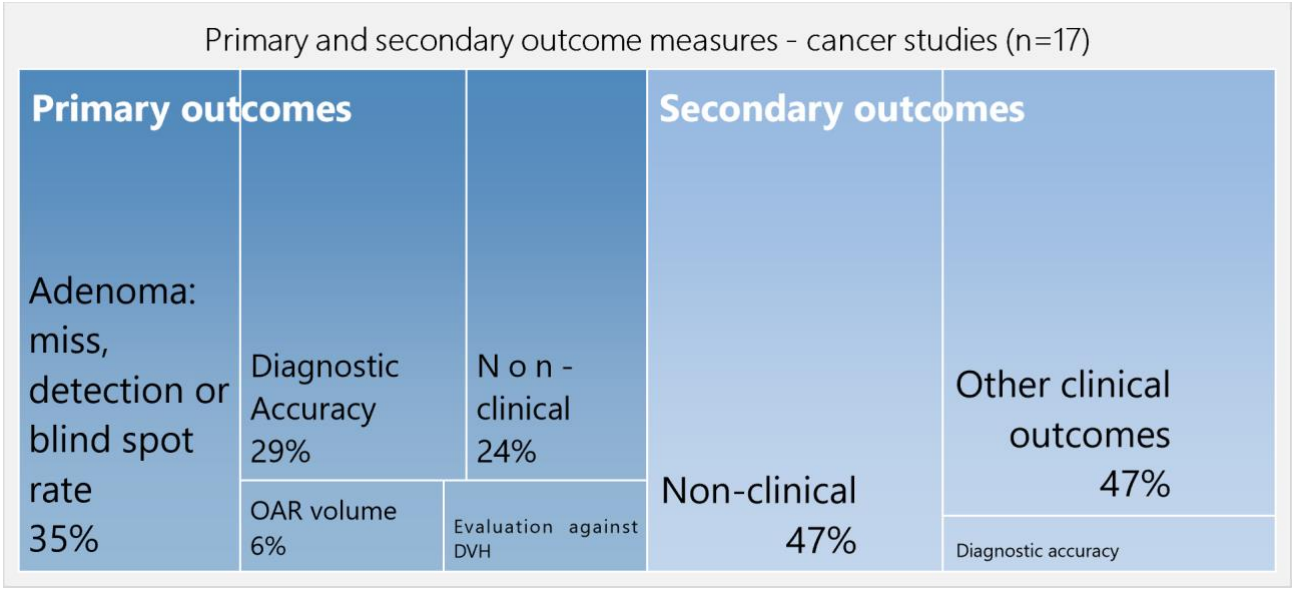
Outcomes in respiratory disease studies

Eight studies were clustered together as respiratory related studies, and were of a diverse range of disease areas: pulmonary nodules (117), pneumonia (130), acute respiratory distress syndrome (136), chest x-rays (any indication) (124, 128) , point of care lung ultrasound (any indication)(43), extubation assessment (137) and pulmonary embolism (138). This group of studies had largely non-clinical primary and secondary outcomes around user experience, image quality, time differences, implementation outcomes and service outcomes. It was not possible to consolidate the clinical outcomes from these studies.

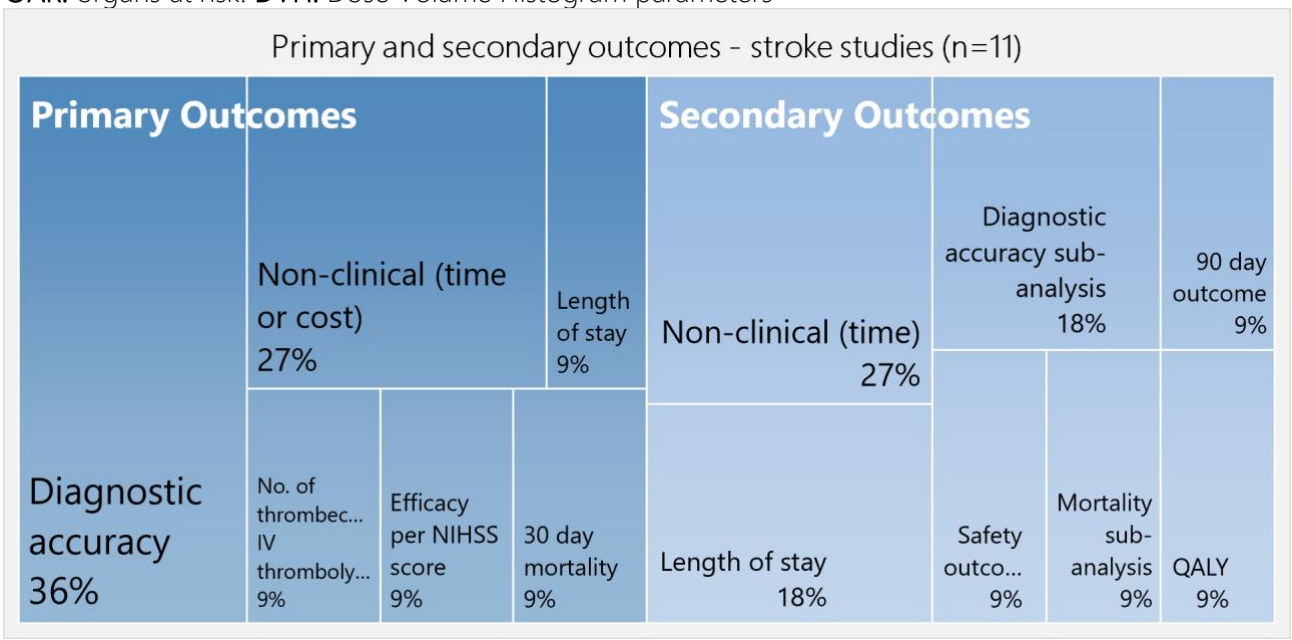
Table 24: A comparison of five colorectal screening studies.

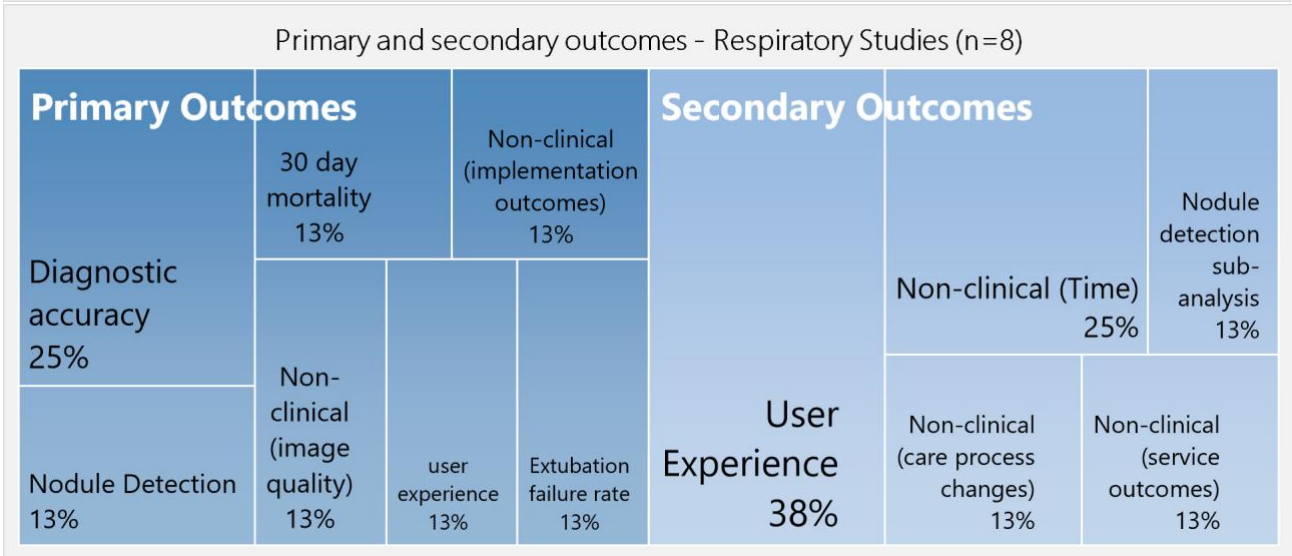
Author, year	Design	Outcome
Glissen-Brown et al. (170) 2022	Consented study subjects were randomised 1:1 either a "standard colonoscopy-first group" or "CADe first group" to undergo back-to-back tandem procedure. 116 participants per arm.	Adenoma miss-rate (AMR) (175): AI group 20.1% vs 31.2% - histology confirmed adenoma's. Polyp miss-rate (PMR): AI 20.7% vs 33.7%. Sessile serrated lesion miss-rate: AI 7.14% vs 42.11%. False positive and false negative rates: 107FP during CADe colonoscopy in the CADe-first group and 96FP during CADe colonoscopy in the HDWL-first group (p=0.2). There were 3FN's in the CADe-first group, defined as polyps detected by the endoscopist that were not recognised by the CADe system.
Kamba et al. (171) 2021	Consented study subjects were randomised 1:1 either a "standard colonoscopy-first group" or "CADe first group" to undergo back-to-back tandem procedure. 176 participants per arm.	The AMR of CADe-assisted colonoscopy was significantly lower than that of standard colonoscopy (13.8% vs 36.7% P<0.0001). The PMR, including non-neoplastic polyps, was also significantly lower in CADe-assisted colonoscopy than in standard colonoscopy (14.2% vs. 40.6%, p<0.0001).
Alessandro et al. (131) 2022	Prior to the procedure, subjects were randomised 1:1 between colonoscopy with or without CADe. Randomisation was stratified by gender, age and personal history of adenomas. 330 participants per arm.	Adenoma Detection Rate (184): Compared with the standard colonoscopy, CADe was associated with a difference in proportion of detected adenomas of 8.8% (95% CI: 2% to 17.9%). This means that ADR in the CADe group was non- inferior to the control group. False positive rate: Overall, 430/660 (65.2%) patients had polyp resections. Of these, 79/430 (18.4%) did not have histologically proven adenomas, SSLs or CRCs. These non- neoplastic polyp rates, representing 'unnecessary' polypectomies, were 12.1% and 11.8% in CADe and control group, respectively.
Quan et al. (180) 2022	300 patients at two centres underwent colonoscopy with CAD system. Their results were compared to 300 historical controls performed by the same endoscopists 12 months prior to the CAD system being piloted.	Mean number of adenomas per colonoscopy: Use of real-time CAD trended towards increased adenoma detection (1.35 vs 1.07, p=0.099) per colonoscopy though this did not achieve statistical significance. Secondary outcomes: Compared to historical controls, use of CAD demonstrated a trend towards increased identification of serrated polyps (0.15 vs 0.07) and all neoplastic (adenomatous and serrated) polyps (1.50 vs 1.14) per procedure. There were significantly more non-neoplastic polyps detected with CAD (1.08 vs 0.57, p<0.0001).
Xu et al. (168) 2021	Eligible patients were randomly assigned to conventional colonoscopy (control group) or AI-assisted colonoscopy (AI group). 1175 participants in control group, 1177 in AI group.	Polyp Detection Rate (PDR): No statistically significant difference between polyp detection rate in either group. Non- first polyps per colonoscopy (PPC- Plus): was significantly higher (0.5 vs. 0.4, p<0.05), meaning AI- assisted colonoscopy detected more diminutive polyps (easy to miss) and flat polyps than conventional colonoscopy.

Figure 3: Primary and secondary outcome measures from cancer, stroke and respiratory studies.



OAR: organs at risk. DVH: Dose Volume Histogram parameters





4.12 Health economics research

Two HE research studies from Europe described the potential cost effectiveness of deployed ML in acute care settings – the Navoy Sepsis prediction model deployed in Sweden's Healthcare service (119) and various CE and/or FDA cleared AI for vessel occlusion detection in acute stroke in the United Kingdom (122).

The Navoy sepsis model, a CE marked SaMD validated previously in a prospective randomised control trial (publication pending at this time: ClinicalTrials.gov identifier: NCT04570618) uses physiological readings and other electronic health record data routinely collected in intensive care units (ICUs) to predict sepsis. The study aimed to quantify Navoy cost savings potential in the short- and long-term effects of sepsis by developing a health economic model based on findings from the RCT and other literature sources. They deduced that the total cost per patient in Sweden was €16436 and €16512 for the algorithm arm and current practice arms respectively – a cost saving per patient of €76. A further cost saving would come from the AI reducing ICU stay by 0.16 days, saving €1009 per ICU patient. With a 3-hour faster sepsis detection time implying a reduction in in-hospital mortality, 356 lives were estimated to be saved per year in Sweden.

The HE research by van Leeuwen et al. (122) was informed by cohort specific data from the UK Stroke registry and pooled outcomes data and cost data from five large randomised trials. The research team did not account for the costs encountered for the innovation of AI, and due to lack of published evidence, assumed base case performance of 50% missed LVO rate of commercial AI products. However, for the projected lifetime per ischemic stroke patient, the incremental costs and incremental efficacy were − $156 (− 0.23%) and + 0.0095 QALYs (+ 0.07%) respectively. Using the reference value of $25,662 per QALY, 0.0095 QALY would translate to $244. For each yearly cohort of patients in the UK this translated to a total cost saving of $11 million and QALY gain of 682 ($17.5 million).

4.13 Exemplar studies

Three case studies were highlighted as exemplars of ML systems deployed into acute healthcare settings. In the first, Martinez et al. (113) described their experience of deploying a ML clinical deterioration model across 19 hospitals in the Kaiser Permanente Northern California Health Network. It was considered an exemplary case because of the approach to implementation (two centre pilot, workflow integration, early insights into clinical utility before staggered deployment across the 19 hospitals) and the post-implementation quality assurance measures put into place – local oversight and performance improvement plan for continuous evaluation. Generalisability is often a challenge with deployed ML models (12) and this case highlights successful utility of a model across 19 centres.

The second exemplar study was a multi-centre pilot study demonstrating the use of AI in a cancer pathology setting (143). It was considered exemplary because of its robust research methodology, attempts at measuring workflow impacts including the time for pathologists to read slides with or without AI assistance and collecting pathologist feedback, inclusive of the System Usability Scale.

The final exemplar study was a multi-centre randomised stepped wedge study using Viz.AI - trained to detect large vessel occlusion acute ischaemic stroke. This study exemplifies pragmatic study design and use of an FDA approved ML system. As the most common healthcare specialty utilising ML according to the literature search, this study provides useful insights into a commercially owned solution integrating into existing radiologist workflow and IT infrastructure.

Case study 1: Deployment of a ML clinical deterioration model across 19 hospitals (113).

Problem statement: Acute inpatient deterioration requires up-transfer to ICU. Failure to identify, communicate, and provide interventions for early clinical indicators of deterioration can lead to delays in care, adverse events, unplanned ICU admissions and unexpected death.

Deployed solution: Advance Alert Monitor AAM to give clinicians 12 hours of lead time before clinical deterioration. Features of this solution: 3 components

i. *The EHR embedded predictive model:* low variance due to the ML statistical modelling capacity to assimilate a large set of predictor variables, and calibrated to a clinically sustainable alert frequency.
ii. *Monitored virtual care program:* a nurse consultant (VQNC) performs expert clinical assessments on the AAM identified high risk patients.
iii. *Multidisciplinary bedside care*: The VQNC triages to the Rapid Response Team nurse, who responds to the alert and collaborates with the hospitalist, bedside nurse, and supportive care team.

Pre-Implementation: Piloted in two hospitals that addressed model performance, workflow integration, stress tested the system and gave early indication of utility of outcomes (decreased mortality, Length of Stay and improved provision of palliative care).

Deployment: staggered over 19 hospitals over 2.5 years following and implementation schedule with key milestones needing to be met before, during and after implementation.

- *Clinical governance* involvement made workflow integration more efficient because they facilitate the standardisation of infrastructure, clinical rescue and palliative care response through managed quality assurance and training.
- *Local site champions and leaders* were involved in the workflow, built a shared safety culture and minimised communication gaps.
- *The team leaders* performed daily 15-minute debriefs two weeks before "go live" and at least two weeks post "go live" at each facility where we deliberately engage the frontline RRT nurses and their direct supervisor to troubleshoot and reinforce best practices in real time.

Post-Implementation:

- Communication channels established to capture real-time feedback from end users.
- Strategic regional support.
- Local oversight and performance improvement plans put in place including quality tracking dashboards with weekly and monthly performance reported to support ongoing improvement and evaluation.

Clinical outcomes: An estimated five hundred deaths were prevented each year with AAM program, with a measured lower in-hospital mortality (9.8% vs 14.4%).

Comments on case:

This case study highlights the far-reaching impacts of deploying ML system in an acute care setting, most of which were quality improvement related, such as the standardisation of monitoring workflows, clinical rescue protocols and coordination of patient care and the creation of new healthcare profession roles for local and regional oversight. The authors also recognised a safety culture shift from reactive to proactive.

Multiple study designs and methods can be utilised to support wide deployment of ML systems – beyond this implementation case report, there were prior studies that included model development and validation, and piloting of the system by means of a three-cohort study comparing 30 day mortality, ICU admissions, LoS in the intervention cohort (alert led to clinical responses) vs comparison cohort (usual care, no alerts) and a historic cohort (1 year pre-implementation of AAM)with 97.7% of case matching achieved.

Case study 2: AI augmented system for histological classification of colorectal polyps (115).

Problem statement: Variable compliance with both colonoscopy and pathology guidelines when screening for colorectal cancer creates inconsistencies in care; with risk, cost, and negative patient outcome ramifications. With the shortage of pathologists continuing into the next decade, an already labour-intensive task of histopathological characterisation of polyps could exacerbate errors and delays.

Pilot solution: An internally and externally validated ResNet-18 neural network deep learning model that classifies by four different types of colorectal polyps was developed into an AI–augmented digital system for whole-slide images of colorectal polyp tissue samples. The regions of each histologic type were colour coded and explained in a legend contained in a sidebar of the screen. This sidebar also included the predicted classes of the whole-slide images as identified by the classifier and the percentage of patches attributed to each class to aid pathologists through quantification, instead of having them rely on visual estimations.

Method of testing workflow integration: A randomised crossover study was initiated to compare the AI-augmented digital system with standard practice of microscopic examination. 100 slides with colorectal polyp samples were read by 15 pathologists in simulated routine clinical practice setting, eight using a microscope first, seven using AI-augmented digital system first, with a washout period of at least 12 weeks before crossing over to the alternate tool. After the digital session, pathologists completed a survey to provide feedback on the digital system using the System Usability Scale (185), Paas mental-effort scale (186) and by providing written comments. The two primary outcomes were accuracy and time taken for evaluation when a pathologist used a standard practice microscope compared with when a pathologist use the AI-augmented system. A three-member board of pathologists provided the gold standard classification of the 100 slides.

Results:
Accuracy: Among the 15 pathologists, accuracy was better with the digital system (80.8%; 95% CI, 78.8%-82.8%) compared with conventional assessment with the microscope (73.9%; 95% CI, 71.7%-76.2%). Accuracy was most improved for identification of a tubulovillous or villous adenoma, for which the digital system improved reading by 21.3% (95% CI, 15.3%-27.3%). The deep learning model without a pathologist user achieved an accuracy of 87.0% (95% CI, 82.2%-91.7%) overall.

Time: The mean time of evaluation for all pathologists was longer when the digital system was used (mean, 21.7 seconds; 95% CI, 20.8-22.7 seconds) compared to microscope (mean, 13.0 seconds;95% CI, 12.4-13.5 seconds) (difference: –8.8 seconds; 95% CI, –9.8 to –7.7 seconds).

User experience: The mean score for the System Usability Scale for the digital system was 68.2 (95% CI, 61.3-75.0), which translates to a good usability. Seven of the 15 pathologists stated that they would use a version of this tool to evaluate slides routinely, 4 pathologists stated that they would possibly use a version of this tool to evaluate slides routinely. The mean Paas mental-effort rating, which ranges from "very, very low mental effort" (70) to "very, very, high mental effort" (9), was 5 (1.3), corresponding to "neither low nor high mental effort" (73).

Comments on case:
This study was an insightful pilot as the research team measured accuracy, workflow integration and user experience prior to the full deployment of the AI digital system. An interesting question is the balance between better accuracy for a longer evaluation time. The 8 second longer read time compared to the manual microscope method was investigated by the authors. They noted read times became faster as the pathologist assimilated the 100 slides, with the last set of 20 slides having a mean read-time difference of 4.8 seconds. They speculated this could be addressed by better training, although given the mid-range Paas mental-effort rating and the mid-positive usability score, there may be other causes to explore.

Case study 3: Automated large vessel occlusion detection software and thrombectomy treatment times (114).

Problem statement: Prompt endovascular thrombectomy (EVT) can drastically improve outcomes for patients with large vessel occlusion (99) acute ischaemic stroke (LVO AIS). Subsequently, accelerating the time from hospital arrival to initiation of EVT has become a cornerstone metric of stroke centre certification. Multiple challenges that contribute to the time delay have been identified and include the challenge for clinicians and radiologists to promptly recognise LVO AIS among the many patients they see, and consequently care coordination to execute emergent EVT.

Solution: A cloud-based AI-algorithm (Viz.AI) trained to detect LVO AIS. Non-contrast CT and CT angiography acquisition protocols were modified during the trial period to allow for possible AIS to be automatically transmitted to the cloud for analysis. Viz.AI analyses CT images and arrives at a decision on the presence or absence of LVO within several minutes of receiving images. The Viz.AI generated decision is then transmitted to a mobile phone application, which the clinical care team were required to download on to their phones and arrived in the form of a pushed alert notification. Within the application, a mobile picture archiving and communication system (PACS) allowed users to verify imaging findings and a secure messaging platform allowed for communication by the entire care team.

Multi-centre RCT design: A randomised, stepped wedge clinical study approach saw each of the four comprehensive stroke centre (CSC) hospitals initiating Viz.AI in pre-determined stepped-time intervals. The researchers hypothesised that initiation of this intervention would lead to a decrease in Door To Groin (D2G) time in patients with LVO AIS.

Workflow integration steps combined with trial 'ready to initiate site' steps: This included cross collaboration with the research team, commercial developer (vendor) of Viz.AI and clinicians: radiologists, radiology technicians and stroke care team members. After IT security review was completed, CT acquisition protocols required modifications to permit transmission to the cloud. Prior to site activation, the vendor would run two-day education and training for physicians and staff, inclusive of downloading apps and login, troubleshooting and running test CT images.

Results: Analysis of 131 patients in the pre-AI period, 9 in the transition period and 103 in the post AI period all who underwent EVT for LVO AIS. D2G time was reduced by 11.2 minutes in the post AI cohort. Time from arrival to IV tPA bolus did not change between the cohorts. Time from CT to start of EVT was reduced (9.8 mins). Length of Stay did not change, neither did the safety outcomes other than mortality, which decreased post-AI use, by 60%.

Comments on case:

This research exemplifies the real-world study of an acute care ML system: in a multi-centre setting, in one of the most common disease areas (stroke), and involving Radiologists, the healthcare professional most exposed to ML systems. It describes the tangible scenario of simultaneously implementing 'new technology' and generating evidence off the back of it. Indeed, this was one of several reasons that were understood to explain the choice of this pragmatic study design:

1. Viz.AI already has FDA clearance for LVO AIS,
2. The potential negative impact on patient care if there were any doubts amongst clinicians as to whether the mobile phone application was permitted for use if randomisation was to be on a per patient based, and
3. That it addresses both the needs of efficient roll out of a ML system across a large health system and the need for robust clinical trial outcomes data.

Despite FDA approval, there was an existing equipoise to warrant this research because implementing an intervention that consensus has determined to be 'beneficial' has other uncertainties to investigate such

as the degree of the effect Viz.AI has at scale. The research team had to further balance the need for time to reach an adequately powered sample size but limit the opportunity for behaviour changes to unduly influence the primary outcome measure (in this case physicians that rotate through the various 4 stroke centres).

4.14 Chapter summary

- The aim of this literature review was to synthesise a contemporaneous summary of ML systems that have been deployed into the acute care hospital setting. It yielded 75 studies that described 76 deployed ML systems.

- All the AI systems identified in the literature search were based on traditional ML techniques. Up until December 2023, no studies had evaluated the implementation of AI in hospital operations or the clinical use of foundation models or generative AI in routine patient care.

- Despite research arising from 20 different countries, the USA and China generated most of the evidence found (n=42).

- 13/76 studies were interventional (as in randomised) and 2/76 were health economic research, meaning that observational studies accounted for 80% of the study designs found in the literature.

- There was a precedence for single site observational study designs, however 36% of all studies (interventional or observational) were multi-centre.

- The health authority approval status or CE mark of the specific ML system was mentioned in several studies; however, with 78% of the literature not stating ML system approval, CE mark or exemption reasons, this regulatory aspect went largely unmentioned in the literature.

- The AI developer i.e. commercial, in-house (hospital), academia or collaboration, was able to be elucidated from much of the literature; with commercial developer being the most common (n=29).

- Radiologists were the most frequent healthcare professional engaging with ML systems (n=25). Principally this was ML systems deployed to aid analysis of CT images that traversed multiple disease areas and a variety of diagnostic tasks; such as diagnosis of stroke or pulmonary embolism. In some instances for stroke, radiologists would leverage ML visual analysis combined with natural language processing of reports to crosscheck the visual findings.

- Oncology-solid tumours (n=17), stroke (n=11) and respiratory (n=8) were the three most common disease areas in which ML was deployed and reported on; although a further 19 disease areas with four or fewer studies indicate ML systems are being deployed broadly.

- Oncology studies accounted for six of the total 13 studies that were of interventional design. Of those six interventional studies, five were a type of randomised control study design in either colorectal cancer screening or gastric cancer screening. As such, these were colonoscopy/endoscopy procedures that were either human alone or human-AI assisted, with the intended effect of reducing the miss rate of polyps or adenomas.

- Stroke was the second most common disease area in the literature (n=11) with ML systems deployed almost exclusively for diagnosis of stroke. All 11 studies used ML image analysis of head CTs, and

seven of the ML systems were classified as having a high level of autonomy (autonomous decision) because analysis of the CTs was automatically triggered upon image acquisition, with any suspected cases automatically prioritised for clinician review either via a securing messaging platform or the existing image database systems such as PACS and RIS. Of these stroke studies, eight described the ML system as either with a CE mark and/or FDA approved, with the remaining three having no explicit statement.

- Nine of the 11 stroke studies were of observational study design, the remaining two were classified as Health Economics Research and interventional. The interventional study was a multi-centre stepped-wedge randomisation where all sites eventually received the intervention.

- Respiratory related conditions were the third most common area (n=8) and consisted of a variety of indications: pulmonary nodule detection during Emergency Department CT scans, pneumonia, ARDS, chest x-ray, point of care lung ultrasound, extubation assessment and pulmonary embolism. Three of the systems supported clinical decisions, with the remaining being still image analysis for diagnosis tasks. The health data modalities were diverse and included EHR data inclusive of pathology reports, radiology reports, progress notes and vital signs. Seven out of eight studies were observational, with one study that again implemented a stepped-wedge cluster-controlled trial.

- Diagnosis tasks were the most frequent clinical task in which AI was deployed (n=40) irrespective of disease area, and predominately diagnosis via imaging modalities analysis (n=31). Nine other studies leveraged electronic health data for diagnosis.

- Two studies yielded insights into ML systems in the field of pathology. Both studies were cancer related (classification of colorectal polyps and atypical urothelial cells). Both studies employed a wash out strategy lasting 4 or 12 weeks, with pathologist reviewing the slides with or without AI assistance, and both studies had ground truth established via expert panel consensus. Case Study 2 examines one of these studies in greater depth.

- Procedure tasks were the next most frequent clinical task (n=14) and there were several examples of ML image registration (i.e. moving images and stills).

- Over two thirds of the literature described ML systems that had limited autonomy (assistive or autonomous information). Eighteen articles described ML systems with autonomous decision-making, a higher level of autonomy. 50% of these were in areas that had a time critical decision-making aspect: stroke (n=8) and acute myocardial infarction (n=1).

- When it came to describing clinical workflow aspects, 41 studies described integration with existing IT infrastructure, 24 studies gave some description of training the algorithm on localised datasets and 23 studies described some form of end user training ahead of deployment. However, only 11 studies described end user engagement during the development or deployment of the ML system, no studies described patient engagement and no studies described engaging a hospital ethics committee or clinical governance board from a workflow integration strategy or responsible use perspective.

- Usability outcomes such as user metrics and user experience data, help to characterise interaction – the first stage of the information value chain analysis of deployed healthcare information systems such as ML systems. Twenty-three studies yielded insights into interaction, with 19 detailing user experience measures and 13 via user adoption metrics.

- Decision change outcomes such as incorrect/correct decisions and decision velocity, help to characterise the effects the ML system has on clinical decision-making, the third stage of the value chain analysis. More than half of all the literature reviewed (62%) described either of these two outcomes of decision change, largely through analysis of false-positive and false-negative rates, which led researchers to detect possible automation bias in one study. Confidence, acceptability and trust were found both in usability studies and a study that tried to elucidate decision-making when physicians engaged with an in-hospital deterioration CDS.

- Care process changes associated with the implemented ML system were not well described in the literature, with 28% of studies commenting or measuring a change in the process of care (such as an increase in anticoagulant therapy prescribing after the deployment of an AI-enabled CDS for VTE risk).

- When comparing clinical, safety and patient reported outcomes in the literature (n=47), clinical outcomes were more commonly reported (n=40) as either primary, secondary or exploratory outcomes. Only one study described a patient reported outcome (QoL as measured by SF-36 scale).

- The diversity of disease states in which ML was found to be deployed in is evident; not least by the several rounds of iterative grouping it took until they were consolidated to twenty-four areas. Because of this diversity, clinical outcomes could not be explored across the literature in its entirety. Instead, the clinical outcomes of the three most common disease areas in the literature Cancer, Stroke and respiratory were all explored.

- ML systems in cancer screening tasks, notably colorectal cancer screening, showed largely non-inferior clinical outcomes by randomised control trials. Stroke studies reported largely positive impacts on both clinical outcomes and reported shortening various timeframes in the identification and treatment of stroke.

- Both these disease areas are somewhat mature in their ML automation journey and could be why they had reported clinical outcomes more consistently than other disease areas, which still included non-clinical study endpoints such as model evaluation metrics or image quality enhancement indices.

- HE research conducted in two European countries for AI enabled sepsis prediction and AI assisted stroke detection both showed cost savings, lives saved and QALY gain.

The literature supports the notion that AI systems deployed in real-world contexts are making headway in the areas of diagnostic accuracy, as a second reader and increasing the speed of execution of clinical tasks, particularly of diagnosis tasks. Evidence of AI systems effecting positive change in clinical outcomes is only just emerging in disease areas which are mature in their AI systems use journey. This next leap hinges on AI systems that provide clear explanation for their outputs to clinicians, are robustly validated, backed by well-designed clinical trials and outcome studies, integrated seamlessly into clinical workflows and used by a workforce that understands how to effectively utilise AI systems in their practice.

5. Safety of AI in acute care

5.1 Introduction

As AI becomes integrated into clinical practice, it is crucial to ensure AI systems are safe and deliver expected benefits. Like the previous generation of digital health systems (187, 188), AI comes with unintended effects that have the potential to disrupt care delivery or risk patient safety (189, 190). When AI is poorly designed, implemented or used, it can lead to patient harm and death. Therefore, it is essential to address these potential risks and ensure proper design, implementation, and use of AI systems in healthcare settings. This chapter identifies and maps emerging safety problems associated with AI in healthcare by reviewing primary studies published in the peer-reviewed literature.

5.2 Method

5.2.1 Study identification and selection

We focused on studies reporting problems with AI and their effects on care delivery or patient outcomes. Studies from Chapter 4 were supplemented with hand searches and cited reference searches using a forward-backward snowballing approach. To be included studies needed to report one or more problems with AI systems or their use and their effects on care delivery or patient outcomes. Only English language studies published in the peer-reviewed literature up to March 2024 were included. Each study was assessed independently by two reviewers against the inclusion criteria. All disagreements were resolved by consensus. While 35 of the 76 studies reviewed in Chapter 4 reported issues with algorithm performance as well as a variety of patient safety-related outcomes, only three of these specifically examined problems with AI and included reporting about effects on care delivery or adverse events (116, 118, 170). After assessment, nine studies remained.

5.2.2 Data extraction and categorisation

For each included study, we extracted information about the authors, year of publication, study period, setting, design, AI, problems, and their effects on care delivery and patient outcomes.

Types of AI safety problems: Information extracted from each study was used to develop an inventory of AI problems. Each identified problem was then labelled based on a simple model of interaction between user and AI (Figure 4) (191). Here the AI system supports a healthcare task the user seeks to accomplish, with the interaction delineated by inputs and outputs.

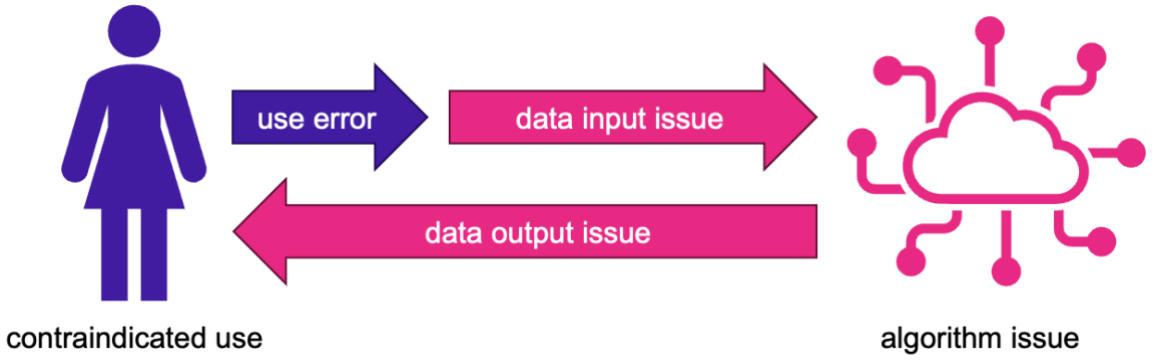


Figure 4: Types of safety problems with AI implemented in healthcare settings (after (191)).

Accordingly, the problems with AI were assigned to the following types (191):

1. Use error: errors in the use of AI systems e.g. a patient was overdosed when data was incorrectly entered into an AI system for radiotherapy planning.
2. Contraindicated use: AI system was not used as intended e.g. users were unaware that a radiological device for triage and notification of intracranial large vessel occlusion did not provide diagnostic information or remove any cases from the imaging clinician's reading queue (192).
3. Data input issue: problems with data acquisition by AI including failure to capture data (no data) or erroneous data e.g. portions of images provided to an AI system were cut-off or contained artifacts.
4. Algorithm issue: arising from the processing and conversion of input data into outputs e.g. an AI system inaccurately calculated heart rate from ECG data.
5. Data output issue: problem with the output provided by AI e.g. an AI system for x-ray interpretation froze while viewing images and stopped responding to user input.

Consequences of AI safety problems: The observable impact of AI problems on care delivery and outcomes was examined using a standard approach and categorised (188), into:

a. *Potential or actual harm to a patient*: An AI problem led to a clinical error that reached the patient, e.g., overdose of radiation or irradiating outside the treatment target when delivering radiotherapy.
b. *Near miss event:* An AI problem led to a clinical error but was detected before reaching the patient, e.g. user recognised and did not administer the insulin dose calculated by an AI tool.
c. *Noticeable consequence but no patient harm:* An AI problem that affected care delivery but involved no harm to a patient, such as delays and rework e.g. scans had to be rescheduled due to non-functional equipment.
d. *No noticeable consequence:* A problem that did not directly affect the delivery of care e.g. an electronic backup copy of patient records was corrupted, but this was detected and the copy was not needed.
e. *Hazardous event or circumstance:* A problem that could potentially lead to an adverse event or a near miss e.g. an AI system provided inaccurate measurements or results that could lead to misdiagnosis.
f. *Complaint:* An expression of user dissatisfaction e.g. a user found that training to use a new AI system was inadequate.

A narrative synthesis then integrated findings into descriptive summaries.

5.3 Results

Our search identified nine studies reporting safety problems associated with AI in healthcare (Appendix G). Four of these were descriptive studies, focusing on the performance of various ML models (191, 193-195). Only five prospectively investigated issues with AI systems during their implementation and use (61, 118, 170, 196).

The studies reporting safety problems with AI were conducted in a wide variety of settings, including various health systems, dermatology clinics, and diabetic retinopathy clinics. All but one examined AI systems in US healthcare settings. Noteworthy among these is an analysis of 266 safety events reported to the US Food and Drug Administration (FDA), the world's largest medical device regulator, as part of routine post-market surveillance of medical devices incorporating ML (191).

Another useful exemplar is the pilot implementation of an ambient AI scribe involving more than 9000 doctors across the Kaiser Permanente integrated healthcare delivery system in Northern California. This study is among the first to document real-world instances of hallucination stemming from clinical use of emerging generative AI technology (61).

In the following sections, we provide a summary of the different types of safety problems with AI as well as their effects on care delivery and patient outcomes.

5.4 Algorithm issues

Six studies reported algorithm issues arising from the processing and conversion of input data into outputs involving a variety of ML models in different healthcare settings.

Of these, two studies documented the effects of *distributional shift* which is seen to be a major issue with the clinical use of ML algorithms (194, 195). Distributional shift arises from a mismatch between the data set the AI is trained on and the data on which it is deployed. Wong and colleagues found the Epic Sepsis model which was widely used in the USA for predicting the onset of sepsis from electronic health record data performed substantially worse in the real-world (AUC, 0.63) than claimed by the manufacturer (AUC, 0.73–0.83) (195). The model identified only 7% of 2552 patients with sepsis who were not treated with antibiotics in a timely fashion and failed to identify 1709 patients with sepsis that the hospital did identify. A follow-up study of model performance across 24 hospitals using the system found sepsis alerts more than doubled in the weeks following the first COVID-19 hospitalisations(194). Changes in patients' demographic characteristics associated with the COVID pandemic i.e. presence of the virus made it difficult for the algorithm to differentiate bacterial sepsis from COVID, thereby limiting the usefulness of alerts.

Another study that investigated the real-world applicability of ML algorithms found evidence of *bias* in three state-of-the-art models designed to triage skin diseases and identify malignancies (193). Testing with the Diverse Dermatology Images dataset, a publicly available image dataset with diverse skin tones revealed a notable disparity: the models did not perform as effectively on individuals with darker skin tones and for detecting uncommon diseases.

Algorithm issues were also reported from an RCT of deep learning models for detection of lesions in colonoscopy involving 223 patients across four academic medical centres (170). The study reported 203 false positives and three false negatives i.e. polyps detected by the endoscopist that were not recognised by the AI. No immediate adverse events were reported.

Of the 227 AI safety events reported to the US FDA, 25 involved a wide variety of algorithm issues (191). These included devices with inaccurate fractional flow reserve derived from CT (FFRCT) values; problems with image enhancement; inaccurate measurements of bladder urine volume, and problems with radiotherapy treatment plans; being unable to classify cardiac rhythms or incorrect measuring of heart rate from ECG; inaccurate measures of cardiac index or cardiac output calculated by patient monitors; calculations of higher than expected insulin doses, and incorrect prediction of ovulation by a contraceptive app.

Algorithmic issues were also identified in the sole study investigating the clinical application of generative AI (61). The study documented instances of hallucination, wherein the AI provided false information without a sound basis. This included misinterpreting clinician statements such as scheduling a prostate examination, and incorrectly summarising that it had already been completed, incorrectly diagnosing conditions based on clinician mentions (e.g., diagnosing hand, foot, and mouth disease instead of simply

noting issues with the patient's hands, feet, and mouth), omitting crucial details from the summary (such as assessments for chest pain and anxiety), and generating summaries inconsistent with established note templates, thereby resulting in discrepancies in the summarisation process.

5.5 Data input issues

Three studies reported data input issues with data acquisition by AI including failure to capture data (no data) or erroneous data. In the first study, issues with the electronic health record and supporting IT infrastructure disrupted the use of two ML algorithms. These algorithms were intended to identify epilepsy patients for surgery and screen emergency department patients for clinical trial eligibility in a large children's hospital (116).

The second study examined sociotechnical considerations for use of a deep learning model for diabetic retinopathy screening at 11 clinics across Thailand (196). Out of 1838 fundus images that were entered into the system 393 (21%) were poor quality and did not meet the system's high standards for grading. Ungradable images had to be re-taken, frustrating both nurses and patients.

Data input issues accounted for 82% (n=219) of the AI safety events reported to the US FDA (191). Of these, the vast majority were failures to capture data due to various mechanical and electrical problems, including broken device components, electrical arcing, overheating, burning, or shocks. Other device failures included failure to power on, scans terminating mid-procedure, devices freezing during operation, error messages, or other failures preventing use. The remaining events involved errors in data capture, including the presence of artifacts in images, portions of images being cut off, as well as known lesions or administered contrast barely visible in scans.

5.6 Data output issues

Only safety events reported to the US FDA documented problems with the output provided by AI (191). This related to a tool designed to aid radiologists in MRI interpretation for diagnosing breast cancer, which would freeze while viewing images and stopped responding to user input.

5.7 Contraindicated use and use errors

Only two studies examined safety problems associated with the use of AI systems (118, 191). Although use problems accounted for only 7% of events reported to the US FDA (n=266), they were 4 times more likely to harm than device problems (relative risk 4.2; 95% CI 2.5–7). Errors in the utilisation of AI systems were often linked to incorrect settings, issues with user calibrations, or improper patient positioning during procedures. Many of these events were associated with radiotherapy planning devices. One instance reported skin burns attributed to a clinician mistakenly adding a 'bubble' outside the tumour, while another, with no direct impact on the patient, was linked to the movement of the target area before the treatment plan was approved.

Events where AI was not used as intended i.e. contraindicated use, mainly involved consumer-facing tools (191). For example, over-the-counter ECG devices indicated a 'normal sinus rhythm' during a heart attack, a condition beyond the device's intended capabilities. Some individuals delayed seeking medical care based on these erroneous results. Exceptions involved insulin dosing software, where clinician failure to adhere to the indicated carbohydrate treatment plan resulted in the patient suffering a hypoglycemic event. Another example is AI tools for triage and notification of intracranial large vessel occlusion (LVO). The FDA published an open letter to clinicians addressing issues with using these AI tools in real-world settings (Box 3)(192).

The second study reported the clinical manifestation of automation bias, which refers to the risk of incorrect decision support systems biasing clinicians and potentially resulting in misdiagnoses (11). This study was a prospective randomised controlled trial (RCT) that found the use of AI improved the accuracy of skeletal age assessment and reduced interpretation times for radiologists across six departments (118). However, diagnostic errors increased when inaccurate AI predictions were presented to radiologists in the AI-assisted group, compared to instances where inaccurate predictions were not presented (absolute difference in skeletal age compared to the gold standard: 10.9 months [AI] vs. 9.4 months [control]; P = .06).

Box 3: Training and support during implementation to ensure the safe and effective use (192).

The importance of training clinicians about the intended use of specific AI tools is highlighted by the US FDA's letter to healthcare providers about AI for triage and notification of intracranial large vessel occlusion (LVO).

LVO is an obstruction of one of the large arteries in the brain and is a common cause of acute ischaemic strokes. Information from real-world use suggested that clinicians may not be aware of the intended use of AI tools to support prioritisation and triage.

The FDA recommended that clinicians:

1. Be aware that AI tools only flag radiological exams with suspected findings and should never be used as a replacement for informed interpretation by a radiologist.
2. Recognise that AI tools cannot rule out the presence of an LVO. If a radiological exam is not flagged by AI, an LVO may still be present.
3. Recognise that when the AI is used as intended (as a prioritisation and triage tool and not a diagnostic device), it can improve workflow by prioritising suspected cases.
4. Recognise that the device does not remove any radiological exams from the queue for interpretation by a radiologist. When used as intended, exams that are not flagged by the device are still interpreted by a radiologist according to the standard of care.
5. Be aware of the design of AI. This includes understanding the vessels (arteries) for which the AI was designed and tested to detect LVO. AI tools may not be designed and tested to evaluate all intracranial vessels.

5.8 Chapter summary

This chapter reviews recent peer-reviewed studies to uncover emerging safety concerns associated with AI systems in healthcare, examining their consequences for care delivery and patients. While there is limited documentation of AI-related adverse events in current literature on AI implementation, analyses of safety events provide valuable insights into emerging problems with AI in healthcare. Our findings reveal instances of algorithmic issues, notably stemming from the inherent limitations of ML, such as susceptibility to biases in training data and the occurrence of distributional shifts over time. These findings underscore the importance of evaluating algorithm performance in real-world healthcare settings as part of implementation and in routine use.

Additionally, our review highlights that many safety concerns are centred around data acquisition processes for algorithmic processing. Importantly, the evidence indicates that issues related to the use of AI are more likely to result in patient harm, a finding consistent with previous research on safety events involving digital health technology. Specifically, human factors issues were proportionally higher in events where patient harm occurred. This highlights the critical need for ongoing monitoring and evaluation of AI systems to mitigate potential risks to patient safety.

6. Key findings from policy review and principles for safe and responsible AI in healthcare

6.1 Introduction

This chapter provides a summary of the key findings, incorporating published legislation, policies, guidelines, and principles for AI implementation in healthcare from government agencies in the UK, US, New Zealand, Singapore and Canada. Additional documents from EU, WHO, and OECD were reviewed. This chapter presents these findings and principles in the context of evaluating and implementing AI in health services.

6.2 Governance and regulation of AI in acute care - structures, systems and principles

Key findings: Common approaches to governance and regulation of AI in acute care based on the international documents included creation of new legislation to govern AI, appeal to existing legislation (e.g., data privacy laws, consumer rights laws, and anti-discrimination laws), and appeal to ethical frameworks. Two of the nine new pieces of legislation proposed or enacted in the US, Canada, UK and EU were explicitly focused on healthcare applications of AI: the US Proposed Rule (87 FR 47824) on *Nondiscrimination in Health Programs and Activities* (24) and the *Final Rule (89 FR 1192) on Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing* (26).

In addition, agencies from international jurisdictions suggested using practical procedural tools, such as risk assessment tools and question-based checklists (see Table 6: List of procedural tools. in Chapter 2 for a summary of tools). Of the nine procedural tools found in the documents, few were explicitly designed for healthcare AI applications (30, 39, 40) . Future research could focus on validating and evaluating the usefulness of these tools, or adapting them to the Australian context.

Australian and international documents showed national ethics frameworks influence how policy is formulated. The US Department of Health and Human Services drew on a national ethics framework to develop a playbook (40) to guide health departments in embedding ethical principles in AI development, acquisition and deployment. Australia's National Ethics Framework (70) is commonly used to frame Australian policy; similarly the NSW Government AI Ethics Principles (71) are embedded in the NSW AI Assurance Framework (67), which applies to uses of AI in the NSW health system.

Australia's existing regulatory landscape provides governance and regulation of AI implementation across industries, including healthcare. Existing regulatory instruments include the Privacy Act, Consumer Law, and TGA regulation of Software as a Medical Device (SaMD). There are, however, calls and proposals for AI-specific guidance and regulation, such as the use of risk assessment tools (15, 68), development of best practice guidelines (86), and a single set of AI use standards (25). The most significant developments in the healthcare sector are policy initiatives by the Royal Australian and New Zealand College of Radiologists (3), Australian Alliance for Artificial Intelligence in Healthcare (73) and the Australian Medical Association (4). The AAAiH National Policy Roadmap Process has recommended, by consensus, that Australia establish an independent National Council to oversee AI governance in health.

Principles:

- A National Council should be established urgently. Its work should be shaped by the National AI Ethics Principles and the recommendations made by consensus in the National Policy Roadmap process.
- The Australian legislative and policy environment for AI is rapidly changing: upcoming developments include changes in cross-sectoral legislation (e.g. privacy law) and an intended national risk-based approach to AI legislation.
- Governance of healthcare in AI is well advanced in other jurisdictions: there are significant opportunities for leadership in Australia in this regard.
- Ensure AI implementation within your organisation complies with existing legislation, including data privacy, consumer law, and cybersecurity policy, among others, and the Australian National Ethics Framework 2019.

6.3 Engagement with consumers, patients and citizens

Key findings: Documents from international jurisdictions provided insights into governance approaches that rest on engagement with consumers, patients and citizens in two ways. First, public engagement was recommended at the level of policymaking to govern AI implementation across sectors, including healthcare (21, 54, 55). In *Ethics and Governance of AI for Health* (54), WHO recommended public engagement and dialogue to ensure that the use of AI for healthcare meets core societal expectations and greater trust. Exemplifying this approach, the UK's Health Data Research, which collects health data, used public engagement workshops to provide a forum for participants to discuss their expectations and concerns about the use of patient data in AI (54). A pilot policy project sponsored by New Zealand contained a six-step plan for holding national conversations to gain social licence or public trust (47). Key aims of a national conversation included involving individuals who traditionally are not included in policymaking, and building consensus on AI ethics and values that underpin AI policy and use.

Fewer documents discuss public engagement in the context of implementation of AI in acute care. Practical strategies to incorporate or institutionalise public engagement in the implementation of AI in healthcare include the following:

- Organisations should develop effective public partnership and communication strategies that involve both informing and listening, and cover issues such as limitations of AI, risks and benefits, current and potential applications, and frameworks for governance (28, 47).
- Organisations should be required to publicise intent to deploy automated decision systems, including use policy, in their websites. Upon publication of use policy, invite the public to submit comments, with comments being incorporated as part of the approval process (18).
- Organisations should coordinate with patient representative groups and other stakeholders to help develop information materials about AI systems that will be understood by patients and other stakeholders (31).

Australian organisations were in the early stages of considering patient and public engagement, with less well-developed recommendations than some international jurisdictions. Several Australian organisations made commitments to public engagement or made recommendations that AI systems should be developed and implemented in consultation with the public, including the Federal Government Department of Industry, Science and Resources (15, 72) and the NSW State Government (71). Despite committing to public engagement, exemplars of Australian activities to involve publics in AI regulation and governance were limited in the Australian documents. The NSW State Government launched a community feedback program in 'Artificial Intelligence – Have your say', which encouraged members of the public to respond to an online survey but did not mention involving the general public in any two-way consultation processes (197).

In the Australian healthcare context, both AAAiH and AMA recommended public engagement (or co-design) in the development and regulation of AI systems (4, 81) but neither organisation provided specific guidance on how public consultation should be conducted or what it should aim to achieve. There is opportunity to develop specific guidelines for community and public engagement in the implementation and governance of AI in healthcare. As recommended by AAAiH (81), this guidance should also consider how Aboriginal and Torres Strait Islander communities should be involved in decision-making about AI in healthcare.

Principles:

- Strengthen engagement with consumers, communities, and stakeholders in healthcare AI implementation to ensure trustworthiness, and to shape implementation and use of consumer- or patient-facing AI.
- Develop communication and engagement strategies in collaboration with stakeholders and patient groups to keep public and service users involved in implementation of AI systems in their health services.
- Implementation of AI in health systems should ensure appropriate Aboriginal and Torres Strait Islander governance, by connecting AI governance processes in health systems to existing Aboriginal and Torres Strait Islander governance structures. Implementation should be in line with principles of Indigenous Data Sovereignty.

6.4 Equity, discrimination and human/patient rights

Key findings: The international documents provided evidence of international government agencies paying serious attention to the risk of AI bias that could lead to discriminatory practices, violation of human and patient rights, and health inequities (50, 198, 199). The risk of bias tended to motivate the principles of fairness, justice, and human dignity (55). While government agencies in most jurisdictions cited existing anti-discrimination laws, US federal and state governments proposed new legislation that explicitly banned discrimination arising from AI systems (8, 16, 17). For example, the *New Jersey Senate Bill 1402* (16) states that "A healthcare provider shall not discriminate through the use of an automated decision system against any person or group of persons who is a member of a protected class."

Several international agencies provided strategies to address bias. New Zealand's Ministry of Health recommended the following steps: acknowledging bias exists, prioritising explainable and auditable algorithms, and following existing data standards (30). Agencies in other jurisdictions suggested adding equity and other societal considerations to performance metrics (42), use of institutional review boards to evaluate the risk of bias (1), inclusion of bias and discrimination assessment in safety credentials criteria (39), and development of compliance and enforcement tools to monitor bias (200).

In the Australian context, policies for the use of administrative or automated decision-making applications of AI focussed on the public's right to contest decisions made about them, including those made by (or assisted by) AI. Organisations including the AGA, the DISR and the Australian Human Rights Commission advised that AI should not prevent citizens from being able to contest decisions (72, 87), that organisations using AI should ensure that systems are designed such that they can provide adequate reasons for any decision that is made (69), and that avenues for contestability should be made easily accessible for all members of the public (69, 72). Aside from the right to contest decisions, non-clinical organisations such as AGA and the Human Technology Institute highlighted that any AI systems which directly or indirectly discriminate against groups with protected attributes are illegal under Australia's Anti-Discrimination laws (64, 87).

Australian clinical organisations made a series of recommendations about how to ensure AI-integrated care remains equitable and upholds patient rights, including the following:

- Allowing patients control over their medical records and how their data is used and disclosed (4);
- Building clinician capacity to ensure clinicians are comfortable informing patients about how AI will be used in their care (88);
- Ensuring that hospitals and clinics have information available to patients about how they are using AI (82);
- Implementing AI systems that have been trained on diverse, inclusive, and relevant data (3, 4);
- Making efforts to ensure that Aboriginal and Torres Strait Islander communities benefit from the introduction of AI (79, 81).

Whilst most organisations indicated their commitment to building and implementing AI that is equitable and upholds human rights, there is need for the development of a more refined best-practice approach. Best-practice guidance should address a range of expectations, including how hospitals should ensure that people are adequately informed when AI is used in their care, how avenues for contestability can be made available and accessible where AI is being used for administrative decisions, the extent to which it is appropriate for patients to have full control over the disclosure of their health information, how hospitals can audit and monitor systems to ensure outcomes are equitable, and how to appropriately assess whether AI systems are benefiting Aboriginal and Torres Strait Islander communities.

Principles:

- Consider the transparency and contestability of decisions involving automation, including administrative decisions such as allocation of bed days or services;
- Implement risk assessment frameworks to address the risk of bias, discrimination or unfairness, being mindful of obligations under anti-discrimination legislation and ethical requirements to prevent unfair outcomes – this requires ongoing monitoring as well as initial evaluation;
- Support clinician capacity to sufficiently inform patients about the use of AI in their care.

6.5 Privacy and confidentiality

Key findings: International documents showed that government agencies were calling for stronger privacy policy and protection, particularly in the context of healthcare data being used in AI development (58). Several agencies recommended for regulators, developers and implementers of AI to review existing data protection regulatory frameworks. Legislation and polices that cover personal data are typically grouped under "privacy policy" in the US, and under "protection policy" in the EU and elsewhere (55). In the EU and the UK, agencies refer to GDPR as one of the key regulations that should guide privacy considerations when implementing AI in healthcare. Other national legislation mentioned by the reviewed documents included the UK's Health and Social Care (National Data Guardian) Act of 2018, Singapore's Personal Data Protection Act of 2012, and New Zealand's Health Information Privacy Code of 2020.

While there were numerous data privacy laws, the OECD noted the lack of harmonised policies between data authorities and health authorities, as well as among authorities from different jurisdictions (58). Authorities are encouraged to work together to continue protecting sensitive health data while recognising the value of national, regional and global data collaboration to realise the benefits of AI. Other practical strategies to address privacy concerns include:

- Implementing privacy-enhancing technologies such as encryption and pseudonymisation of data (58).

- Developing a Code of Conduct that includes measures to maintain privacy and confidentiality in the implementation of AI in healthcare (58).
- Completing a Data Protection Impact Assessment (1, 31).
- Creating a legally binding written data processing contract (information sharing agreement) between developers and deployers/implementers (39).

Australian organisations often highlighted the Privacy Act as the piece of legislation that should govern how data is held and used for AI projects (25, 41, 70, 87, 90), with the DISR noting that ongoing work on privacy law reforms will address important privacy issues associated with AI (72). Non-clinical documents framed guidance for the development of AI around the principles in the Privacy Act. Guidance from the Commonwealth Ombudsman (41), the Human Technology Institute (64), and OVIC (66) provides specific recommendations for how Australia's existing privacy laws should be interpreted for those developing and implementing AI systems.
There is a current and ongoing review of the federal Privacy Act by the Australian Government Attorney-General's Department (AGD), in part driven by the use of AI and automated decision making. However, more work is required to fully canvas the impact of privacy legislation on the deployment and use of AI in the Australian context.

Beyond existing Australian privacy laws, the AAAiH and the MTAA advocated for a data governance framework specifically for the sharing of health-related information (81, 92). Both organisations identified a need to govern private, secure, and efficient transfer of health information between clinical contexts and AI developers for the development of AI systems. The AAAiH advised that this data sharing process should be 'consent-based' but did not provide further guidance on the practicalities of how consent should be sought from individuals (81). The MTAA (92) highlighted the Canadian Institute for Health Information's approach to data governance as an exemplar (201), advocating for the need for a similarly comprehensive health data management framework in Australia to improve efficiency and enhance public trust in data governance.

Principles:

- Ensure that AI systems that require and interact with patient data comply with existing legislation and policy on patient privacy and confidentiality.
- At a national level, detailed legal analysis of privacy requirements with respect to AI implementation in healthcare may be warranted, as this is not as well resolved in Australia as in some other jurisdictions. This could potentially support legal reform.
- At a national level, the development of a formal data governance framework may be warranted. As previously noted, this must uphold Indigenous Data Sovereignty.

6.6 Evaluation, monitoring and maintenance as an issue for governance

Key findings: Most of the international documents included governance strategies to guide the evaluation, monitoring and maintenance of AI systems deployed in healthcare. In *A Buyer's Guide to AI in Health and Care* (39), UK's NHS provided a comprehensive list of implementation, procurement, and delivery considerations designed to guide health organisations that are planning to deploy AI. Several government agencies outlined the roles and responsibilities of organisations procuring an AI solution (39, 48). The Government of Singapore proposed an AI governance framework (48) that contained a list of roles and responsibilities that can be allocated to qualified personnels within an organisation's internal governance structure. These roles include implementation of a risk management framework, maintenance and review of deployed AI, and establishment of communication channels with stakeholders.

Several international documents showed that government agencies are concerned that existing internal governance structures may need to be modified to address AI deployment in healthcare (30, 48). For example, New Zealand's Ministry of Health suggested establishing an AI-dedicated governance structure with diverse areas of expertise, including methodological (governance and data science), data structure, organisational strategy, clinical, and advocacy (30).

A key challenge is that while government agencies in international jurisdictions emphasised the need for regular and continued monitoring of AI post-deployment, there was no clear guidance on the frequency or specific intervals of monitoring or review.

Most of the Australian documents included in the review mentioned the need for evaluation, monitoring and maintenance of AI systems. Aside from the AI Assurance Framework, which necessitated ongoing self-assessment for AI projects in NSW State Government agencies, there were no mandatory monitoring processes for AI systems in Australia. Organisations recommended that AI systems be assessed at regular intervals to ensure that systems were still meeting the outcomes for which they were implemented, were still delivering community benefits, and were not degrading in performance over time (15, 70-72). The AATSE and Human Rights Commission both recommended that auditing of any government uses of AI should be done independently (25, 69).

From the AMA, ACD, and RANZCR, there were strong recommendations for independent real-world evaluations to demonstrate an AI tool's effectiveness before widespread clinical use (3, 4, 79, 82). The ACD recommend only implementing tools with performance at least equivalent to healthcare professionals (79), whereas the AMA recommended that tools demonstrate improved health outcomes for patients (4). Further, RANZCR recommended additional evaluations be conducted where AI systems are being imported from other countries to ensure the system works in the Australian context (3) and made substantial recommendations for how ongoing performance audits for clinical AI systems should be conducted (82). In addition, both RANZCR and AAAiH highlighted the need for guidance to address whether and how to implement AI that continues to learn after implementation (81, 82).

The AAAiH identified a need for better governance of AI safety in Australian healthcare and made a series of recommendations for establishing a consistent national approach to ensuring that AI is implemented safely and ethically in healthcare. These recommendations include the establishment of a risk-based safety framework that necessitates that vendors provide real-world evidence of performance, a better mechanism for post-market safety monitoring, and the development of minimum AI safety standards for healthcare organisations using AI (81).

In federal systems, such as Australia, and multinational systems, such as the EU, it is particularly important to work towards harmonising legislation, frameworks, and guidance on issues such as the development, deployment and use of AI because the technology moves and is used across borders. The EU have tackled this challenge by developing the EU AI Act, as part of its digital strategy, which will apply in all EU countries. The aim of the AI Act is to ensure that AI systems are across the EU are safe, transparent, non-discriminatory, and environmentally friendly. The approach taken in the Act is risk-based and horizontal, rather than sectoral, to ensure a consistent approach to AI across sectors.

In Australia, the Australian Government is also seeking to take a leadership role, having identified AI as a 'critical technology in the national interest', that is, a technology that can impact Australia's national interest in areas such as economic prosperity, national security, and social cohesion. The Australian

Government is also best placed to work at the international level, for example, as a member of the Global Partnership on Artificial Intelligence (GPAI).

Principles:

- Ensure high quality, local, practice-relevant evidence of AI system performance before implementation.
- Consider establishing a governance framework for AI implementation that clearly sets out tasks, roles, and responsibilities in the evaluation, monitoring and maintenance of AI systems.
- Ensure use of existing patient safety and quality systems for monitoring AI incidents and safety events (including hazards and near miss events) as well as post-market safety monitoring so that cases of AI-related patient risk and harm are rapidly detected, reported and managed.
- At a national level, consider development and implementation of a risk-based safety framework and minimum standards of practice, overseen by an independent National Council. Authorities/agencies should consider harmonising legislation, frameworks, and guidance on issues such as the development, deployment and use of AI.

6.7 Transparency

Key findings: International documents offered insights into various conceptions of transparency in the governance of AI in healthcare. One key conception was algorithmic transparency (or algorithmic explainability), which refers to a characteristic that enables users to understand the details or reasons a model made a decision.[US 11] Government departments across jurisdictions recognised this conception of transparency as a matter of safety and performance (21, 30, 31, 55). Algorithmic transparency recommendations included ensuring access of external stakeholders (e.g. regulators and the public) to technology assumptions, limitations, operating procedures, data properties, and algorithmic model development (1, 42, 54).

Another key conception was process transparency, which refers to individuals interacting with or affected by the AI decision being able to understand the implementation practices that lead to an AI-supported outcome (38). In addition to being a matter of safety, this conception was generally associated with the principles of trust and trustworthiness. To ensure process transparency, Singapore's Ministry of Health provided a list of suggested information depending on the type of end-user (49). Other process transparency considerations include commercial and institutional transparency. To ensure commercial transparency, the UK NHS suggested for agencies and organisations to make commercial contracts (or at the very least information about commercial partnerships) publicly available (31). For institutional transparency, the UK Government proposed using the Algorithmic Transparency Recording Standard (ATRS), which established a way for government departments to publish information about how and why they are using AI (44).

Table 25: Types of transparency requirements and purpose

Transparency of what	Transparency for what purpose
1. Transparency to support consent	To provide meaningful information to support informed consent, such as limitations of AI, adverse events, and alternative (non-AI) solutions.(49)
2. Transparency of using AI in patient care	To improve user and service recipient awareness and understanding of AI; and ensure organisational accountability; enable adequate regulation of safety.(44, 55)
3. Transparency in data use	To ensure compliance with privacy and data protection laws, such as EU's GDPR.(50)

Transparency of what	Transparency for what purpose
4. Transparency with respect to governance (including performance of AI system and organisational governance structures)	To build confidence and trust, ensure interoperability, enable independent audits. (44)

The Australian organisations acknowledged that full algorithmic transparency or explainability was often not possible or desirable when implementing AI systems, but that a certain degree of transparency was important to ensure that systems were auditable (15, 70). The guidance for administrative AI systems often referred again to citizens' rights to contest AI-informed decisions, with guidelines typically indicating that AI systems should be transparent enough to generate 'reasons' to justify an automated decision (41, 66, 69, 70, 86). The Human Rights Commission (69) further advised that those reasons should be understandable to a person with relevant expertise, and that organisations should seek support for how to ensure their systems generate reasons that meet those requirements.

In addition to transparency about the decision-making process, several organisations recommended that agencies are transparent about when AI is being used (66, 67, 86, 91, 94). The clinical organisations were strong advocates for transparent reporting of algorithm performance. These recommendations intended to ensure that practices are aware, when deciding whether to implement AI systems, of potential issues with transferability of systems to new healthcare settings, and of the potential for algorithmic bias when populations are underrepresented in training data. Organisations recommended transparency about an AI tool's performance, evaluation processes, and any under- or over-represented populations in the training and testing data (3, 4, 81, 82). As part of a risk-based safety framework, the AAAiH recommend mandatory transparency reporting requirements for vendors, including information about performance and training datasets (81). However, AAAiH recommendations do not address transparent reporting of reasons for administrative (or clinical) decisions to allow for contestability.

Principles:

- Governance of transparency should draw on existing expertise and governance systems in healthcare organisations, including clinical ethics committees, research ethics committees, digital health committees, consumer governance committees and risk management structures.
- Ensure transparency to consumers and clinicians about the fact that AI is being used, as this is generally seen as a baseline requirement.
- Risk-based assessment could require greater transparency for higher-risk applications.
- Clinical policies and guidelines tend to lean towards a requirement for more, rather than less, transparency than administrative applications.
- In clinical contexts, consider ensuring transparency regarding training data, including data bias, and transparency regarding AI system performance and evaluation methods – multiple organisations have suggested these should be required.

6.8 Consent considerations

Key findings: The review of international legislation and polices showed two separate consent obligations: 1) patient consent to use AI in their care and 2) patient consent for their health data being used to develop AI. In the *Ethics and Governance of AI for Health* (57), WHO noted that hospitals and healthcare providers are unlikely to inform patients about the use of AI in their care given the absence of precedence for seeking consent when using technologies for diagnosis (e.g., use of x-ray) or treatment. Despite this practice gap, some agencies clarified the legal imperatives of consent based on existing legislation and policies. The UK NHS cited articles in GDPR about consent, as well as the common law duty of confidentiality (39). The *EU AI Act* (22) mandates that individuals must be informed when they

interact with an AI system so they can make an informed decision to continue or decline the interaction, with exceptions made for cases where it is clear from the context that notification is unnecessary.

Key recommendations from the international documents included:

- Upholding existing consent requirements and practices for other medical procedures performed by physicians (49).
- Upholding meaningful consent, which requires use of understandable and plain language explanation of how AI works or how patient data might be used to develop AI, and information about the limitations of AI and the availability of alternative options (54)
- Developing mechanisms for dynamic consent or ongoing consent to adapt to the evolving nature and scope of AI usage in a patient's care (28, 37).
- Promoting AI literacy among stakeholders, such as patients and healthcare professionals to facilitate informed decision-making and ensure informed communication with patients (22, 26).

There are challenges in complying with consent requirements. WHO acknowledge that patients may not be able to anticipate and consent to all the ways their health data might be used in the future, such as population-level analysis or risk modelling (54). There is also a clear tension between protecting individual privacy and ensuring sufficient data for reliable and representative datasets. Finally, reviewed documents discussing consent considerations did not provide details about how to implement consent policies in clinical practice.
Fundamental requirements for consent in clinical contexts—that a person must have capacity, consent voluntarily and specifically, and have sufficient information about their condition, options, and material risks and benefits—remain unchanged by the use of AI. The National Safety and Quality Health Service Standards definition of informed consent is in Box 4. There is limited guidance available regarding requirements for consent for use of AI as an element in clinical care.

Box 4: Definition of informed consent.

> Informed consent: a process of communication between a patient and clinician about options for treatment, care processes or potential outcomes. This communication results in the patient's authorisation or agreement to undergo a specific intervention or participate in planned care. The communication should ensure that the patient has an understanding of the care they will receive, all the available options and the expected outcomes, including success rates and side effects for each option.
>
> From the National Safety and Quality Health Service Standards 2nd Edition (202)

Australian organisations frequently referred to the Privacy Act in making recommendations about when it is appropriate to seek consent for the use or collection of data. OVIC and the NSW Government identified that seeking consent for the collection of information was ethical and helped promote public trust (66, 71), but guidance from OVIC (66) advised that seeking consent was not always necessary or practical for organisations developing or testing AI systems. For example, OVIC highlighted the challenges in allowing individuals to revoke their consent when their data may have already been used to train an AI system (66). OVIC advised that, whilst consent may be necessary under certain circumstances, Australian privacy law contains certain exemptions that allow for the collection and use of sensitive data without consent.

Both RANZCR and the AMA recommended a more stringent approach to seeking patient consent to use health data in AI development than that outlined in Australia's Privacy Act. Whilst Australia's existing

privacy laws outline certain conditions where consent is not necessary for the collection of sensitive information, including where that information is being used to manage or monitor a health service, the AMA recommended that patients have full jurisdiction over the use and disclosure of their data (4), and RANZCR recommended that no patient data should be transferred outside the clinical environment without permission from an Ethics Committee or patient consent—unless required by law (3). Similarly, AAAiH recommend a consent-based framework for industry access to healthcare data (81). Despite the clinical organisations' commitment to seek explicit consent for the use and disclosure of patient data, the documents contained no further information on how the practicalities of consent would be managed and balanced with the practicalities of AI system development.

Principles:

- Patient consent for using AI in their care should build on existing consent requirements and practices for other medical procedures.
- There is currently disagreement across existing policy regarding consent for data use by AI systems and for AI development.
- Further work is required to establish a national consensus that would provide firm foundations for the development and deployment of AI systems in healthcare in Australia, with attention to community views, privacy legislation, and clinical and research ethics requirements.

6.9 Accountability and liability

Key findings: Accountability was one of the key principles found in the international documents included in the review. Intergovernmental organisations such as OECD and WHO recognised the evolving challenge of accountability, understood as attribution of responsibility, in healthcare AI (55, 58). WHO proposed a faultless responsibility model ("collective responsibility") in which all agents involved in the development and deployment of AI are held responsible to promote integrity and minimise harm (54). The US Government Accountability Office published an accountability framework (42) that identified key practices and audit procedures to promote accountability in governance structures and processes to manage, operate and oversee the implementation of AI systems across federal agencies and other entities.

International government agencies expressed concern about the lack of clarity in establishing legal liability, and apportioning blame when mishaps occur (21, 50). One reason provided by some government agencies was the black box problem in AI systems, where there is lack of explainability or clarity in how AI make decisions (21, 31). In addition, there is a lack of legally defined actors in the AI system lifecycle that creates considerable uncertainty regarding which party or parties might be liable if harms from AI arise (31, 203). While there is ongoing policy work to establish fair and effective allocation of liability throughout the AI life cycle (31, 203), the review noted liability interpretations based on the international documents:

- If equipment can be proven to be faulty, then the manufacturer is liable (50).
- Use of AI-MD does not change the liability of the implementing institution or the individual medical professional in their provision of appropriate and safe care (49). If the AI is used as an aide, human experts remain the liable party. If the medical staff relied solely on the AI without applying their specialist knowledge, that could be a negligent act (50).
- In areas such as health, liability "must ultimately lie with a natural or legal person" (50), or specifically "practitioners supervising AI" (28).

Other strategies to address accountability and liability concerns include implementing existing liability rules (22, 39, 50), incorporating liability questions in assessment checklists (32), and establishing clear frameworks for liability (28).The New Zealand Government released a report that included guiding principles and practices for adoption of AI in healthcare settings. One of the recommendations was for organisations to establish a framework for liability that should: i/ distinguish by application/output; ii/ distinguish by level of supervision; iii/ distinguish by level of associated risk; and iv/ establish clear criteria for insurance coverage.

Both the non-clinical and clinical Australian documents recommended that accountability for all decisions should remain with responsible individuals and organisations, even when those decisions were informed by AI. The non-clinical organisations often recommended human-in-the-loop processes for AI systems, where human intervention points were identified to ensure that responsible parties had oversight over decisions (15, 70, 72, 87). Several organisations, including the DISR and RANZCR, recommended that the person or people responsible for oversight over an AI system should be clearly identified to ensure that accountability for the system's performance is always clear (67, 71, 72, 75, 81, 82).

The clinical organisations advocated for ensuring that physicians retained authority and control over clinical decisions. The ACD and the AMA recommended that AI should only be used to augment, and not to replace, physician judgement (4, 79), and both the AMA and RANZCR recommended that the decision to use or not to use an AI tool should always rest with the physician and never with the hospital or clinic (4, 82). Only RANZCR advocated for shared responsibility for an AI systems ethical implementation, between the physician, the Practice or hospital, and the AI developer (3, 82). Whilst it was common for the organisations to place responsibility for AI-informed decisions with an identified individual, there was no standard approach to assigning accountability for AI-informed decisions across the Australian organisations.

Principles:

- AI governance should build on existing governance processes in healthcare organisations to ensure safe and responsible use of AI, as well as clarify lines of individual and organisation responsibility over AI-assisted clinical and administrative decision-making that comply with existing liability rules.
- At a national level, detailed legal analysis of liability and accountability with respect to AI implementation in healthcare may be warranted, as these are less well-resolved in Australia than in other jurisdictions. This legal analysis may support legal reform.

6.10 Worker training and support

Key findings: The review of international documents showed that agencies and organisations acknowledged the need for continuous training and education for healthcare workers to keep pace with evolving AI technologies (54, 57, 58, 198). Recommended training programs for healthcare workers should include promoting AI and digital literacy (1, 22, 50), improving skills in effective communication regarding the use of AI technologies (58), and increasing awareness about the adoption of AI technologies within the organisation (1). Some practical strategies that could be considered for implementing worker training and support initiatives related to AI in healthcare include:

- Developing tailored training modules for different healthcare roles that cover AI functionalities, limitations, potential biases, and ethical considerations (22, 54, 58).
- Collaborating with AI developers to design user-friendly AI systems to facilitate access for workers with various skills, ensuring healthcare workers are equipped to effectively communicate with patients about AI technologies (20, 39, 54).

- Implementing mechanisms for evaluating the effectiveness of training programs and gathering feedback to continuously improve training initiatives thereby enabling feedback and dialogue about the real-life effects of the AI system's use (38, 54, 55).

Some recommendations about worker training and support go beyond technical capacity building. New Zealand agencies recommended that AI deployers/implementers should understand the comfort levels of healthcare staff regarding the use of AI in healthcare delivery in New Zealand (28, 30). The UK NHS (31) emphasised the importance of supporting staff to understand the ethical considerations and regulatory procedures required for approving and monitoring AI solutions. However, the documents from NHS did not include details about training programmes that incorporate ethical and regulatory considerations.

Non-clinical Australian documents often recommended that organisations should invest in capacity-building for staff, so that all relevant employees understand how to utilise AI systems safely and ethically (25, 41, 64, 87). Guidance on public sector use of generative AI recommended that staff are trained in how to use generative AI safely and without compromising privacy or using the tools for tasks like administrative decision-making (86, 91, 93). See Chapter 5 on the link between training and patient safety (Box 3).

The clinical documents had a strong focus on clinician and healthcare worker capacity-building. The Victorian Department of Health recommended that healthcare workers be provided with guidance and training on how to safely use generative AI (80). RANZCR and ACD recommended clinician training for use of AI in clinical workflows (3, 79). The TGA recommend clinicians build capacity in understanding how software-based medical devices might compromise security, so that they can safely use the tools and communicate risks to patients. The organisations identified the need for physician training in the use of AI, but none of the clinical organisations mentioned general or discipline-specific guidelines to inform the development of training or support programs.

Principle: There is broad consensus in both international and Australian jurisdictions that significant training and support for clinicians and other health workers are required prior to the implementation of AI tools or systems integrated into existing clinical information systems or digital health solutions (e.g., electronic medical records). Implement training programs for healthcare workers to improve skills in using AI, as well as understand the ethical and liability considerations. This is a responsibility for professional colleges, regulators, and health services employing clinicians and health workers.

6.11 Cybersecurity

Key findings: Documents from various international jurisdictions highlighted the importance of robust security measures to protect AI systems and patient data (20, 38, 49, 50, 55, 199). There was a consistent emphasis on considering cybersecurity risks throughout the development and deployment of AI systems in healthcare settings (54, 56, 59). Specifically, documents highlighted cybersecurity breaches such as compromised patient privacy, manipulated data, disrupted critical systems, incorrect recommendations and errors based on inaccurate results (54, 59).

Recommendations to address cybersecurity risks included adopting a risk-based approach (55), developing comprehensive data security plans (42), and prioritising transparency to address cybersecurity concerns effectively (32). Governmental agencies are encouraged to promote transparency in AI systems to ensure accountability and trust in AI (21, 32). Furthermore, most documents emphasised adherence to existing cybersecurity regulatory measures to enhance security standards in the healthcare sector, such as the EU Cybersecurity Act (32), the GDPR (39), the Healthcare Cybersecurity Essentials

Guidelines (49), and the Directive on Security Management (34). Additionally, it was recommended to employ privacy laws or regulations applicable to AI applications, such as the Privacy Impact Assessment decision tool and the Digital Technology Assessment Criteria, to ensure compliance with cybersecurity standards and requirements (1, 204).

The Australian documents frequently mentioned cybersecurity, although organisations typically referred to other relevant legislation and policies rather than identifying bespoke principles for cybersecurity for AI. The Human Technology Institute (64) referred to cybersecurity guidance in the Privacy Act, the NSW Government (67) referred to the NSW Cyber Security Policy, OVIC (66) referred to the Victorian Data Security Frameworks, and the Commonwealth Ombudsman (41) referred to the Digital Service Standards. In the clinical context, only RANZCR and the TGA made specific recommendations pertaining to cybersecurity. RANZCR advised that Practices should implement a user registry to track access to patient information (82), and the TGA advised that physicians should understand cybersecurity risks for software-based medical devices and report any cybersecurity issues to the TGA (77).

Principles:

- Ensure that your organisation has a comprehensive cybersecurity plan to protect against data breach.
- Consider ensuring that clinicians understand cybersecurity, implementing a user registry, and supporting timely security upgrades.

6.12 Guidance specific to pathology tests and medical imaging

Key findings: Overall, there was a lack of explicit guidance on implementing AI for pathology and medical imaging in most of the international documents. However, AI for diagnostics in these fields was mentioned in some documents from international jurisdictions, including WHO's *Ethics and Governance of Artificial Intelligence for Health* (54), and New Zealand's *Emerging Health Technology: Introductory Guidance* (30). Effective data governance and regular monitoring were indicated as essential for handling medical image data and ensuring the ongoing performance of AI-based precision medicine tools (1, 54).

The UK's *Using Machine Learning in Diagnostic Services* report focused on the use of ML applications for diagnostic purposes in healthcare services. It aimed to identify the essential requirements for providing excellent care within services utilising these applications and conduct a thorough evaluation of associated risks (33).

Three of the Australian documents provided specific guidance for medical imaging: the AHPRA Medical Radiation Practice Board's Guidance for Clinical Imaging and Therapeutic Radiology Professionals (88), the Australian College of Dermatologists' Position Statement on Use of Artificial Intelligence in Dermatology in Australia (79), and Chapter 9 of RANZCR's Standards of Practice for Clinical Radiology. None of the Australian documents contained specific guidance for pathology services.

6.13 Other legislative and policy considerations

Key findings: In various documents reviewed, several noteworthy points have been raised that can provide valuable insights into the broader discussions surrounding AI ethics and governance. For instance, there was a notable emphasis on the environmental impact considerations of AI, particularly regarding sustainability. The reviewed documents indicated that AI development, particularly large language models, can have a significant environmental footprint due to their energy consumption (50, 55). On the other hand, AI has the potential to contribute to waste management and conservation efforts, thereby offering environmental benefits (50). Some the documents contained warning against

"technological solutionism", or overreliance on AI as a quick fix (18), and the risk of "data colonialism", particularly in data collection efforts targeting underrepresented groups (24). Data colonialism involves the exploitation of these groups during the data collection process, posing ethical and social risks.

Another concern highlighted in some documents was related to fair labour practices. Concerns were raised about potential job displacement due to automation in certain sectors such as healthcare (19). It was emphasised that fair labour practices throughout AI development are essential to ensure equitable treatment of workers involved in data processing (55). In line with this, the US Secretary of Health and Human Services (HHS) is advised to prioritise efforts to support the development of AI technologies that promote the welfare of patients and healthcare workers to advance responsible AI innovation in the healthcare sector (20).

The potential of citizen science, where non-professionals contribute to AI development through data collection or tasks, was suggested as a means to promote inclusivity (54). Additionally, concerns were also raised about existing liability frameworks lacking legal clarity and failing to uphold patients' rights to seek legal recourse in cases such as misdiagnosis or incorrect treatment facilitated by AI (52). However, the upcoming legislative proposal on AI liability is welcomed as a step toward addressing these concerns.

6.14 Chapter summary

This chapter brought together insights from the national and international environmental scan of legislation and policy relevant to the implementation of AI in acute care. Australian and international policies showed that national ethics frameworks are common. While some international jurisdictions have introduced new AI-specific laws, primary legislation and policy (e.g. privacy laws) remain important in the governance of AI across sectors. Internationally, governance approaches include establishing dedicated regulatory and oversight authorities, requiring impact and risk-based assessments, provisions to increase transparency and prohibit discrimination, regulatory sandboxing, and implementing formal tools or checklists. Current developments in Australian governance and regulation of AI in healthcare include governance via existing cross-sectoral approaches (e.g. privacy and consumer law), regulation of software as a medical device, and specific health governance proposals from health-related research organisations and professional bodies.

7. Key findings from literature review and principles for safe and responsible AI in healthcare

7.1 Introduction

This chapter provides a summary of key findings, incorporating evidence from the literature about AI systems implemented in acute health settings. It presents these findings and principles in the context of implementing specific AI systems in acute health services.

7.2 AI in acute care settings

Key finding 1: AI technologies are being applied in a wide variety of clinical areas. Whilst the literature revealed the top five disease areas in which AI systems were deployed (cancer, stroke, respiratory disease areas, COVID-19 and sepsis), there were a total of 24 disease areas, indicating the spread of AI beyond areas that are more mature in their journey of implementing systems into their respective clinical settings.

Diagnosis and procedures were the most common clinical tasks that were supported by AI systems, predominantly via image analysis but with some that leveraged EHR data inclusive of text data and physiological signal data. Studies often identified a clear clinical use case for implementing AI.

All the AI systems identified in the literature search were based on traditional machine learning (ML) techniques and most were *assistive* requiring clinicians to confirm or approve AI provided information or decisions. Up until December 2023, no studies had evaluated the implementation of AI in hospital operations or the clinical use of foundation models or generative AI in routine patient care.

Principle 1: Take a problem-driven approach to AI implementation, an AI system should address specific clinical needs. Confirm the specific clinical use case before implementation i.e. the types of patients and condition where the AI system is intended to improve care delivery and patient outcomes.

7.3 Approach to AI implementation

Key finding 2: The literature demonstrated multiple ways in which health services implemented AI systems such as to: i/ develop AI systems in-house; ii/ co-develop in partnership with technology companies; and iii/ purchase AI systems from commercial vendors (including AI systems subject to medical device regulation). Evidence of engagement with hospital ethics committees or clinical governance boards from a responsible use perspective was poorly reported in the studies reviewed.

Principle 2: Deployment of AI systems that have been developed externally or internally, is a highly complex process and should be undertaken in partnership with key stakeholders including healthcare professionals and patients. Consultation should occur with those who have specialist skills traversing clinical safety, governance, ethics, IT system architecture legal and procurement, and include the specific healthcare professionals as well as patient representatives and/or patient liaison officers.

Principle 3: When purchasing AI systems from commercial vendors, assess clinical applicability and feasibility of implementation in the care setting. Consider the system performance and whether the ML model will transport from its training and validation environment to the local clinical setting of interest. Consider feasibility of testing the AI using localised de-identified data sets or localised synthetic datasets to illicit utility and performance of the AI system in the local clinical area of interest, before conducting pilot implementation projects.

7.4 AI system performance

Key finding 3: AI system performance was usually assessed against a comparator (e.g. human or another device). Evaluation metrics such as sensitivity, specificity, positive predictive value, accuracy and F1 score were commonplace amongst the literature.

Principle 4: Ensure AI is fit for clinical purposes by assessing evidence for system performance against a comparator. Evaluate performance in the local context of interest using localised de-identified datasets or synthetic datasets, before conducting pilot implementation projects to measure AI system performance and answer any evidence gaps in prior assessments.

Key finding 4: Emerging evidence highlights the impact of distributional shift, stemming from disparities between the dataset on which AI systems are trained and deployment datasets. However, studies describing implementation lacked any reported quality assurance measures, such as post-deployment monitoring, auditing, or performance reviews.

Principle 5: Monitor AI system performance in-situ post deployment, by means of electronic dashboards or other performance monitoring/auditing methods to rapidly detect and mitigate the effects of distributional shift. This should be underpinned by technical support as well as processes around planned and unplanned system downtime.

7.5 Safety of AI in healthcare

Key finding 5: Emerging evidence underscores safety concerns associated with AI systems and their impact on patient care. Although literature reporting on AI-related adverse events has been limited, evidence from the US FDA's post-market safety monitoring emphasises the necessity of examining issues with AI systems beyond the known limitations of ML algorithms. Predominantly, issues with data acquisition were observed, while problems with use i.e. the misapplication of AI and its intended purposes were four times more likely to lead to patient harm that technical issues.

Principle 6: A whole-of-system approach to safe AI implementation is needed. Ensure that AI systems are effectively integrated into IT infrastructure as they are highly reliant on data and integration with the IT infrastructure and other clinical information systems. Data quality and requirements for any accompanying changes to the EMR and other supporting clinical information systems need to be assessed to ensure data provided to the AI system is fit for purpose and its output is accurately displayed to users.

7.6 Role of AI in clinical task, clinical workflow, usability, and safe use

Key finding 6: AI systems in the literature were predominantly assistive or providing autonomous information meaning users were required to confirm or approve AI provided information or decisions, and still had overall autonomy over the task at hand. However, problems with the use of AI were more likely to harm patients compared to algorithm issues in safety events reported to the US FDA's post-market safety monitoring.

Principle 7: Ensure that users are aware of the intended use of AI systems (see Box 3). Training around the intended use and safe use of AI should be developed in consultation with the AI developer, clinical governance, patient safety and clinical leaders. The training should be maintained and updated throughout the life cycle of the AI system.

Key finding 7: End user engagement to devise clinical workflows and training ahead of deployment were less well reported in the literature. When understanding interaction and adoption of AI systems into healthcare workflows, user experience data and user metrics uncovered facilitators and barriers.

Principle 8: Integrate AI systems with clinical workflow. Devise clinical workflows for AI systems in a real-world care setting to ensure AI is seamlessly integrated into practice. Evaluate early to ensure AI fits local requirements and address any issues. A pilot implementation can be used to test and refine integration with clinical workflow and supporting systems.

Principle 9: Identify issues with system usability via user metrics and short, regular survey requests. Address these issues promptly by collaboration with the AI developer and clinicians using the system.

7.7 Clinical utility and effects on decision-making

Key finding 8: Decision change outcomes such as incorrect/correct decisions and the rate at which clinicians make decisions, their decision velocity, help to characterise effects of AI systems on clinical decision-making. More than half of all the literature reviewed (62%) described either of these two outcomes of decision change, largely through analysis of false-positive and false-negative rates, which led researchers to detect possible automation bias in one study. Confidence, acceptability and trust in the AI system were important factors in decision change.

Principle 10: Limitations of the AI system abilities must be made clear to all staff engaging with the AI system. This can be fostered by collaboration with the AI developer and strong engagement with clinicians in both pre-deployment and post deployment phases. Safety events should be easy to report and escalate.

Principle 11: Before-and-after studies or historical cohort studies can be utilised to assess the clinical utility and safety of AI compared to a time period when AI was not implemented. Other prospective methods include AI vs. ground truth by expert consensus; AI vs other gold standard system; AI silent deployment vs visible deployment; and AI vs human corrected designs. Randomised control studies, cross over studies or wash out phase studies are also appropriate study designs.

7.8 Effects on care delivery and patient outcomes

Key finding 9: Care process changes were not well described in the literature, with 28% of studies commenting on or measuring a change in the process of care. When comparing clinical, safety and patient reported outcomes in the literature (n=47), clinical outcomes were more commonly reported (n=40) as either primary, secondary or exploratory outcomes. Only one study described a patient reported outcome (QoL as measured by SF-36 scale).

Principle 12: Ensure AI systems are suitably embedded i.e. their use and clinical utility in a particular context is established using formative evaluation methods during implementation before conducting clinical trials to assess impact on care delivery and patient outcomes.

7.9 Chapter summary

This chapter draws together the main findings from the literature review to identify a general set of principles for practically implementing an AI system at the health service level. The literature supports the notion that AI systems deployed in real-world contexts are making headway in the areas of diagnostic accuracy, as a second reader, and increasing the speed of execution of clinical tasks, particularly of diagnosis tasks. Evidence of AI systems effecting positive change in clinical outcomes is only just

emerging in disease areas which are mature in their AI systems use journey. This next leap hinges on AI systems that provide clear explanation for their outputs to clinicians, are robustly validated, backed by well-designed clinical trials and outcome studies, integrated seamlessly into clinical workflows and used by a workforce that understands how to effectively utilise AI systems in their practice.

8. Conclusion

An overview of findings in the context of Australian health services

The aim of AI technologies in healthcare is to improve care delivery and patient outcomes safely and responsibly. The goal of implementing AI systems in health services should be no different to this, and the findings of this report are designed to support this goal.

The Australian legislative and policy environment is less well-resolved than in some other jurisdictions. Significant opportunities for leadership and coordination exist, and national legislation and policy are in a period of rapid development. Legislation, including cross-sectoral legislation such as privacy and liability law, ethics frameworks, and recognised imperatives such as Indigenous data sovereignty are important reference points. Current proposals for policy change may indicate a way forward: e.g. proposals for an independent National Council to oversee AI implementation in healthcare, and establishment of a formal data governance framework, a risk-based safety framework, and minimum standards of practice. As legislation and policy rapidly develop in this field, high quality community and consumer engagement to ensure that changes reflect public values will be critical.

Despite the current state of flux, the issues discussed in Chapters 2, 3 and 6 demonstrate broad global agreement on key issues for responsible AI governance and implementation. The challenge ahead is to address these issues in the Australian context by implementing consensus guidance, and developing consensus where needed. Key challenges are to establish the forms of governance, oversight and accountability that can ensure that healthcare AI systems best serve consumers, members of the Australian public, and society.

The findings of the literature review suggest there are a wide variety of clinical areas where AI technologies can be applied to improve care delivery and patient outcomes. Many aspects of AI implementation discussed in this report, in particular around the need to ensure usability and integrate AI systems with the local clinical workflow and existing IT infrastructure, are already evident in current practices to implement digital health technologies. One relevant example is medications management in hospitals where a software system is deployed state-wide but needs to be specifically configured to local clinical needs and workflow requirements; so too with AI systems there is a need to tailor AI workflows to local needs.

The challenge ahead is to address the specific requirements of AI systems to ensure they are fit for purpose and their performance is maintained in real-world clinical settings. This applies whether AI is based on traditional ML or foundation models with generative capabilities. AI can report erroneous findings due to differences between training data and real-world populations, as well as differences in the way data is captured in different health services. Managing such algorithm issues requires AI systems to be rigorously evaluated before deployment, and their performance monitored in routine use. Generative AI are prone to hallucination, wherein the AI can provide false information without a sound basis. One possible approach for AI that assists with diagnosis and procedures is to validate their performance against experts using routinely collected data. In parallel, users also need to be trained to operate AI within the bounds of its design or delegated authority, as well as in the procedures needed to verify AI provided information, monitor AI performance and intervene when AI fails. Users must be particularly vigilant with generative AI, as it is rapidly appearing in current clinical information systems.

There are limitations of the findings presented in this report, which aimed to undertake a scoping review of recent studies reporting AI implementation and undertake an environmental scan to identify principles that enable the safe and responsible implementation of AI in healthcare. First, this review is limited to the

published research literature about AI systems in acute care settings. We did not include grey literature, such as white papers and reports. Second, our analysis of the level of AI autonomy was limited to the information that was reported in the papers and prior validation studies. This information was less structured than the indications of use in medical device approval documents which informed the development of the level of autonomy. Finally, there was considerable heterogeneity in the study designs and outcome measures which prevented quantitative examination of the effects on decision-making, care delivery, and patient outcomes.

There are several important related areas which emerged from the study, and further work is recommended to consider these other areas. For example, there is a need to examine the implementation science literature for approaches and frameworks that are applicable to AI. Another important area for further work is around guidance for Australian health services about how to govern the safe and responsible implementation and use of AI. While there are a multitude of theoretical frameworks for AI, little is known about how they are operationalised for the governance of AI systems which are of varying levels of technical maturity; can incorporate many different types of computational reasoning methods including traditional and generative AI; and be used in a wide variety of clinical and non-clinical areas. Effective governance at the health service level is critical not only to ensure safe and effective deployment, but to foster clinician trust that leads to meaningful adoption as well as improvements in care delivery and patient outcomes. Further work is also required to examine the organisational capacity to implement complementary innovations in culture, leadership, and workforce that are required to effectively harness AI.

AI technologies are just beginning to be used in Australian healthcare. By implementing AI safely and responsibly, building on existing governance processes, strengthening engagement with consumers, and creating strong evaluation processes to assess its performance and clinical usefulness according to current best practices, Australian health services can get ready for the future. This readiness is essential as AI systems advance from providing recommendations to autonomously carrying out clinical tasks. Furthermore, Australia could provide valuable guidance to other countries seeking to use modern AI systems safely and effectively to improve patient care and outcomes.

References

1. US Department of Health and Human Services. Artificial Intelligence Strategy. 2021. Available from: https://www.hhs.gov/sites/default/files/final-hhs-ai-strategy.pdf. [Accessed March 2024].
2. Australian Government Department of Industry, Science and Resources. Safe and responsible AI in Australia: Discussion paper. 2023. Available from: https://storage.googleapis.com/converlens-au-industry/industry/p/prj2452c8e24d7a400c72429/public_assets/Safe-and-responsible-AI-in-Australia-discussion-paper.pdf. [Accessed March 2024].
3. The Royal Australian and New Zealand College of Radiologists. Ethical Principles for AI in Medicine. 2023. Available from: https://www.ranzcr.com/search/ethical-principles-for-ai-in-medicine. [Accessed March 2024].
4. Australian Medical Association. Artificial Intelligence in Healthcare. 2023. Available from: https://www.ama.com.au/sites/default/files/2023-08/Artificial%20Intelligence%20in%20Healthcare%20-%20AMA.pdf. [Accessed March 2024].
5. Coiera E. Guide to health informatics, third edition: CRC Press; 2015.
6. Coiera E. The fate of medicine in the time of AI. Lancet. 2018. 10.1016/s0140-6736(18)31925-1
7. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56. https://doi.org/10.1038/s41591-018-0300-7
8. Lyell D, Coiera E, Chen J, Shah P, Magrabi F. How machine learning is embedded to support clinician decision making: an analysis of FDA-approved medical devices. BMJ Health Care Inform. 2021;28(1). https://doi.org/10.1136/bmjhci-2020-100301
9. Domingos P. The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World: Penguin Books Limited; 2015.
10. Parasuraman R, Sheridan TB, Wickens CD. A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans. 2000;30(3):286-97. https://doi.org/10.1109/3468.844354
11. Coiera E, Liu S. Evidence synthesis, digital scribes, and translational challenges for artificial intelligence in healthcare. Cell Rep Med. 2022;3(12):100860. https://doi.org/10.1016/j.xcrm.2022.100860
12. Coiera E. The Last Mile: Where Artificial Intelligence Meets Reality. J Med Internet Res. 2019;21(11):e16323. https://doi.org/10.2196/16323
13. Morley J, Murphy L, Mishra A, Joshi I, Karpathakis K. Governing Data and Artificial Intelligence for Health Care: Developing an International Understanding. JMIR Form Res. 2022;6(1):e31623. https://doi.org/10.2196/31623
14. Ulnicane I, Eke DO, Knight W, Ogoh G, Stahl BC. Good governance as a response to discontents? Déjà vu, or lessons for AI from other emerging technologies. Interdisciplinary Science Reviews. 2021;46(1-2):71-93. https://doi.org/10.1080/03080188.2020.1840220
15. Australian Government Department of Industry, Science and Resources. Safe and responsible AI in Australia consultation - Australian Government's interim response. 2024. Available from: https://storage.googleapis.com/converlens-au-industry/industry/p/prj2452c8e24d7a400c72429/public_assets/safe-and-responsible-ai-in-australia-governments-interim-response.pdf. [Accessed March 2024].

16. State of New Jersey US. Bill S1402 An Act concerning discrimination and automated decision systems and supplementing P.L.1945, c.169 (C.10:5-1 et seq.). 2022. Available from: https://www.njleg.state.nj.us/bill-search/2022/S1402/bill-text?f=S1500&n=1402_I1. [Accessed March 2024].

17. Council of the District of Columbia US. Bill 25-0114 - Stop Discrimination by Algorithms Act of 2023. 2023. Available from: https://lims.dccouncil.gov/Legislation/B25-0114. [Accessed March 2024].

18. New York State Senate US. Bill A3308 Digital Fairness Act. 2023. Available from: https://www.nysenate.gov/legislation/bills/2023/A3308. [Accessed March 2024].

19. US Congress. H.R.6216 National Artificial Intelligence Initiative Act of 2020. 2020. Available from: https://www.congress.gov/bill/116th-congress/house-bill/6216/text. [Accessed March 2024].

20. US White House. Executive Order: the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. 2023. Available from: https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/. [Accessed March 2024].

21. Office of Management and Budget. M-21-06: Memorandum for the heads of executive deparments and agencies: Guidance for Regulation of Artificial Intelligence Applications. 2020. Available from: https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf. [Accessed March 2024].

22. European Union. Artificial Intelligence Act (EU). 2024. Available from: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN. [Accessed March 2024].

23. Government of Canada. Bill C-27: An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts. 2022. Available from: https://www.justice.gc.ca/eng/csj-sjc/pl/charter-charte/c27_1.html. [Accessed March 2024].

24. US Department of Health and Human Services. Nondiscrimination in Health Programs and Activities (87 FR 47824). Federal Register. 2022. Available from: https://www.federalregister.gov/documents/2022/08/04/2022-16217/nondiscrimination-in-health-programs-and-activities. [Accessed March 2024].

25. Australian Academy of Technological Sciences and Engineering. Submission to the inquiry into artificial intelligence in New South Wales. 2023. Available from: https://www.atse.org.au/wp-content/uploads/2023/10/SBM-2023-10-20-NSW-AI-submission.pdf. [Accessed March 2024]

26. US Department of Health and Human Services. Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing (89 FR 1192). 2024. Available from: https://www.federalregister.gov/documents/2024/01/09/2023-28857/health-data-technology-and-interoperability-certification-program-updates-algorithm-transparency-and. [Accessed March 2024].

27. UK Government. Artifical Intelligence (Regulation) Bill [HL]. 2023. Available from: https://bills.parliament.uk/bills/3519. [Accessed March 2024].

28. New Zealand Department of the Prime Minister and Cabinet. Capturing the benefits of AI in healthcare for Aotearoa New Zealand, Office of the Prime Minister's Chief Science Advisor. 2023. Available from: https://www.dpmc.govt.nz/sites/default/files/2024-01/PMCSA-23-12-03-V3-PMCSA-AI-healthcare-LONG-REPORT-FINAL-%28pdf-version%29-v3.pdf. [Accessed March 2024].

29. New Zealand Government. Algorithm Charter for Aotearoa New Zealand. 2020. Available from: https://www.data.govt.nz/toolkit/data-ethics/government-algorithm-transparency-and-accountability/algorithm-charter/. [Accessed March 2024].
30. New Zealand Ministry of Health. Emerging Health Technology: Introductory Guidance. 2019. Available from: https://www.tewhatuora.govt.nz/assets/Uploads/introductory_guidance_-_algorithms_v0.4_-_web.pdf. [Accessed March 2024].
31. UK National Health Services. Artificial Intelligence: How to get it right. 2019. Available from: https://transform.england.nhs.uk/ai-lab/explore-all-resources/understand-ai/artificial-intelligence-how-get-it-right/. [Accessed March 2024].
32. European Union. Understanding algorithmic decision-making- Opportunities and challenges. 2019. Available from: https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2019)624261. [Accessed March 2024].
33. UK Care Quality Commission Using machine learning in diagnostic services: a report with recommendations from CQC's regulatory sandbox. 2022. Available from: https://www.cqc.org.uk/sites/default/files/20200324%20CQC%20sandbox%20report_machine%20learning%20in%20diagnostic%20services.pdf. [Accessed March 2024].
34. Government of Canada. Directive on Automated Decision-Making. 2020. Available from: https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592. [Accessed March 2024].
35. Infocommunications Media Development Authority of Singapore. AI Verify: AI Governance Testing Framework and Toolkit. 2022. Available from: https://aiverifyfoundation.sg/what-is-ai-verify/. [Accessed March 2024].
36. European Union. Ethics Guidelines for Trustworthy AI. 2019. Available from: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai. [Accessed March 2024].
37. Infocommunications Media Development Authority of Singapore. Companion to the Model AI Governance Framework. 2020. Available from: https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGIsago.pdf. [Accessed March 2024].
38. Alan Turing Institute. Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. 2019. Available from: https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf. [Accessed March 2024].
39. UK National Health Services. A Buyer's Guide to AI in Health and Care. 2020. Available from: https://transform.england.nhs.uk/ai-lab/explore-all-resources/adopt-ai/a-buyers-guide-to-ai-in-health-and-care/. [Accessed March 2024].
40. US Department of Health and Human Services. Trustworthy AI Playbook. 2021. Available from: https://www.hhs.gov/sites/default/files/hhs-trustworthy-ai-playbook.pdf. [Accessed March 2024].
41. Commonwealth Ombudsman. Automated Decision-Making - Better Practice Guide. 2019. Available from: https://www.ombudsman.gov.au/__data/assets/pdf_file/0029/288236/OMB1188-Automated-Decision-Making-Report_Final-A1898885.pdf. [Accessed March 2024]
42. US Government Accountability Office. Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities. 2021. Available from: https://www.gao.gov/products/gao-21-519sp. [Accessed March 2024].
43. Nhat PTH, Van Hao N, Tho PV, Kerdegari H, Pisani L, Thu LNM, et al. Clinical benefit of AI-assisted lung ultrasound in a resource-limited intensive care unit. Critical Care. 2023;27(1):257. https://doi.org/10.1186/s13054-023-04548-w

44. UK Department for Science Innovation and Technology. A pro-innovation approach to AI regulation: government response. 2024. Available from: https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response. [Accessed March 2024].

45. UK Government. The Bletchley Declaration by Countries Attending the AI Safety Summit. 2023. Available from: https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023#contents. [Accessed March 2024].

46. Health New Zealand. Advice on the use of Large Language Models and Generative AI in Healthcare. 2023. Available from: https://www.tewhatuora.govt.nz/our-health-system/digital-health/national-ai-and-algorithm-expert-advisory-group-naiaeag-te-whatu-ora-advice-on-the-use-of-large-language-models-and-generative-ai-in-healthcare/. [Accessed March 2024].

47. World Economic Forum. Reimaging Regulation for the Age of AI: New Zealand Pilot Project. 2020. Available from: https://www3.weforum.org/docs/WEF_Reimagining_Regulation_Age_AI_2020.pdf. [Accessed March 2024].

48. Personal Data Protection Commission of Singapore. Model AI Governance Framework, Second Edition. 2020. Available from: https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf. [Accessed March 2024].

49. Singapore Ministry of Health. Artificial Intelligence in Healthcare Guidelines. 2021. Available from: https://www.moh.gov.sg/licensing-and-regulation/artificial-intelligence-in-healthcare. [Accessed March 2024].

50. European Union. The ethics of artificial intelligence: Issues and initiatives. 2020. Available from: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf. [Accessed March 2024].

51. European Commission. Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics. 2020. Available from: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52020DC0064. [Accessed March 2024].

52. European Union. Motion for a European Parliament Resolution on artificial intelligence in a digital age 2020/2266(INI). 2022. [Accessed March 2024].

53. European Union. Framework of ethical aspects of artificial intelligence, robotics and related technologies. 2020. Available from: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_EN.html. [Accessed March 2024].

54. World Health Organization. Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models. 2024. Available from: https://www.who.int/publications/i/item/9789240084759. [Accessed March 2024].

55. World Health Organization. Regulatory considerations on artificial intelligence for health. 2023. Available from: https://www.who.int/publications/i/item/9789240078871. [Accessed March 2024].

56. World Health Organization. Generating Evidence for Artificial Intelligence Based Medical Devices: A Framework for Training Validation and Evaluation. 2021. Available from: https://www.who.int/publications/i/item/9789240038462. [Accessed March 2024].

57. World Health Organization. Ethics and governance of artificial intelligence for health. 2021. Available from: https://www.who.int/publications/i/item/9789240029200. [Accessed March 2024].

58. Organisation for Economic Co-operation and Development. Collective action for responsible AI in health. OECD. 2024. Available from: https://www.oecd.org/publications/collective-action-for-responsible-ai-in-health-f2050177-en.htm. [Accessed March 2024].

59. Organisation for Economic Co-operation and Development. OECD Framework for the Classification of AI systems. OECD. 2022. Available from: https://www.oecd.org/publications/oecd-framework-for-the-classification-of-ai-systems-cb6d9eca-en.htm. [Accessed March 2024].

60. Chonde DB, Pourvaziri A, Williams J, McGowan J, Moskos M, Alvarez C, et al. RadTranslate: an artificial intelligence–powered intervention for urgent imaging to enhance care equity for patients with limited English proficiency during the COVID-19 pandemic. Journal of the American College of Radiology. 2021;18(7):1000-8. https://doi.org/10.1016/j.jacr.2021.01.013

61. Tierney AA, Gayre G, Hoberman B, Mattern B, Ballesca M, Kipnis P, et al. Ambient Artificial Intelligence Scribes to Alleviate the Burden of Clinical Documentation. NEJM Catalyst. 2024;5(3):CAT.23.0404. https://doi.org/10.1056/CAT.23.0404

62. Australian Government Department of Industry Science and Resources. Australia's Artificial Intelligence Ethics Framework [Available from: https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework]. [Accessed March 2024]

63. Australian Government Department of Industry, Science and Resources. Australia's Artificial Intelligence Action Plan. 2021. Available from: https://www.industry.gov.au/publications/australias-artificial-intelligence-action-plan. [Accessed March 2024].

64. Solomon L, Davis PN. The State of AI Governance in Australia. The Human Technology Institute, The University of Technology Sydney 2023. Available from: https://www.uts.edu.au/sites/default/files/2023-05/HTI%20The%20State%20of%20AI%20Governance%20in%20Australia%20-%2031%20May%202023.pdf. [Accessed March 2024]

65. Office of the Australian Information Commissioner. OAIC submission to the Department of Industry, Science and Resources – Safe and responsible AI in Australia discussion paper. 2023. Available from: https://www.oaic.gov.au/engage-with-us/submissions/oaic-submission-to-the-department-of-industry-science-and-resources-safe-and-responsible-ai-in-australia-discussion-paper. [Accessed March 2024].

66. Office of the Victorian Information Commissioner. Artificial Intelligence – Understanding Privacy Obligations. 2021. Available from: https://ovic.vic.gov.au/privacy/resources-for-organisations/artificial-intelligence-understanding-privacy-obligations/. [Accessed March 2024].

67. NSW Government, Artificial Intelligence Assurance Framework. Available from: https://www.digital.nsw.gov.au/sites/default/files/2022-09/nsw-government-assurance-framework.pdf [Accessed March 2024]. 2022. [Accessed March 2024]

68. Therapeutic Goods Administration. Examples of regulated and unregulated software (excluded) software based medical devices. 2021. Available from: https://www.tga.gov.au/sites/default/files/examples-regulated-and-unregulated-software-excluded-software-based-medical-devices.pdf. [Accessed March 2024].

69. Australian Human Rights Commission. Human Rights and Technology: Final Report. 2021. Available from: https://humanrights.gov.au/sites/default/files/document/publication/ahrc_rightstech_2021_final_report_10.pdf. [Accessed March 2024]

70. Data61. Artificial Intelligence - Australia's Ethics Framework. 2019. [Accessed March 2024]

71. Digital.NSW. Mandatory Ethical Principles for the use of AI. 2021. Available from: https://www.digital.nsw.gov.au/policy/artificial-intelligence/artificial-intelligence-ethics-policy/mandatory-ethical-principles. [Accessed March 2024].

72. Australian Government Department of Industry, Science and Resources. Australia's AI Ethics Principles. 2022. Available from: https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles. [Accessed March 2024].

73. Soltan AAS, Yang J, Pattanshetty R, Novak A, Yang Y, Rohanian O, et al. Real-world evaluation of rapid and laboratory-free COVID-19 triage for emergency care: external validation and pilot deployment of artificial intelligence driven screening. The Lancet Digital Health. 2022;4(4):e266-e78. https://doi.org/10.1016/S2589-7500(21)00272-7

74. Australian Government Architecture. Interim guidance on government use of public generative AI tools - November 2023. 2023. Available from: https://architecture.digital.gov.au/guidance-generative-ai. [Accessed March 2024].

75. Weatherall K, Henman P, Bello y Villarino J-M, Matulionyte R, Sleep L, Trezise M, et al. Executive Report Automated decision-making in NSW: Mapping and analysis of the use of ADM systems by state and local governments (Research Report). ADM+S 2024. Available from: https://www.ombo.nsw.gov.au/__data/assets/pdf_file/0005/145094/Executive-Report-ADMS.pdf. [Accessed March 2024]

76. Therapeutic Goods Administration. Clinical decision support software. 2021. Available from: https://www.tga.gov.au/resources/resource/guidance/clinical-decision-support-software. [Accessed March 2024].

77. Therapeutic Goods Administration. Digital mental health: Software based medical devices. 2022. Available from: https://www.tga.gov.au/sites/default/files/digital-mental-health-software-based-medical-devices.pdf. [Accessed March 2024].

78. Therapeutic Goods Administration. Regulation of software-based medical devices 2023 [Available from: https://www.tga.gov.au/how-we-regulate/manufacturing/medical-devices/manufacturer-guidance-specific-types-medical-devices/regulation-software-based-medical-devices]. [Accessed March 2024]

79. The Australasian College of Dermatologists. Position statement: Use of Artificial Intelligence in Dermatology in Australia. 2022. Available from: https://www.dermcoll.edu.au/wp-content/uploads/2022/11/ACD-Position-Statement-Use-of-Artifical-Intelligence-in-Dermatology-in-Australia-Nov-2022.pdf. [Accessed March 2024].

80. Victoria State Government Department of Health. Health service use of unregulated Artificial Intelligence (AI) - Health Service Advisory. 2023. Available from: https://www.safercare.vic.gov.au/sites/default/files/2023-07/Advisory%20-%20ChatGPT%20and%20Generative%20AI%20July%202023%20FINAL.pdf. [Accessed March 2024].

81. Australian Alliance for Artificial Intelligence in Healthcare. A National Policy Roadmap for Artificial Intelligence in Healthcare. 2023. Available from: https://aihealthalliance.org/wp-content/uploads/2023/11/AAAiH_NationalPolicyRoadmap_FINAL.pdf. [Accessed March 2024]

82. The Royal Australian and New Zealand College of Radiologists. Standards of Practice for Clinical Radiology. 2020.Report No.: 11.2. Available from: https://www.ranzcr.com/college/document-library/ranzcr-standards-of-practice-for-diagnostic-and-interventional-radiology. [Accessed March 2024]

83. Digital.NSW. Artificial Intelligence Strategy - Strategy Overview. 2021. Available from: https://www.digital.nsw.gov.au/policy/artificial-intelligence/artificial-intelligence-strategy/strategy-overview. [Accessed March 2024].

84. Carter SM, Aquino YSJ, Carolan L, Frost E, Degeling C, Rogers WA, et al. How should artificial intelligence be used in Australian health care? Recommendations from a citizens' jury. Medical Journal of Australia. 2024. https://doi.org/10.5694/mja2.52283

85. Digital.NSW. Ethical Policy Statement. 2022. Available from: https://www.digital.nsw.gov.au/policy/artificial-intelligence/artificial-intelligence-ethics-policy/ethical-policy-statement. [Accessed March 2024].

86. Australian Government Architecture. Artificial Intelligence policy (Position). 2023. Available from: https://architecture.digital.gov.au/artificial-intelligence-policy-position. [Accessed March 2024].

87. Australian Government Architecture. Adoption of Artificial Intelligence in the Public Sector. 2022. Available from: https://architecture.digital.gov.au/adoption-artificial-intelligence-public-sector-0. [Accessed March 2024].

88. AHPRA Medical Radiation Practice Board. Artificial Intelligence: Guidance for clinical imaging and therapeutic radiography professionals, a summary by the Society of Radiographers AI working group. 2022. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1078817421001085. [Accessed March 2024]

89. Digital.NSW. Artificial Intelligence Ethics Policy - Key considerations 2021. Available from: https://www.digital.nsw.gov.au/policy/artificial-intelligence/artificial-intelligence-ethics-policy/key-considerations. [Accessed March 2024].

90. Department of the Premier and Cabinet Office of Digital Government. Interim Guidance for WA Public Sector Agencies on Adoption of Artificial Intelligence. 2023. Available from: https://www.wa.gov.au/system/files/2023-08/aiinterimguidance.pdf. [Accessed March 2024].

91. Queensland Government. Use of generative AI for government - information sheet. 2023. Available from: https://www.forgov.qld.gov.au/__data/assets/pdf_file/0028/416647/Use-of-generative-AI-for-gov-information-sheet.pdf. [Accessed March 2024].

92. Medical Technology Association of Australia. Digital Health: Breaking Barriers to Deliver Better Patient Outcomes. 2023. Available from: https://www.mtaa.org.au/sites/default/files/uploaded-content/website-content/digital_health_-_breaking_barriers_to_deliver_better_patient_outcomes_report.pdf. [Accessed March 2024]

93. NSW Government. Using public generative artificial intelligence (AI) tools safely. Available from: https://www.digital.nsw.gov.au/sites/default/files/2023-10/Using-public-generative-artificial-intelligence-AI-tools-safely.pdf 2023. [Accessed March 2024]

94. Digital.NSW. Generative AI: basic guidance. Available from: https://www.digital.nsw.gov.au/policy/artificial-intelligence/generative-ai-basic-guidance. [Accessed March 2024].

95. Royal Australasian College of Medical Administrators. Position Statement: Digital Health. RACMA2020. Available from: https://racma.edu.au/advocacy/position-statements/2020-position-statements/digital-health/. [Accessed March 2024].

96. Therapeutic Goods Administration. Medical device cyber security information for users. 2022. Available from: https://www.tga.gov.au/resources/resource/guidance/medical-device-cyber-security-information-users. [Accessed March 2024].

97. Plana D, Shung DL, Grimshaw AA, Saraf A, Sung JJY, Kann BH. Randomized Clinical Trials of Machine Learning Interventions in Health Care: A Systematic Review. JAMA Netw Open. 2022;5(9):e2233946. https://doi.org/10.1001/jamanetworkopen.2022.33946

98. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. The Lancet. 2020;395(10236):1579-86. https://doi.org/10.1016/S0140-6736(20)30226-9

99. Howell MD, Corrado GS, DeSalvo KB. Three Epochs of Artificial Intelligence in Health Care. JAMA. 2024;331(3):242-4. https://doi.org/10.1001/jama.2023.25057

100. Evans RS. Electronic Health Records: Then, Now, and in the Future. Yearb Med Inform. 2016;Suppl 1(Suppl 1):S48-61. https://doi.org/10.15265/IYS-2016-s006

101. Jang H-J, Cho K-O. Applications of deep learning for the analysis of medical data. Archives of Pharmacal Research. 2019;42(6):492-504. https://doi.org/10.1007/s12272-019-01162-9

102. Avula V, Wu KC, Carrick RT. Clinical Applications, Methodology, and Scientific Reporting of Electrocardiogram Deep-Learning Models: A Systematic Review. JACC Adv. 2023;2(10). https://doi.org/10.1016/j.jacadv.2023.100686

103. Flanary J, Daly SR, Bakker C, Herman AB, Park MC, McGovern R, et al. Reliability of visual review of intracranial electroencephalogram in identifying the seizure onset zone: A systematic review and implications for the accuracy of automated methods. Epilepsia. 2023;64(1):6-16. https://doi.org/10.1111/epi.17446

104. Chapalain X, Huet O. Is artificial intelligence (AI) at the doorstep of Intensive Care Units (ICU) and operating room (OR)? Anaesth Crit Care Pain Med. 2019;38(4):337-8. https://doi.org/10.1016/j.accpm.2019.05.003

105. Smith LA, Oakden-Rayner L, Bird A, Zeng M, To MS, Mukherjee S, et al. Machine learning and deep learning predictive models for long-term prognosis in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. Lancet Digit Health. 2023;5(12):e872-e81. https://doi.org/10.1016/s2589-7500(23)00177-2

106. Susanto AP, Lyell D, Widyantoro B, Berkovsky S, Magrabi F. Effects of machine learning-based clinical decision support systems on decision-making, care delivery, and patient outcomes: a scoping review. J Am Med Inform Assoc. 2023;30(12):2050-63. https://doi.org/10.1093/jamia/ocad180

107. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthc J. 2021;8(2):e188-e94. https://doi.org/10.7861/fhj.2021-0095

108. Magrabi F, Lyell D, Coiera E. Automation in Contemporary Clinical Information Systems: a Survey of AI in Healthcare Settings. Yearb Med Inform. 2023;32(01):115-26. https://doi.org/10.1055/s-0043-1768733

109. World Health Organization. World health statistics 2023: monitoring health for the SDGs, Sustainable Development Goals. Geneva 2023. Available from: https://www.who.int/publications/i/item/9789240074323. [Accessed March 2024]

110. Bagci U, Irmakci I, Demir U, Keles E. Building Blocks of AI. AI in Clinical Medicine: Wiley Online Books; 2023. p. 5-6, 156. https://doi.org/10.1002/9781119790686.ch6

111. Medical Board of Australia. The current COAG Health Council Approved List of specialties, fields and related titles. Medical Board of Australia. 2018. Available from: https://www.medicalboard.gov.au/Registration/Recognition-of-medical-specialties.aspx. [Accessed March 2024].

112. Coiera E. Assessing Technology Success and Failure Using Information Value Chain Theory. Stud Health Technol Inform. 2019;263:35-48. https://doi.org/10.3233/shti190109

113. Martinez VA, Betts RK, Scruth EA, Buckley JD, Cadiz VR, Bertrand LD, et al. The Kaiser Permanente Northern California Advance Alert Monitor Program: An Automated Early Warning System for Adults at Risk for In-Hospital Clinical Deterioration. The Joint Commission Journal on Quality and Patient Safety. 2022;48(8):370-5. https://doi.org/10.1016/j.jcjq.2022.05.005

114. Martinez-Gutierrez JC, Kim Y, Salazar-Marioni S, Tariq MB, Abdelkhaleq R, Niktabe A, et al. Automated Large Vessel Occlusion Detection Software and Thrombectomy Treatment Times: A Cluster Randomized Clinical Trial. JAMA Neurology. 2023;80(11):1182-90. https://doi.org/10.1001/jamaneurol.2023.3206

115. Nasir-Moin M, Suriawinata AA, Ren B, Liu X, Robertson DJ, Bagchi S, et al. Evaluation of an Artificial Intelligence–Augmented Digital System for Histologic Classification of Colorectal Polyps. JAMA Network Open. 2021;4(11):e2135271-e. https://doi.org/10.1001/jamanetworkopen.2021.35271

116. Kanbar LJ, Wissel B, Ni Y, Pajor N, Glauser T, Pestian J, et al. Implementation of Machine Learning Pipelines for Clinical Practice: Development and Validation Study. JMIR Med Inform. 2022;10(12):e37833. https://doi.org/10.2196/37833

117. Cavallo JJ, de Oliveira Santo I, Mezrich JL, Forman HP. Clinical Implementation of a Combined Artificial Intelligence and Natural Language Processing Quality Assurance Program for Pulmonary Nodule Detection in the Emergency Department Setting. Journal of the American College of Radiology. 2023;20(4):438-45. https://doi.org/10.1016/j.jacr.2022.12.016

118. Eng DK, Khandwala NB, Long J, Fefferman NR, Lala SV, Strubel NA, et al. Artificial Intelligence Algorithm Improves Radiologist Performance in Skeletal Age Assessment: A Prospective Multicenter Randomized Controlled Trial. Radiology. 2021;301(3):692-9. https://doi.org/10.1148/radiol.2021204021

119. Ericson OH, J; Sjovall,Fredrik; Soderberg, J; Persson I. The Potential Cost and Cost-Effectiveness Impact of Using a Machine Learning Algorithm for Early Detection of Sepsis in Intensive Care Units in Sweden. Journal of Health Economics and Outcomes Research. 2022;9(1):101-10. https://doi.org/10.36469/jheor.2022.33951

120. Gunda B, Neuhaus A, Sipos I, Stang R, Böjti PP, Takács T, et al. Improved Stroke Care in a Primary Stroke Centre Using AI-Decision Support. Cerebrovascular Diseases Extra. 2022;12(1):28-32. https://doi.org/10.1159/000522423

121. Rawson TM, Hernandez B, Moore LSP, Herrero P, Charani E, Ming D, et al. A Real-world Evaluation of a Case-based Reasoning Algorithm to Support Antimicrobial Prescribing Decisions in Acute Care. Clin Infect Dis. 2021;72(12):2103-11. https://doi.org/10.1093/cid/ciaa383

122. van Leeuwen KG, Meijer FJA, Schalekamp S, Rutten MJCM, van Dijk EJ, van Ginneken B, et al. Cost-effectiveness of artificial intelligence aided vessel occlusion detection in acute stroke: an early health technology assessment. Insights into Imaging. 2021;12(1):133. https://doi.org/10.1186/s13244-021-01077-4

123. Cheema BS, Walter J, Narang A, Thomas JD. Artificial Intelligence–Enabled POCUS in the COVID-19 ICU: A New Spin on Cardiac Ultrasound. JACC: Case Reports. 2021;3(2):258-63. https://doi.org/10.1016/j.jaccas.2020.12.013

124. Lee S, Shin HJ, Kim S, Kim E-K. Successful Implementation of an Artificial Intelligence-Based Computer-Aided Detection System for Chest Radiography in Daily Clinical Practice. Korean J Radiol. 2022;23(9):847-52. https://doi.org/10.3348/kjr.2022.0193

125. Maheshwarappa HM, Mishra S, Kulkarni AV, Gunaseelan V, Kanchi M. Use of Handheld Ultrasound Device with Artificial Intelligence for Evaluation of Cardiorespiratory System in COVID-19. Indian J Crit Care Med. 2021;25(5):524-7. https://doi.org/10.5005/jp-journals-10071-23803

126. Peng S, Liu Y, Lv W, Liu L, Zhou Q, Yang H, et al. Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study. The Lancet Digital Health. 2021;3(4):e250-e9. https://doi.org/10.1016/S2589-7500(21)00041-8

127. Edalati M, Zheng Y, Watkins MP, Chen J, Liu L, Zhang S, et al. Implementation and prospective clinical validation of AI-based planning and shimming techniques in cardiac MRI. Medical Physics. 2022;49(1):129-43. https://doi.org/10.1002/mp.15327

128. Rabinovich D, Mosquera C, Torrens P, Aineseder M, Benitez S. User Satisfaction with an AI System for Chest X-Ray Analysis Implemented in a Hospital's Emergency Setting. IOS Press; 2022. https://dx.doi.org/10.3233/shti220386

129. King W. HeRO: AI with Evidence. Neonatal Intensive Care. 2022;35:48 - 51.

130. Dean NC, Vines CG, Carr JR, Rubin JG, Webb BJ, Jacobs JR, et al. A Pragmatic, Stepped-Wedge, Cluster-controlled Clinical Trial of Real-Time Pneumonia Clinical Decision Support. American Journal of Respiratory and Critical Care Medicine. 2022;205(11):1330-6. https://doi.org/10.1164/rccm.202109-2092oc

131. Alessandro R, Marco S, Giulio A, Loredana C, Roberta M, Piera Alessia G, et al. Artificial intelligence and colonoscopy experience: lessons from two randomised trials. Gut. 2022;71(4):757. https://doi.org/10.1136/gutjnl-2021-324471

132. Martins Jarnalo CO, Linsen PVM, Blazís SP, van der Valk PHM, Dieckens DBM. Clinical evaluation of a deep-learning-based computer-aided detection system for the detection of pulmonary nodules in a large teaching hospital. Clinical Radiology. 2021;76(11):838-45. https://doi.org/10.1016/j.crad.2021.07.012

133. Cerminara SE, Cheng P, Kostner L, Huber S, Kunz M, Maul J-T, et al. Diagnostic performance of augmented intelligence with 2D and 3D total body photography and convolutional neural networks in a high-risk population for melanoma under real-world conditions: A new era of skin cancer screening? European Journal of Cancer. 2023;190:112954. https://doi.org/10.1016/j.ejca.2023.112954

134. Hong JC, Eclov NCW, Stephens SJ, Mowery YM, Palta M. Implementation of machine learning in the clinic: challenges and lessons in prospective deployment from the System for High Intensity EvaLuation During Radiation Therapy (SHIELD-RT) randomized controlled study. BMC Bioinformatics. 2022;23(12):408. https://doi.org/10.1186/s12859-022-04940-3

135. Hu M, Chen N, Zhou X, Wu Y, Ma C. Deep Learning-Based Computed Tomography Perfusion Imaging to Evaluate the Effectiveness and Safety of Thrombolytic Therapy for Cerebral Infarct with Unknown Time of Onset. Contrast Media & Molecular Imaging. 2022;2022:1-8. https://doi.org/10.1155/2022/9684584

136. Knighton AJ, Kuttler KG, Ranade-Kharkar P, Allen L, Throne T, Jacobs JR, et al. An alert tool to promote lung protective ventilation for possible acute respiratory distress syndrome. JAMIA Open. 2022;5(2):ooac050. https://doi.org/10.1093/jamiaopen/ooac050

137. Sarti AJ, Katina Z, Christophe LH, Stephanie S, Nathan BS, Irene W, et al. Feasibility of implementing Extubation Advisor, a clinical decision support tool to improve extubation decision-making in the ICU: a mixed-methods observational study. BMJ Open. 2021;11(8):e045674. https://doi.org/10.1136/bmjopen-2020-045674

138. Schmuelling L, Franzeck FC, Nickel CH, Mansella G, Bingisser R, Schmidt N, et al. Deep learning-based automated detection of pulmonary embolism on CT pulmonary angiograms: No significant effects on report communication times and patient turnaround in the emergency department nine months after technical implementation. European Journal of Radiology. 2021;141. https://doi.org/10.1016/j.ejrad.2021.109816

139. Chen J, Gao Y. The Role of Deep Learning-Based Echocardiography in the Diagnosis and Evaluation of the Effects of Routine Anti-Heart-Failure Western Medicines in Elderly Patients with Acute Left Heart Failure. Journal of Healthcare Engineering. 2021;2021:4845792. https://doi.org/10.1155/2021/4845792

140. Howell RS, Liu HH, Khan AA, Woods JS, Lin LJ, Saxena M, et al. Development of a Method for Clinical Evaluation of Artificial Intelligence–Based Digital Wound Assessment Tools. JAMA Network Open. 2021;4(5):e217234-e. https://doi.org/10.1001/jamanetworkopen.2021.7234

141. Pangti R, Mathur J, Chouhan V, Kumar S, Rajput L, Shah S, et al. A machine learning-based, decision support, mobile phone application for diagnosis of common dermatological diseases. Journal of the European Academy of Dermatology and Venereology. 2021;35(2):536-45. https://doi.org/10.1111/jdv.16967

142. Hao S, Liu C, Li N, Wu Y, Li D, Gao Q, et al. Clinical evaluation of AI-assisted screening for diabetic retinopathy in rural areas of midwest China. PLOS ONE. 2022;17(10):e0275983. https://doi.org/10.1371/journal.pone.0275983

143. Ou Y-C, Tsao T-Y, Chang M-C, Lin Y-S, Yang W-L, Hang J-F, et al. Evaluation of an artificial intelligence algorithm for assisting the Paris System in reporting urinary cytology: A pilot study. Cancer Cytopathology. 2022;130(11):872-80. https://doi.org/10.1002/cncy.22615

144. Liu W-C, Lin C, Lin C-S, Tsai M-C, Chen S-J, Tsai S-H, et al. An Artificial Intelligence-Based Alarm Strategy Facilitates Management of Acute Myocardial Infarction. Journal of Personalized Medicine. 2021;11(11). https://doi.org/10.3390/jpm11111149

145. Hwang J, Lee T, Lee H, Byun S. A Clinical Decision Support System for Sleep Staging Tasks With Explanations From Artificial Intelligence: User-Centered Design and Evaluation Study. J Med Internet Res. 2022;24(1):e28659. https://doi.org/10.2196/28659

146. Alrajhi AA, Alswailem OA, Wali G, Alnafee K, AlGhamdi S, Alarifi J, et al. Data-Driven Prediction for COVID-19 Severity in Hospitalized Patients. International Journal of Environmental Research and Public Health. 2022;19(5):2958. https://doi.org/10.3390/ijerph19052958

147. Li Y-Y, Wang J-J, Huang S-H, Kuo C-L, Chen J-Y, Liu C-F, et al. Implementation of a machine learning application in preoperative risk assessment for hip repair surgery. BMC Anesthesiology. 2022;22(1):116. https://doi.org/10.1186/s12871-022-01648-y

148. Lipatov K, Daniels CE, Park JG, Elmer J, Hanson AC, Madsen BE, et al. Implementation and evaluation of sepsis surveillance and decision support in medical ICU and emergency department. The American Journal of Emergency Medicine. 2022;51:378-83. https://doi.org/10.1016/j.ajem.2021.09.086

149. Ivanov O, Wolf L, Brecher D, Lewis E, Masek K, Montgomery K, et al. Improving ED Emergency Severity Index Acuity Assignment Using Machine Learning and Clinical Natural Language Processing. Journal of Emergency Nursing. 2021;47(2):265-78.e7. https://doi.org/10.1016/j.jen.2020.11.001

150. Garzon-Chavez D, Romero-Alvarez D, Bonifaz M, Gaviria J, Mero D, Gunsha N, et al. Adapting for the COVID-19 pandemic in Ecuador, a characterization of hospital strategies and patients. PLOS ONE. 2021;16(5):e0251295. https://doi.org/10.1371/journal.pone.0251295

151. Hinson JS, Klein E, Smith A, Toerper M, Dungarani T, Hager D, et al. Multisite implementation of a workflow-integrated machine learning system to optimize COVID-19 hospital admission decisions. npj Digital Medicine. 2022;5(1):94. https://doi.org/10.1038/s41746-022-00646-1

152. Wang Y-C, Chen K-W, Tsai B-Y, Wu M-Y, Hsieh P-H, Wei J-T, et al. Implementation of an All-Day Artificial Intelligence-Based Triage System to Accelerate Door-to-Balloon Times. Mayo Clinic Proceedings. 2022;97(12):2291-303. https://doi.org/10.1016/j.mayocp.2022.05.014

153. Jordan M, Hauser J, Cota S, Li H, Wolf L. The Impact of Cultural Embeddedness on the Implementation of an Artificial Intelligence Program at Triage: A Qualitative Study. Journal of Transcultural Nursing. 2023;34(1):32-9. https://doi.org/10.1177/10436596221129226

154. Wu L, He X, Liu M, Xie H, An P, Zhang J, et al. Evaluation of the effects of an artificial intelligence system on endoscopy quality and preliminary testing of its performance in detecting early gastric cancer: a randomized controlled trial. Endoscopy. 2021;53(12):1199-207. https://doi.org/10.1055/a-1350-5583

155. Byun HK, Chang JS, Choi MS, Chun J, Jung J, Jeong C, et al. Evaluation of deep learning-based autosegmentation in breast cancer radiotherapy. Radiation Oncology. 2021;16(1):203. https://doi.org/10.1186/s13014-021-01923-1

156. Cha E, Elguindi S, Onochie I, Gorovets D, Deasy JO, Zelefsky M, et al. Clinical implementation of deep learning contour autosegmentation for prostate radiotherapy. Radiotherapy and Oncology. 2021;159:1-7. https://doi.org/10.1016/j.radonc.2021.02.040

157. Liu Y, Cheng L. Ultrasound Images Guided under Deep Learning in the Anesthesia Effect of the Regional Nerve Block on Scapular Fracture Surgery. Journal of Healthcare Engineering. 2021;2021:6231116. https://doi.org/10.1155/2021/6231116

158. Choudhury A. Factors influencing clinicians' willingness to use an AI-based clinical decision support system. Frontiers in Digital Health. 2022;4. https://doi.org/10.3389/fdgth.2022.920662

159. Kneepkens E, Bakx N, van der Sangen M, Theuws J, van der Toorn P-P, Rijkaart D, et al. Clinical evaluation of two AI models for automated breast cancer plan generation. Radiation Oncology. 2022;17(1):25. https://doi.org/10.1186/s13014-022-01993-9

160. Winslow CJ, Edelson DP, Churpek MM, Taneja M, Shah NS, Datta A, et al. The Impact of a Machine Learning Early Warning Score on Hospital Mortality: A Multicenter Clinical Intervention Trial. Critical Care Medicine. 2022;50(9). https://doi.org/10.1097/ccm.0000000000005492

161. Schwartz JM, George M, Rossetti SC, Dykes PC, Minshall SR, Lucas E, et al. Factors Influencing Clinician Trust in Predictive Clinical Decision Support Systems for In-Hospital Deterioration: Qualitative Descriptive Study. JMIR Hum Factors. 2022;9(2):e33960. https://doi.org/10.2196/33960

162. Park C, Loza-Avalos SE, Harvey J, Hirschkorn C, Dultz LA, Dumas RP, et al. A Real-Time Automated Machine Learning Algorithm for Predicting Mortality in Trauma Patients: Survey Says it's Ready for Prime-Time. The American Surgeon™. 2023;90(4):655-61. https://doi.org/10.1177/00031348231207299

163. Kermani F, Zarkesh MR, Vaziri M, Sheikhtaheri A. A case-based reasoning system for neonatal survival and LOS prediction in neonatal intensive care units: a development and validation study. Scientific Reports. 2023;13(1):8421. https://doi.org/10.1038/s41598-023-35333-y

164. Wilimitis D, Turer RW, Ripperger M, McCoy AB, Sperry SH, Fielstein EM, et al. Integration of Face-to-Face Screening With Real-time Machine Learning to Predict Risk of Suicide Among Adults. JAMA Network Open. 2022;5(5):e2212095-e. https://doi.org/10.1001/jamanetworkopen.2022.12095

165. Chen X, Huang X, Yin M. Implementation of Hospital-to-Home Model for Nutritional Nursing Management of Patients with Chronic Kidney Disease Using Artificial Intelligence Algorithm

Combined with CT Internet +. Contrast Media & Molecular Imaging. 2022;2022:1183988. https://doi.org/10.1155/2022/1183988

166. Maeda Y, Kudo S-e, Ogata N, Misawa M, Iacucci M, Homma M, et al. Evaluation in real-time use of artificial intelligence during colonoscopy to predict relapse of ulcerative colitis: a prospective study. Gastrointestinal Endoscopy. 2022;95(4):747-56.e2. https://doi.org/10.1016/j.gie.2021.10.019

167. Zhou S, Ma X, Jiang S, Huang X, You Y, Shang H, et al. A retrospective study on the effectiveness of Artificial Intelligence-based Clinical Decision Support System (AI-CDSS) to improve the incidence of hospital-related venous thromboembolism (VTE). Annals of Translational Medicine. 2021;9(6):491. https://doi.org/10.21037/atm-21-1093

168. Xu L, He X, Zhou J, Zhang J, Mao X, Ye G, et al. Artificial intelligence-assisted colonoscopy: A prospective, multicenter, randomized controlled trial of polyp detection. Cancer Medicine. 2021;10(20):7184-93. https://doi.org/10.1002/cam4.4261

169. Zhang B, Jia C, Wu R, Lv B, Li B, Li F, et al. Improving rib fracture detection accuracy and reading efficiency with deep learning-based detection software: a clinical evaluation. British Journal of Radiology. 2021;94(1118):20200870. https://doi.org/10.1259/bjr.20200870

170. Glissen Brown JR, Mansour NM, Wang P, Chuchuca MA, Minchenberg SB, Chandnani M, et al. Deep Learning Computer-aided Polyp Detection Reduces Adenoma Miss Rate: A United States Multi-center Randomized Tandem Colonoscopy Study (CADeT-CS Trial). Clinical Gastroenterology and Hepatology. 2022;20(7):1499-507.e4. https://doi.org/10.1016/j.cgh.2021.09.009

171. Kamba S, Tamai N, Saitoh I, Matsui H, Horiuchi H, Kobayashi M, et al. Reducing adenoma miss rate of colonoscopy assisted by artificial intelligence: a multicenter randomized controlled trial. Journal of Gastroenterology. 2021;56(8):746-57. https://doi.org/10.1007/s00535-021-01808-w

172. Liu X, Wu D, Xie H, Xu Y, Liu L, Tao X, et al. Clinical evaluation of AI software for rib fracture detection and its impact on junior radiologist performance. Acta Radiologica. 2021;63(11):1535-45. https://doi.org/10.1177/02841851211043839

173. Adams R, Henry KE, Sridharan A, Soleimani H, Zhan A, Rawat N, et al. Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis. Nature Medicine. 2022;28(7):1455-60. https://doi.org/10.1038/s41591-022-01894-0

174. Chien H-WC, Yang T-L, Juang W-C, Chen Y-YA, Li Y-CJ, Chen C-Y. Pilot Report for Intracranial Hemorrhage Detection with Deep Learning Implanted Head Computed Tomography Images at Emergency Department. Journal of Medical Systems. 2022;46(7):49. https://doi.org/10.1007/s10916-022-01833-z

175. Boussina A, Shashikumar S, Amrollahi F, Pour H, Hogarth M, Nemati S, editors. Development & Deployment of a Real-time Healthcare Predictive Analytics Platform. Annual International Conference of the IEEE Engineering in Medicine & Biology Society; 2023; 24-27 July 2023. https://doi.org/10.1109/EMBC40787.2023.10340351

176. Madsen M, Gregor SD. Measuring Human-Computer Trust. 2000.

177. Seyam M, Weikert T, Sauter A, Brehm A, Psychogios M-N, Blackham KA. Utilization of Artificial Intelligence–based Intracranial Hemorrhage Detection on Emergent Noncontrast CT Images in Clinical Workflow. Radiology: Artificial Intelligence. 2022;4(2):e210168. https://doi.org/10.1148/ryai.210168

178. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform. 2010;43(1):159-72. https://doi.org/10.1016/j.jbi.2009.07.002

179. Duan X, Su D, Yu H, Xin W, Wang Y, Zhao Z. Adoption of Artificial Intelligence (AI)-Based Computerized Tomography (CT) Evaluation of Comprehensive Nursing in the Operation Room in Laparoscopy-Guided Radical Surgery of Colon Cancer. Computational Intelligence and Neuroscience. 2022;2022:1-11. https://doi.org/10.1155/2022/2180788

180. Quan SY, Wei MT, Lee J, Mohi-Ud-Din R, Mostaghim R, Sachdev R, et al. Clinical evaluation of a real-time artificial intelligence-based polyp detection system: a US multi-center pilot study. Scientific Reports. 2022;12(1):6598. https://doi.org/10.1038/s41598-022-10597-y

181. Wong J, Huang V, Wells D, Giambattista J, Giambattista J, Kolbeck C, et al. Implementation of deep learning-based auto-segmentation for radiotherapy planning structures: a workflow study at two cancer centers. Radiation Oncology. 2021;16(1):101. https://doi.org/10.1186/s13014-021-01831-4

182. Kotovich D, Twig G, Itsekson-Hayosh Z, Klug M, Simon AB, Yaniv G, et al. The impact on clinical outcomes after 1 year of implementation of an artificial intelligence solution for the detection of intracranial hemorrhage. International Journal of Emergency Medicine. 2023;16(1):50. https://doi.org/10.1186/s12245-023-00523-y

183. Lucas E, David D, III, Christopher N, Andrei A, Violiza I-A, Adam SA, et al. Automated emergent large vessel occlusion detection by artificial intelligence improves stroke workflow in a hub and spoke stroke system of care. Journal of NeuroInterventional Surgery. 2022;14(7):704. https://doi.org/10.1136/neurintsurg-2021-017714

184. Wang D, Jin R, Shieh C-C, Ng AY, Pham H, Dugal T, et al. Real world validation of an AI-based CT hemorrhage detection tool. Frontiers in Neurology. 2023;14. https://doi.org/10.3389/fneur.2023.1177723

185. Brooke J. SUS: A quick and dirty usability scale. Usability Eval Ind. 1995;189. https://doi.org/10.1201/9781498710411-35

186. Paas FGWC. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. 84. US: American Psychological Association; 1992. p. 429-34. https://doi.org/10.1037/0022-0663.84.4.429

187. Committee on Patient Safety and Health Information Technology; Institute of Medicine. Health IT and Patient Safety: Building Safer Systems for Better Care. 2012. https://doi.org/10.17226/13269

188. Kim MO, Coiera E, Magrabi F. Problems with health information technology and their effects on care delivery and patient outcomes: a systematic review. J Am Med Inform Assoc. 2017;24(2):246-50. https://doi.org/10.1093/jamia/ocw154

189. Amodei D, Olah C, Steinhardt J, Christiano P, Schulman J, Mané D. Concrete problems in AI safety. arXiv preprint. 2016. https://doi.org/10.48550/arXiv.1606.06565

190. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Qual Saf. 2019;28(3):231-7. https://doi.org/10.1136/bmjqs-2018-008370

191. Lyell D, Wang Y, Coiera E, Magrabi F. More than algorithms: an analysis of safety events involving ML-enabled medical devices reported to the FDA. J Am Med Inform Assoc. 2023;30(7):1227-36. https://doi.org/10.1093/jamia/ocad065

192. U.S. Food & Drug Administration. Intended Use of Imaging Software for Intracranial Large Vessel Occlusion—Letter to Health Care Providers. 2022. Available from: https://www.fda.gov/medical-devices/letters-health-care-providers/intended-use-imaging-software-intracranial-large-vessel-occlusion-letter-health-care-providers#publication Accessed March 2024. [Accessed March 2024]

193. Daneshjou R, Vodrahalli K, Novoa RA, Jenkins M, Liang W, Rotemberg V, et al. Disparities in dermatology AI performance on a diverse, curated clinical image set. Science Advances. 2022;8(32):eabq6147. https://doi.org/10.1126/sciadv.abq6147

194. Wong A, Cao J, Lyons PG, Dutta S, Major VJ, Ötles E, et al. Quantification of Sepsis Model Alerts in 24 US Hospitals Before and During the COVID-19 Pandemic. JAMA Netw Open. 2021;4(11):e2135286. https://doi.org/10.1001/jamanetworkopen.2021.35286

195. Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O, et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. JAMA Intern Med. 2021;181(8):1065-70. https://doi.org/10.1001/jamainternmed.2021.2626

196. Beede E, Baylor E, Hersch F, Iurchenko A, Wilcox L, Ruamviboonsuk P, et al. A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems; Honolulu, HI, USA: Association for Computing Machinery; 2020. p. 1–12. https://doi.org/10.1145/3313831.3376718

197. NSW Government. Artificial Intelligence: Have Your Say. 2021. Available from: https://www.haveyoursay.nsw.gov.au/artificial-intelligence. [Accessed March 2024].

198. World Health Organization. Ageism in artificial intelligence for health. WHO. 2022. Available from: https://www.who.int/publications/i/item/9789240040793. [Accessed March 2024].

199. Organisation for Economic Co-operation and Development. Recommendation of the Council on Artificial Intelligence. OECD. 2020. Available from: https://one.oecd.org/document/C/MIN(2019)3/FINAL/en/pdf. [Accessed March 2024].

200. UK Centre for Data Ethics and Innovation. Review into bias in algorithmic decision-making. 2020. Available from: https://www.gov.uk/government/publications/cdei-publishes-review-into-bias-in-algorithmic-decision-making. [Accessed March 2024].

201. Canadian Institute for Health Information. CIHI's Health Data and Information Governance and Capability Framework. 2020. Available from: https://www.cihi.ca/sites/default/files/document/health-data-info-capability-framework-en.pdf. [Accessed March 2024].

202. Australian Commission on Safety and Quality in Health Care. National Safety and Quality Health Service Standards. 2nd ed. - version 2. Sydney: ACSQHC. 2021. Available from: https://www.safetyandquality.gov.au/publications-and-resources/resource-library/national-safety-and-quality-health-service-standards-second-edition. [Accessed March 2024].

203. UK Department of Science Innovation and Technology. UK Artificial Intelligence Regulation Impact Assessment. 2023. Available from: https://assets.publishing.service.gov.uk/media/6424208f3d885d000cdadddf/uk_ai_regulation_impact_assessment.pdf. [Accessed March 2024].

204. UK National Institute for Health and Care Excellence. Evidence standards framework for digital health technologies. 2023. Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies. [Accessed March 2024].

205. US Food and Drug Administration. Artificial Intelligence and Machine Learning Software as a Medical Device Action Plan. 2021. Available from: https://www.fda.gov/media/145022/download?attachment. [Accessed March 2024].

206. US Congress. Artificial Intelligence: Overview, Recent Advances, and Considerations for the 118th Congress. 2023. Available from: https://crsreports.congress.gov/product/pdf/R/R47644. [Accessed March 2024].

207. US White House. Blueprint for an AI Bill of Rights. 2022. Available from: https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf. [Accessed March 2024].

208. UK Government. National AI Strategy. 2021. Available from: https://www.gov.uk/government/publications/national-ai-strategy. [Accessed March 2024].

209. UK Department of Health and Social Care. The future of healthcare: Our vision for digital, data and technology in health and care. 2018. Available from: https://www.gov.uk/government/publications/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care. [Accessed March 2024].

210. European Union. The impact of the General Data Protection Regulation (GDPR) on artificial intelligence. 2020. Available from: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf. [Accessed March 2024].

211. 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. [Accessed March 2024]

212. Yahav-Dovrat A, Saban M, Merhav G, Lankri I, Abergel E, Eran A, et al. Evaluation of Artificial Intelligence–Powered Identification of Large-Vessel Occlusions in a Comprehensive Stroke Center. American Journal of Neuroradiology. 2021;42(2):247. https://doi.org/10.3174/ajnr.A6923

213. Zia A, Fletcher C, Bigwood S, Ratnakanthan P, Seah J, Lee R, et al. Retrospective analysis and prospective validation of an AI-based software for intracranial haemorrhage detection at a high-volume trauma centre. Scientific Reports. 2022;12(1):19885. https://doi.org/10.1038/s41598-022-24504-y

Glossary

Terminology / Abbreviation	Definition
Acute care setting	Acute care hospital – an establishment that provides care in which the intent is to perform surgery, diagnostic or therapeutic procedures in the treatment of illness or injury
AI	Artificial Intelligence
EHR/EMR	Electronic health record /electronic medical record
False Positive	The number of samples that were incorrectly identified as positive by the algorithm
False Negative	The number of samples that were incorrectly identified as negative by the algorithm
ML	Machine Learning
PACS	Picture Archiving and Communications System
RIS	Radiology Information System
Regulatory sandbox	A regulatory sandbox for Artificial Intelligence is a controlled and supervised environment where developers and innovators, under the guardianship of the governmental authorities, can test and deploy AI systems in real-world scenarios, with some regulatory flexibility.
Reinforcement learning	Machine learning by policy that maximises the cumulative reward over time by trial and error.
SaMD	Software as a Medical Device
Supervised machine learning	Ground truth labelled outcome training data
True positive	The number of positive samples that have been correctly identified by the algorithm
True negative	The number of samples that were accurately identified as negative by the algorithm
Unsupervised machine learning	Ground truth not provided

Appendices

Appendix	Document
A	Example of an impact assessment tool
B	List of international legislation and policy reviewed
C	List of Australian legislation and policy reviewed
D	Chapter 4 Primary literature review search strategy
E	Chapter 4 PRISMA Flow chart
F	Chapter 4 Summary table of studies about AI in acute settings included in report
G	Chapter 5 Studies reporting the effects of AI problems on care delivery and patient outcomes

Appendix A: Impact assessment level

Adopted from Canada's "Directive on Automated Decision Making" (34).

Level	Description
I	The decision will likely have little to no impact on: • the rights of individuals or communities; • the equality, dignity, privacy, and autonomy of individuals; • the health or well-being of individuals or communities; • the economic interests of individuals, entities, or communities; • the ongoing sustainability of an ecosystem. Level I decisions will often lead to impacts that are reversible and brief.
II	The decision will likely have moderate impacts on: • the rights of individuals or communities; • the equality, dignity, privacy, and autonomy of individuals; • the health or well-being of individuals or communities; • the economic interests of individuals, entities, or communities; • the ongoing sustainability of an ecosystem. Level II decisions will often lead to impacts that are likely reversible and short-term.
III	The decision will likely have high impacts on: • the rights of individuals or communities; • the equality, dignity, privacy, and autonomy of individuals; • the health or well-being of individuals or communities; • the economic interests of individuals, entities, or communities; • the ongoing sustainability of an ecosystem. Level III decisions will often lead to impacts that can be difficult to reverse and are ongoing.
IV	The decision will likely have very high impacts on: • the rights of individuals or communities;

Level	Description
	• the equality, dignity, privacy, and autonomy of individuals; • the health or well-being of individuals or communities; • the economic interests of individuals, entities, or communities; • the ongoing sustainability of an ecosystem. Level IV decisions will often lead to impacts that are irreversible and perpetual.

Appendix B: List of reviewed documents from international jurisdictions

Reference no.	Title	Authoring agency/ organisation	Jurisdiction	Year	Relevance to acute care
(205)	Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan	US Food and Drug Administration (FDA)	US	2021	Healthcare-specific
(16)	Bill S1402 An Act concerning discrimination and automated decision systems and supplementing P.L.1945, c.169 (C.10:5-1 et seq.).	New Jersey State Legislature	US	2022	Sector-agnostic
(17)	Stop Discrimination by Algorithms Act of 2023	Council of the District of Columbia	US	2023	Sector-agnostic
(18)	Digital Fairness Act	State of New York	US	2023	Sector-agnostic
(20)	Executive Order: Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence	US White House	US	2023	Sector-agnostic
(206)	Artificial Intelligence: Overview, Recent Advances, and Considerations for the 118th Congress	US Congressional Research Service	US	2023	Sector-agnostic
(24)	Nondiscrimination in Health Programs and Activities (Section 1557)	US Department of Health and Human Services (HHS) Office of Civil Rights (OCR)	US	2022	Healthcare-specific
(26)	Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing	US Department of Health and Human Services (HHS)	US	2024	Healthcare-specific
(19)	US National AI Initiatives Act	US Office of Science and Technology Policy (OSTP)	US	2021 (NAII Act)	Sector-agnostic
(207)	Blueprint for an AI bill of rights making automated systems work for the American people	The White House Office of Science and Technology Policy	US	2022	Sector-agnostic
(42)	Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities	US Government Accountability Office (GAO)	US	2021	Sector-agnostic

References

Reference no.	Title	Authoring agency/ organisation	Jurisdiction	Year	Relevance to acute care
(1)	Artificial Intelligence (AI) Strategy	US Department of Health and Human Services	US	2021	Healthcare-specific
(21)	OMB M-21-06 Memorandum for the heads of executive departments and agencies (Guidance for Regulation of Artificial Intelligence Applications)	Office of Management and Budget	US	2020	Sector-agnostic
(40)	Trustworthy AI Playbook	US Department of Health and Human Services	US	2021	Healthcare-specific
(38)	Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector	Alan Turing Institute, UK's national institute for data science and artificial intelligence	UK	2019	Sector-agnostic
(39)	A Buyer's Guide to AI in Health and Care	NHS	UK	2020	Healthcare-specific
(31)	Artificial Intelligence: How to get it right	NHS	UK	2019	Healthcare-specific
(27)	Artificial Intelligence (Regulation) Bill [HL]: A Bill to make provision for the regulation of artificial intelligence; and for connected purposes	House of Lords	UK	2023	Sector-agnostic
(204)	Evidence standards framework for digital health technologies	National Institute for Health and Care Excellence (NICE)	UK	2023	Healthcare-specific
(203)	UK Artificial Intelligence Regulation Impact Assessment	Department of Science, Innovation & Technology	UK	2023	Sector-agnostic
(44)	A pro-innovation approach to AI regulation: government response	Department of Science, Innovation & Technology	UK	2024	Sector-agnostic
(200)	Review into bias in algorithmic decision-making	Centre for Data Ethics and Innovation	UK	2020	Sector-agnostic
(208)	National AI Strategy	Department of Digital, Culture, Media and Sport	UK	2021	Sector-agnostic
(209)	The future of healthcare: Our vision for digital, data and technology in health and care	Department of Health and Social Care, NHS	UK	2018	Healthcare-specific

Reference no.	Title	Authoring agency/ organisation	Jurisdiction	Year	Relevance to acute care
(33)	Using machine learning in diagnostic services	Care Quality Commission	UK	2022	Healthcare-specific
(45)	The Bletchley Declaration by Countries Attending the AI Safety Summit	UK Government	UK and international	2023	Sector-agnostic
(29)	NZ Algorithm Charter for Aotearoa New Zealand	NZ government	New Zealand	2020	Sector-agnostic
(47)	Reimagining Regulation for the Age of AI: New Zealand Pilot Project	World Economic Forum and New Zealand Government	New Zealand	2020	Sector-agnostic
(46)	Advice on the use of Large Language Models and Generative AI in Healthcare	Health New Zealand, National Artificial Intelligence (AI) and Algorithm Expert Advisory Group (NAIAEAG)	New Zealand	2023	Healthcare-specific
(30)	Emerging Health Technology: Introductory Guidance	Ministry of Health	New Zealand	2019	Healthcare-specific
(28)	REPORT: Capturing the benefits of AI in healthcare for Aotearoa New Zealand - Key messages	Office of the Prime Minister's Chief Science Advisor	New Zealand	2023	Healthcare-specific
(34)	Directive on Automated Decision-Making	Treasury Board of Canada	Canada	2020	Sector-agnostic
(23)	Bill C27 An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts	House of Commons	Canada	2022	Sector-agnostic
(35)	AI Verify: AI Governance Testing Framework and Toolkit	Personal Data Protection Commission	Singapore	2023	Sector-agnostic
(48)	Model Artificial Intelligence Governance Framework, Second Edition	Infocomm Media Development Authority and Personal Data Protection Commission	Singapore	2020	Sector-agnostic
(49)	Artificial Intelligence in Healthcare	Ministry of Health, the Health Sciences Authority, Integrate Health Information Systems	Singapore	2021	Healthcare-specific

References

Reference no.	Title	Authoring agency/ organisation	Jurisdiction	Year	Relevance to acute care
(37)	Companion to the Model AI Governance Framework	Infocomm Media Development Authority and Personal Data Protection Commission	Singapore	2020	Sector-agnostic
(50)	Ethics of artificial intelligence: Issues and Initiatives	European Parliament	EU	2020	Sector-agnostic
(53)	Framework of ethical aspects of artificial intelligence, robotics and related technologies	European Parliament	EU	2020	Sector-agnostic
(32)	Understanding algorithmic decision-making- Opportunities and challenges	European Parliamentary Research Service (EPRS)	EU	2019	Sector-agnostic
(22)	The EU AI Act (ARTIFICIAL INTELLIGENCE ACT)	European Parliament	EU	2021	Sector-agnostic
(210)	The impact of the General Data Protection Regulation (GDPR) on artificial intelligence	European Parliamentary Research Service (EPRS)	EU	2020	Sector-agnostic
(52)	Motion for a European Parliament Resolution on artificial intelligence in a digital age	European Parliament	EU	2022	Sector-agnostic
(36)	Ethics Guidelines for Trustworthy AI	European Commission	EU	2019	Sector-agnostic
(51)	Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics	European Commission	EU	2020	Sector-agnostic
(198)	Ageism in artificial intelligence for health: WHO policy brief	World Health Organization	International	2022	Healthcare-specific
(54)	Ethics and governance of artificial intelligence for health Guidance on large multi-modal models	World Health Organization	International	2024	Healthcare-specific
(55)	Regulatory considerations on artificial intelligence for health	World Health Organization	International	2023	Healthcare-specific
(56)	Generating evidence for artificial intelligence-based medical devices: a framework for training, validation and evaluation	World Health Organization	International	2021	Healthcare-specific
(57)	Ethics and governance of artificial intelligence for health	World Health Organization	International	2021	Healthcare-specific

References

Reference no.	Title	Authoring agency/ organisation	Jurisdiction	Year	Relevance to acute care
(59)	Framework for the classification of AI systems	Organisation for Economic Co-operation and Development	International	2022	Sector-agnostic
(58)	Collective action for responsible AI in health	Organisation for Economic Co-operation and Development	International	2024	Sector-agnostic
(199)	Recommendation of the Council on Artificial Intelligence	Organisation for Economic Co-operation and Development	International	2019	Sector-agnostic

Appendix C: List of reviewed policy documents from Australia

Ref no.	Title	Authorship organisations	Year	Relevance to acute care
(66)	Artificial Intelligence - Understanding Privacy Obligations	Office of the Victorian Information Commissioner	2021	Sector-agnostic
(95)	Position Statements: Digital Health	Royal Australasian College of Medical Administrators	2020	Healthcare specific
(68)	Examples of regulated and unregulated software (excluded) and software-based medical devices	Therapeutic Goods Administration	2021	Healthcare specific
(90)	Interim Guidance for WA Public Sector Agencies on Adoption of Artificial Intelligence	Department of the Premier and Cabinet Office of Digital Government	2023	Sector-agnostic
(85)	Ethical Policy Statement	Digital.NSW	2022	Sector-agnostic
(71)	Mandatory Ethical Principles for the use of AI	Digital.NSW	2021	Sector-agnostic
(89)	Artificial Intelligence Ethics Policy | Key Considerations	Digital.NSW	2021	Sector-agnostic
(72)	Australia's AI Ethics Principles	Department of Industry, Science and Resources	2019	Sector-agnostic
(15)	Safe and Responsible AI in Australia Consultation - Australian Government's Interim Response	Department of Industry, Science and Resources	2024	Sector-agnostic
(78)	Regulation of Software-Based Medical Devices	Therapeutic Goods Administration	2023	Healthcare specific
(88)	Artificial Intelligence: Guidance for Clinical Imaging and Therapeutic Radiography Professionals, a summary by the Society of Radiographers AI working group	AHPRA Medical Radiation Practice Board	2022	Healthcare specific
(70)	Artificial Intelligence - Australia's Ethics Framework	Data61	2019	Sector-agnostic
(86)	Artificial Intelligence Policy (Position)	Australian Government Architecture	2023	Sector-agnostic
(80)	Health Service Use of Unregulated Artificial Intelligence (AI) - Health Service Advisory	Victoria State Government Department of Health	2023	Healthcare specific
(92)	Digital Health: Breaking Barriers to Deliver Better Patient Outcomes	Medical Technology Association of Australia	2023	Healthcare specific

References

Ref no.	Title	Authorship organisations	Year	Relevance to acute care
(41)	Automated Decision-Making - Better Practice Guide	Commonwealth Ombudsman	2019	Sector-agnostic
(4)	Artificial Intelligence in Healthcare - Position Statement	Australian Medical Association	2023	Healthcare specific
(67)	Artificial Intelligence Assurance Framework	NSW Government	2022	Sector-agnostic
(81)	A National Policy Roadmap for Artificial Intelligence in Healthcare	Australian Alliance for Artificial Intelligence in Healthcare	2023	Healthcare specific
(79)	Position Statement: Use of Artificial Intelligence in Dermatology in Australia	Australasian College of Dermatologists	2022	Healthcare specific
(69)	Human Rights and Technology: Final Report	Australian Human Rights Commission	2021	Sector-agnostic
(25)	Submission to the inquiry into artificial intelligence in New South Wales	Australian Academy of Technological Sciences and Engineering	2023	Sector-agnostic
(91)	Use of Generative AI for Government - Information Sheet	Queensland Government	2023	Sector-agnostic
(64)	The State of AI Governance in Australia	The Human Technology Institute (Solomon & Davis)	2023	Sector-agnostic
(75)	Automated Decision-Making in NSW	ADM+S (Weatherall et al.)	2024	Sector-agnostic
(2)	Safe and Responsible AI in Australia - Discussion Paper	Department of Industry, Science and Resources	2023	Sector-agnostic
(76)	Clinical decision support software - Scope and Examples	Therapeutic Goods Administration	2021	Healthcare specific
(77)	Digital mental health: Software based medical devices	Therapeutic Goods Administration	2022	Healthcare specific
(96)	Medical device cyber security information for users	Therapeutic Goods Administration	2022	Healthcare specific
(211)	A Roadmap for Artificial Intelligence in Healthcare in Australia	Australian Alliance for Artificial Intelligence in Healthcare	2021	Healthcare specific
(74)	Interim Guidance on government use of public generative AI tools - November 2023	Australian Government Architecture	2023	Sector-agnostic

References

Ref no.	Title	Authorship organisations	Year	Relevance to acute care
(87)	Adoption of Artificial Intelligence in the Public Sector	Australian Government Architecture	2022	Sector-agnostic
(3)	Ethical Principles for AI in Medicine	Royal Australian and New Zealand College of Radiologists	2023	Healthcare specific
(82)	Standards of practice for Clinical Radiology - Chapter 9: Artificial Intelligence	Royal Australian and New Zealand College of Radiologists	2020	Healthcare specific
(83)	Artificial Intelligence Strategy	Digital.NSW	2021	Sector-agnostic
(94)	Generative AI: basic guidance	NSW Government	2023	Sector-agnostic
(93)	Using public generative artificial intelligence (AI) tools safely	NSW Government	2023	Sector-agnostic

Appendix D: Primary literature review search strategy

ACQSHC – AI in Acute Care 2024
Medline Search 11/01/2024
Total: 2259 articles
https://simsrad.net.ocs.mq.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&NEWS=N&PAGE=main&SHAREDSEARCHID=7CeFEdtjb0os4pHoUzQ9WBeCE0xPeUkC6TTMY3E5AW5qYhxmDgdaLXZvg8thjlU9c

1.	exp artificial intelligence/ or exp deep learning/ or exp machine learning/ or (AI or "artificial intelligence" or "classification algorithm*" or "computer heuristic*" or "decision support system*" or "decision tree" or "deep learning" or "data science" or "feature detection" or "generative pre-trained transformer" or "language learning model*" or "large language model*" or "learning algorithm*" or "machine learning" or (Markov adj3 model*) or ((multifactor* or multicriteria) adj3 ("decision analysis" or "decision making")) or "natural language process*" or "nearest neighbor*" or "neural network*" or "outlier detection" or "pattern recognition" or "random forest" or "representation learning" or "support vector machine*" or "transfer learning" or "Bing chat" or ChatGPT* or "Chat GPT" or "Google* Bard" or "IBM Watson" or "Microsoft* Bing" or OpenAI or "Open AI" or PathAI or "Path AI").mp.
2.	((("artificial intelligence" or AI) adj2 generat*) or GenAI or ((large or natural or generative or machine or deep learning) adj3 (language or text) adj3 model*) or AlexaTM or (Amazon* and Alexa) or Anthropic or Bard or Bardeen or BERT or "Bing chat" or BioGPT or BLOOM or BloombergGPT or Cerebras-GPT or ChatGPT* or "Chat GPT" or chatbot* or Chatsonic or Chinchilla or Claude or DALL-E or EinsteinGPT or Ernie or Falcon or Galactica or "Generative Fill" or "GitHub Copilot" or GLaM or "Google* Assistant" or "Google* Bard" or Gopher or GPT-1 or GPT-2 or GPT-3* or GPT-4* or GPTNeo or GPT-NEoX or GPT-J* or "IBM Watson" or LaMDA or LLaMA or "Megatron-Turing NLG" or "Microsoft* Bing" or Midjourney or Minerva or NeevaAI or Nvidia or OpenAI or "Open AI" or OpenAssistant or OPT or PaLM or PanGu-E or PathAI or "Path AI" or Perplexity or "pre-trained transformer*" or "pretrained transformer*" or (Apple* and Siri) or SlackGPT or StyleGAN or Synthesia or XLNet or YaLM 100B or YouChat).mp.
3.	or/1-2
4.	critical care/ or early goal-directed therapy/ or hospitalization/ or intensive care, neonatal/ or life support care/ or advanced cardiac life support/ or advanced trauma life support care/ or perioperative care/ or postoperative care/ or preoperative care/
5.	intensive care units/ or burn units/ or coronary care units/ or intensive care units, pediatric/ or intensive care units, neonatal/ or recovery room/ or respiratory care units/
6.	((intensive care or burn or coronary care or NICU or neonatal intensive care or pediatric intensive care or medical assessment or respiratory care) adj1 unit*).ti,ab.
7.	(recovery room or hospital care or acute care).ti,ab.
8.	or/4-7
9.	7 not 8
10.	limit 9 to english language
11.	exp animals/ not humans.sh.
12.	10 not 11
13.	limit 12 to yr="2021 -Current"
14.	limit 13 to (editorial or letter or "review" or "systematic review")
15.	13 not 14

References

Embase Search 11/01/2024
Total: 4 articles (all non-relevant)
https://simsrad.net.ocs.mq.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&NEWS=N&PAGE=main&SHAREDSEARCHID=12aOK6YYCJewERVVnKjMolvO4tmGRW5GwKYdDfyWWEI0d4j47tKZVv74tBx2HxoNc

1.	exp artificial intelligence/ or exp deep learning/ or exp machine learning/ or (AI or "artificial intelligence" or "classification algorithm*" or "computer heuristic*" or "decision support system*" or "decision tree" or "deep learning" or "data science" or "feature detection" or "generative pre-trained transformer" or "language learning model*" or "large language model*" or "learning algorithm*" or "machine learning" or (Markov adj3 model*) or ((multifactor* or multicriteria) adj3 ("decision analysis" or "decision making")) or "natural language process*" or "nearest neighbor*" or "neural network*" or "outlier detection" or "pattern recognition" or "random forest" or "representation learning" or "support vector machine*" or "transfer learning" or "Bing chat" or ChatGPT* or "Chat GPT" or "Google* Bard" or "IBM Watson" or "Microsoft* Bing" or OpenAI or "Open AI" or PathAI or "Path AI").mp.
2.	((("artificial intelligence" or AI) adj2 generat*) or GenAI or ((large or natural or generative or machine or deep learning) adj3 (language or text) adj3 model*) or AlexaTM or (Amazon* and Alexa) or Anthropic or Bard or Bardeen or BERT or "Bing chat" or BioGPT or BLOOM or BloombergGPT or Cerebras-GPT or ChatGPT* or "Chat GPT" or chatbot* or Chatsonic or Chinchilla or Claude or DALL-E or EinsteinGPT or Ernie or Falcon or Galactica or "Generative Fill" or "GitHub Copilot" or GLaM or "Google* Assistant" or "Google* Bard" or Gopher or GPT-1 or GPT-2 or GPT-3* or GPT-4* or GPTNeo or GPT-NEoX or GPT-J* or "IBM Watson" or LaMDA or LLaMA or "Megatron-Turing NLG" or "Microsoft* Bing" or Midjourney or Minerva or NeevaAI or Nvidia or OpenAI or "Open AI" or OpenAssistant or OPT or PaLM or PanGu-E or PathAI or "Path AI" or Perplexity or "pre-trained transformer*" or "pretrained transformer*" or (Apple* and Siri) or SlackGPT or StyleGAN or Synthesia or XLNet or YaLM 100B or YouChat).mp.
3.	or/1-2
4.	emergency care/
5.	intensive care/
6.	advanced trauma life support/
7.	((intensive care or burn or coronary care or NICU or neonatal intensive care or pediatric intensive care or medical assessment or respiratory care) adj1 unit*).ti,ab.
8.	(recovery room or emergency care or intensive care).ti,ab.
9.	or/4-8
10.	and/3,9
11.	limit 10 to english language
12.	exp animals/ not humans.sh.
13.	11 not 12
14.	limit 13 to yr="2021 -Current"
15.	limit 13 to (editorial or letter or "review" or "systematic review")
16.	14 not 15

References

WOS Search 11/01/2024
Total: 706 articles
https://www.webofscience.com/wos/alldb/summary/993638f0-6bf4-4e05-a49d-fe94ae45e448-c44dd040/relevance/1

PsycInfo Search 11/01/2024
Total: 16 articles
https://simsrad.net.ocs.mq.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&NEWS=N&PAGE=main&SHAREDSEARCHID=44SUT0G9uTz8LMsz8XkwNGPQAeQ2Uqj9zFpOPifkmQSGHPIRAnyrS0lGEUxehK0Jd

Search History (9)

	# ▲	Searches	Results	Type
☐	1	exp artificial intelligence/ or exp deep learning/ or exp machine learning/ or (AI or "artificial intelligence" or "classification algorithm*" or "computer heuristic*" or "decision support system*" or "decision tree" or "deep learning" or "data science" or "feature detection" or "generative pre-trained transformer" or "language learning model*" or "large language model*" or "learning algorithm*" or "machine learning" or (Markov adj3 model*) or ((multifactor* or multicriteria) adj3 ("decision analysis" or "decision making")) or "natural language process*" or "nearest neighbor*" or "neural network*" or "outlier detection" or "pattern recognition" or "random forest" or "representation learning" or "support vector machine*" or "transfer learning" or "Bing chat" or ChatGPT* or "Chat GPT" or "Google* Bard" or "IBM Watson" or "Microsoft* Bing" or OpenAI or "Open AI" or PathAI or "Path AI").mp.	150625	Advanced
☐	2	((("artificial intelligence" or AI) adj2 generat*) or GenAI or ((large or natural or generative or machine or deep learning) adj3 (language or text) adj3 model*) or AlexaTM or (Amazon* and Alexa) or Anthropic or Bard or Bardeen or BERT or "Bing chat" or BioGPT or BLOOM or BloombergGPT or Cerebras-GPT or ChatGPT* or "Chat GPT" or chatbot* or Chatsonic or Chinchilla or Claude or DALL-E or EinsteinGPT or Ernie or Falcon or Galactica or "Generative Fill" or "GitHub Copilot" or GLaM or "Google* Assistant" or "Google* Bard" or Gopher or GPT-1 or GPT-2 or GPT-3* or GPT-4* or GPTNeo or GPT-NEoX or GPT-J* or "IBM Watson" or LaMDA or LLaMA or "Megatron-Turing NLG" or "Microsoft* Bing" or Midjourney or Minerva or NeevaAI or Nvidia or OpenAI or "Open AI" or OpenAssistant or OPT or PaLM or PanGu-E or PathAI or "Path AI" or Perplexity or "pre-trained transformer*" or "pretrained transformer*" or (Apple* and Siri) or SlackGPT or StyleGAN or Synthesia or XLNet or YaLM 100B or YouChat).mp.	10068	Advanced
☐	3	or/1-2	159375	Advanced
☐	4	intensive care/	5491	Advanced
☐	5	3 and 4	92	Advanced
☐	6	limit 5 to (english language and yr="2021 -Current")	21	Advanced
☐	7	limit 6 to (editorial or letter or "review" or "systematic review")	4	Advanced
☐	8	6 not 7	17	Advanced
☐	9	limit 8 to peer reviewed journal	16	Advanced

References

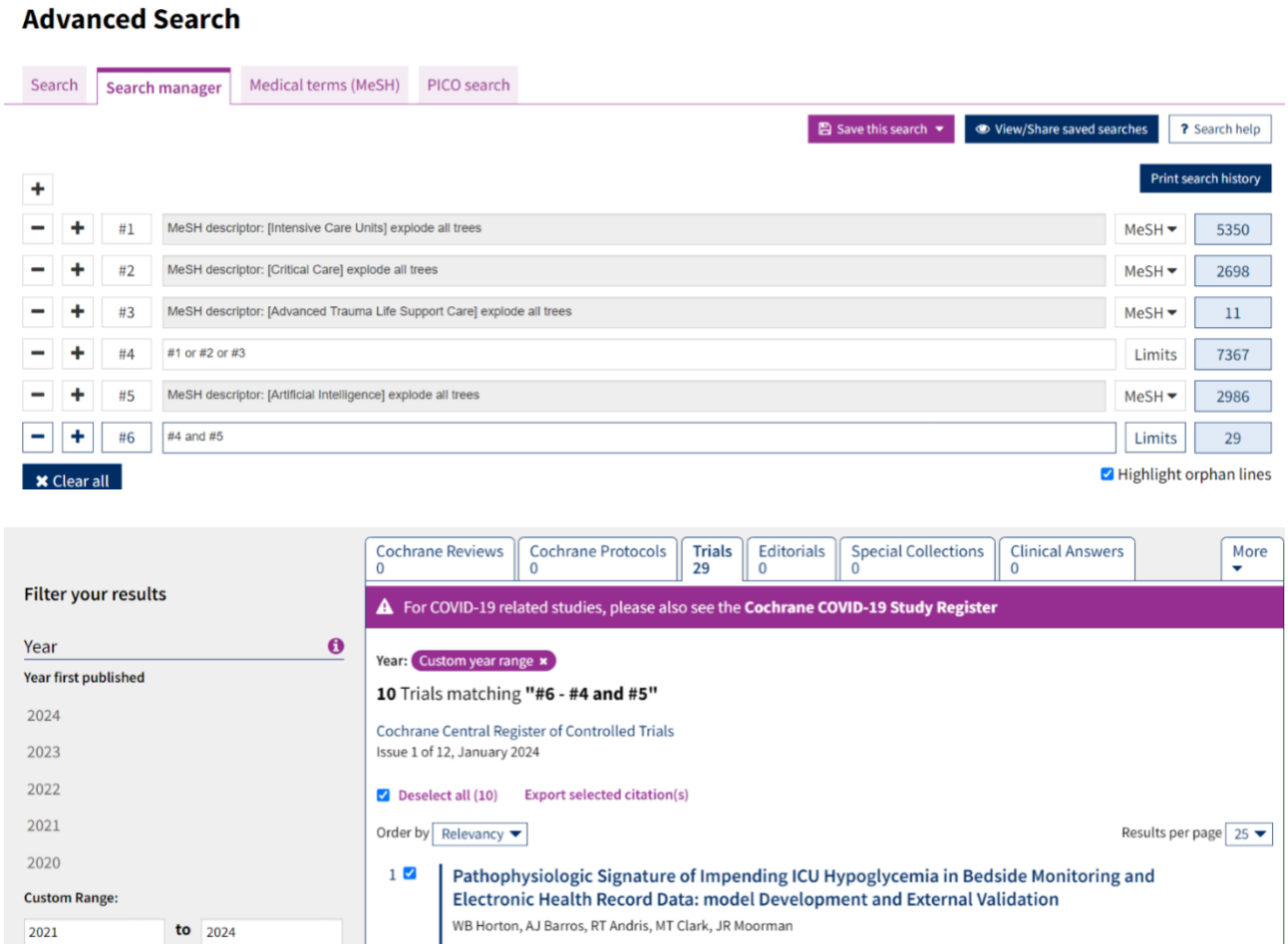
PubMed Search 11/01/2024
Total: 203 articles (114 imported into EndNote after removal of duplicates)
Search Terms: (artificial intelligence OR AI OR generative artificial intelligence OR generative AI OR machine learning OR supervised machine learning OR large language model OR natural language process* OR deep learning OR classification algorithm OR generative pre-trained transformer) AND ((acute care[Title/Abstract] OR emergency care[Title/Abstract] OR intensive care[Title/Abstract] OR preoperative care[Title/Abstract] OR perioperative care[Title/Abstract] OR postoperative care[Title/Abstract]))

CINAHL Search 12/01/2024
Total – 285 articles (221 imported into EndNote after removal of duplicates)
https://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=ccm&bquery=((((MH+%26quot%3bNeural+Networks+(Computer)%26quot%3b)+OR+(MH+%26quot%3bMachine+Learning%2b%26quot%3b)+OR+(MH+%26quot%3bDeep+Learning%26quot%3b)+OR+(MH+%26quot%3bArtificial+Intelligence%2b%26quot%3b))+OR+%26quot%3bS2%26quot%3b)+OR+%26quot%3bS4%26quot%3b+OR+((MH+%26quot%3bDecision+Trees%2b%26quot%3b)))+AND+(((MH+%26quot%3bIntensive+Care+Units%2b%26quot%3b))+OR+(%26quot%3bperioperative+or+peri-operative+or+pre-operative+or+preoperative+or+post-operative+or+postoperative+or+surgical%26quot%3b+OR+(MH+%26quot%3bPerioperative+Care%2b%26quot%3b))+OR+((MH+%26quot%3bLife+Support+Care%2b%26quot%3b)+OR+(MH+%26quot%3bPediatric+Advanced+Life+Support%26quot%3b)+OR+(MH+%26quot%3bAdvanced+Cardiac+Life+Support%2b%26quot%3b)+OR+(MH+%26quot%3bEmergency+Treatment%2b%26quot%3b)))&cli0=DT1&clv0=202101-202412&type=1&searchMode=Standard&site=ehost-live&scope=site&custid=s8434881

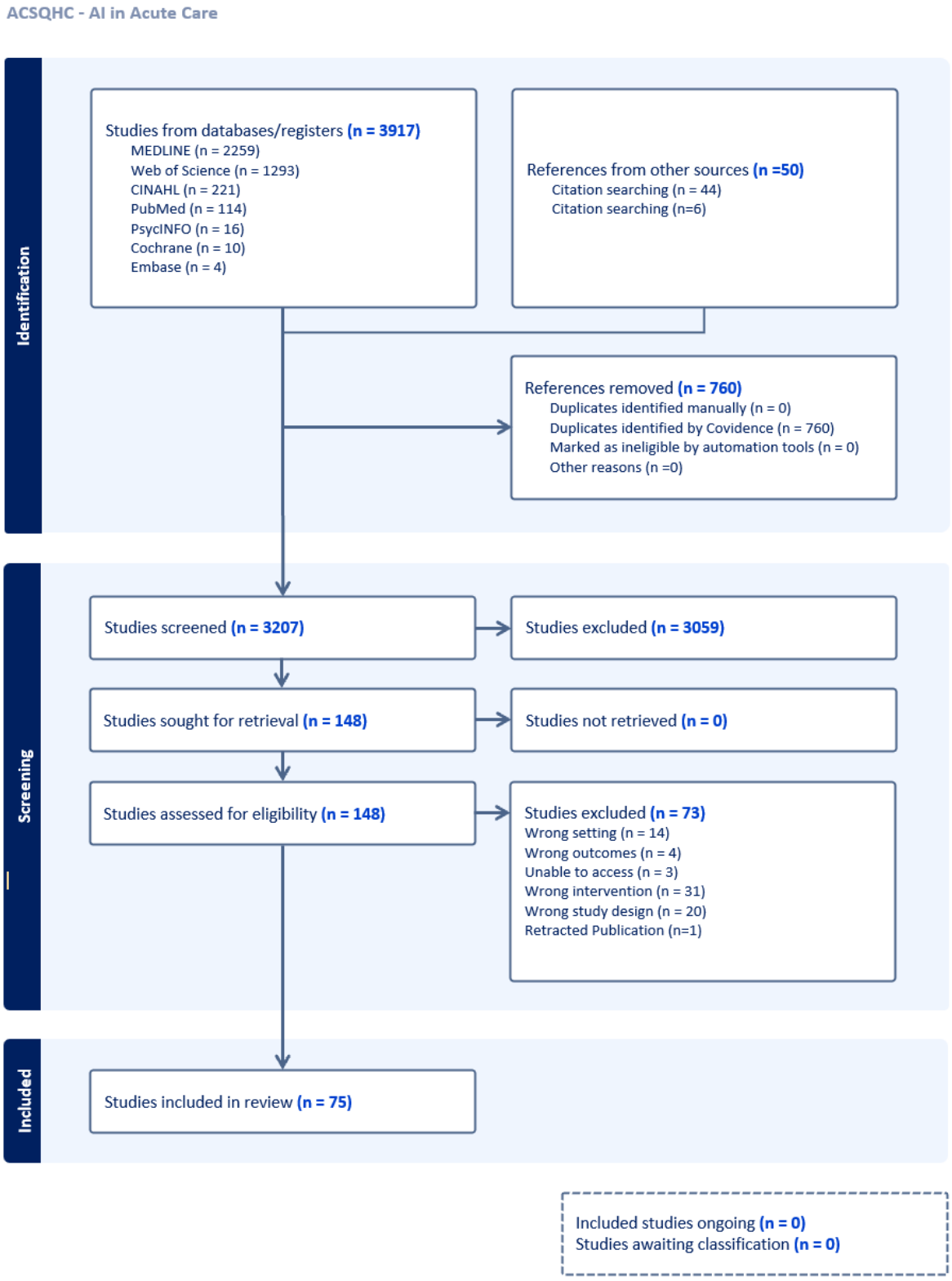
1.	(MH "Neural Networks (Computer)") OR (MH "Machine Learning+") OR (MH "Deep Learning") OR (MH "Artificial Intelligence+")
2.	"artificial intelligence or ai or a.i. or machine learning or deep learning"
3.	S1 OR S2
4.	"generative artificial intelligence or gai or generative ai or chatgpt"
5.	(MH "Decision Trees+")
6.	S3 OR S4 OR S5
7.	(MH "Intensive Care Units+")
8.	"perioperative or peri-operative or pre-operative or preoperative or post-operative or postoperative or surgical" OR (MH "Perioperative Care+")
9.	(MH "Life Support Care+") OR (MH "Pediatric Advanced Life Support") OR (MH "Advanced Cardiac Life Support+") OR (MH "Emergency Treatment+")
10.	S7 OR S8 OR S9
11.	S6 AND S10
12.	S6 AND S10 Limiters - Publication Date: 20210101-20241231
13.	S6 AND S10 Limiters - Publication Date: 20210101-20241231 Expanders - Apply equivalent subjects Narrow by Language: - english Search modes - Boolean/Phrase

References

Cochrane Library Search 15/1/2024
Total: 0 Reviews, 29 trials (10 exported into EndNote after applying date limit 2021 - 2024)



Appendix E: Chapter 4 PRISMA Flowchart



Appendix F: Summary table of studies about AI in acute settings included in report (n=75)

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Cancer (n=17)							
Byun et al. 2021 (155)	South Korea	Cancer: breast	Cancer Centre and Hospital Radiotherapy Department	Procedure	An Auto Contouring System (120) that performs Organ At Risk (OAR) (111) Delineation during 3D CT based planning in breast cancer radiotherapy.	**Observational, multi-centre**. *Description:* 10 cases of women undergoing adjuvant radiotherapy were reviewed by 11 experts. The 11 experts manually delineated the OARs ('manual contouring'). The ACS contoured the same CT scans ('auto-contoured') and then the experts were asked to correct the auto-contoured CT images as needed ('corrected auto-contoured').	OAR volume accuracy was similar between all three respectively based on dice similarity coefficient (0.88 vs 0.90 vs 0.90). Time saving mean manual contouring time 37mins vs 6.4 minutes for corrected auto-contoured CTs. Auto-contoured time was <10mins. Three user satisfaction survey questions all had high mean scores, revealing good user satisfaction.
Cerminara et al. 2023 (133)	Switzerland	Melanoma	Dermatology Department	Diagnosis	Deep learning CNN-based malignancy risk assessment. It categorises each lesion from 0.0 to 10.0 and FotoFinder's Moleanalyzer Pro from 0.0 to 1.0. The higher the score, the higher the risk of malignancy.	**Observational, single centre.** *Description:* Two types of Total Body Photography (TBP) devices utilised in this research: 2D and 3D, each with a dermoscopy camera fitted to take image of skin lesion and CNN embedded. 143 patients with a total 1690 melanocytic skin lesions (mean of 12 per patient) assessed manually (dermatologist alone), then reassessed with the knowledge of the AI risk assessment (Dermatologist +AI).	75 mole excisions occurred, and in those, the sensitivity was 90% for the dermatologist alone, the dermatologist+AI and the 3D TBP CNN. It was 70% 2D TBP CNN. Specificity was highest for dermatologist alone, followed by dermatologist+AI, 3D TBP CNN and lastly 2D TBP CNN at 92%, 86%, 64% and 40% respectively (Ground truth being histopathology result). Total nevi count (mole count) mean was 210 by dermatologists, 469 by 3D TBP CNN, and more than 6.3 times more by 2D (1324).
Cha et al. 2021 (156)	USA	Cancer: prostate	Radiology Department	Procedure	In-house developed, MRI based deep learning auto-segmentation algorithm for both Organs At Risk (OaR) and Clinical Target Volumes (CTV) in short-course prostate radiation therapy.	**Observational, single centre.** *Description:* 173 patients eligible for inclusion, 167 had deep-learning auto-segmentation clinical target volume data available. Geometric indices compared contour accuracy. Clinicians completed user experience survey and time spent contouring was reported.	For the 167 cases for which complete CTV data, the median Surface Dice-Sorensen Co-efficient (DSC) and Volume DSC for CTV final vs. initial automated contours was 0.91 and 0.89, respectively. Physicians reported a median of 28min spent contouring, reflecting a 12-minute reduction in time compared to historic controls (median 40min). Physicians completed surveys for 43/55 patients. Of the 43 auto contours, 13% required major, "clinically significant" edits.

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Duan et al. 2022 (179)	China	Cancer: colon	Operating Theatres	Procedure	Nonlocal mean algorithm (NLM) AI optimises CT images taken of the colon. In this study, the NLM AI was further enhanced (iNLM) and used for CT imaging of the colon to assist identification/position of the cancer.	**Observational, single centre.** *Description:* 100 colon cancer patients underwent CT imaging, these images were analysed three times: by traditional NLM, iNLM and Filter Back Projection (FBP) Algorithms.	Imaging Performance of the iNLM Algorithm was superior as measured by the Structural Similarity Index Measure (SSIM) and Figure of Merit (FOM) index. The average running time of iNLM algorithm was quoted as being statistically significant and less than that of FBP algorithm and NLM algorithm.
Glissen Brown et al. 2022 (170)	USA	Colorectal Cancer Screening	Gastro-enterology Department	Procedure	EndoScreener: ML Computer aided polyp detection (25) system. Detects adenoma's that appear in the visual field but are missed by endoscopist.	**Interventional, multi-centre.** *Description:* Patients were randomised via computer generated randomisation to receive either CADe colonoscopy first or High-definition white light (HDWL) colonoscopy first, followed immediately by the other procedure. Patients were blinded to the result of their randomisation. Provider participants were informed of group allocation directly prior to the start of the colonoscopy procedure.	116 patient per group. In the CADe-first group, 34 adenomas were missed out of 169 total. Adenoma Miss Rate (175) of 20.12% (34/169) compared with an AMR of 31.25% (45/144) in the HDWL-first group, with an OR of 1.8. The Polyp Miss Rate (PMR) was also significantly lower in the CADe-first group compared with the HDWL-first group (20.70% vs 33.71%). There were no immediate adverse events in the CADe-first group or the HDWL-first group. False positives and false negatives rates were calculated.
Hong et al. 2022 (134)	USA	Cancer	Radiology Department	Triage	An electronic Medical Records (EMR) based ML approach to identify patients at high risk for emergency department visits and/or hospitalisation during cancer radiation therapy.	**Interventional, single centre.** *Description:* It was previously reported that a ML system could appropriately identify high-risk patients from EHR analysis, guide clinical evaluation and reduce the rate of acute care events in the high-risk population from 22.3% to 12.3%. This study focused on the implementation barriers encountered during the randomised controlled trial (RCT).	Data extraction and the need for manual review required significant time (5hrs per week). Limited data availability through the standard clinical workflow and commercial products. Aggregating data from multiple sources and logistical challenges from altering the standard clinical workflow to deliver adaptive care were other barriers.

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Kamba et al. 2021 (171)	Japan	Colorectal Cancer Screening	Endoscopy Unit	Procedure	CADe: a CNN based computer-aided detection system that assists endoscopists to detect colorectal lesions during colonoscopy.	Interventional, multi-centre. *Description:* Consented study subjects were randomised 1:1 either a "standard colonoscopy-first group" or "CADe first group" to undergo back-to-back tandem procedure.	176 patients per arm. The AMR of CADe-assisted colonoscopy was significantly lower than that of standard colonoscopy (13.8% vs 36.7% P<0.0001). The PMR, including non-neoplastic polyps, was also significantly lower in CADe-assisted colonoscopy than in standard colonoscopy (14.2% vs. 40.6%, P<0.0001). After starting colonoscopy, three patients were excluded due to irrecoverable malfunction of the CADe system.
Kneepkens et al. 2022 (159)	The Netherlands	Cancer: breast	Radiology Department	Treatment	Two ML systems predicted radiation dose distribution, generating treatment plans for breast radiotherapy.	Observational, single centre. *Description:* In this study, two previously developed ML and Deep Learning models for whole breast radiotherapy are evaluated for clinical appropriateness in a blinded review procedure, by four physicians, in addition to quantitative review (20 patients). The two AIs were compared to the corresponding manual plan.	The in-house U-net model generated higher average and maximal doses to the Planned Target Volume (PTV), and slightly higher Mean Heart Dose (MHD). The vendor developed contextual Atlas Regression Forest Model (cARF) also had higher average and maximum doses to the PTV and slightly highly MHD. Despite this, both AI plans were shown to be clinically acceptable (AI: 90-95% vs. manual: 90%). Plan preparation time was comparable between the U-net model and the manual plan (287 s vs 253 s) while the cARF model took longer (471 s).
Martins Jarnalo et al. 2021 (132)	The Netherlands	Cancer: Lung	Radiology Department	Diagnosis	DL computer aided detection (DL-CAD) system to identify and calculate size of pulmonary nodules from CT images.	Observational, single centre. *Description:* Retrospective diagnostic evaluation study: A retrospective analysis was performed of 145 chest CT examinations by comparing the output of the DL-CAD software with a reference standard based on the consensus reading of three radiologists.	Performance matched the vendor specification; the system had sensitivity of 88%, a false-positive rate of 1.04 false positives/scan and a negative predictive value of 95%.

References

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Nasir-Moin et al. 2021 (115)	USA	Cancer: Colorectal	Pathology Department	Diagnosis	AI augmented Digital System to identify clinically relevant regions of interest and classify colorectal polyps.	**Observational, multi-centre.** *Description:* Randomised Crossover Study: 100 slides with colorectal polyp samples were read by 15 pathologists using a microscope and an AI-augmented digital system, with a washout period of at least 12 weeks between use of each modality.	The use of AI for interpretation of 100 colorectal polyp samples significantly improved pathologists' classification accuracy from 74% to 81% compared with standard microscopic assessment. Time of evaluation of each slide was measured. The mean time of evaluation for all pathologists was longer when the digital system (mean 21.7seconds) than when the microscope was used (mean 13.0 seconds). Difference: –8.8 seconds; 95%CI, –9.8 to –7.7 seconds). System Usability Scale survey feedback by the 15 pathologists: The mean score for the SUS for the digital system was 68.2(95%CI, 61.3-75.0),good usability.
Ou et al. 2022 (143)	Taiwan	Cancer: Bladder	Pathology Department	Diagnosis	An AI system to automatically classify and provide quantification about atypical urothelial cells from whole-slide images (WSI) for urine cytology.	**Observational, single centre**. *Description:* Retrospective model performance study followed by prospective observational, workflow embedded, diagnostic performance study. Part 1: performance of the AI-assisted urine cytology for clinical users. Three staff reviewed the AI-inferred WSIs. The review results were compared with the expert panel consensus and the individual performance of three staff-assisted AI was evaluated. Part 2: Two staff made diagnosis' with AI, then a 4 week wash out period was implemented before staff looked again manually (conventional arm).	Ou et al. 2022 demonstrated that AI-assisted analysis of urine cytology outperformed the conventional method, with an increase of 5% in sensitivity (92% vs. 87%) and 2% in NPV (97% vs. 95%).

References

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Peng et al. 2021 (126)	China	Cancer: Thyroid	Radiology Department	Diagnosis	ThyNet: Analyse ultrasound images and videos to detect and provide a clinical grading for thyroid nodules in accordance with the ACR TI-RADS.	Observational, multi-centre. *Description:* Multi-phase mixed methods diagnostic validation study. Three stages: Test set A was diagnostic performance of ThyNet compared to 12 radiologists. Test set B was radiologists assisted with ThyNet. Test set C was real world clinic deployment.	Peng et al. 2021 conducted a prospective cohort study to demonstrate the utility of an AI system for detecting malignant thyroid nodules in a real-world clinical setting in China. Use of the system by 12 radiologists to interpret 366 ultrasound images and videos with and without AI assistance was shown to improve accuracy (AUROC: 0.837 to 0.875) and reduce the number of fine needle aspirations from 62% to 35%, and decrease missed malignancy from 19% to 17%.
Quan et al. 2022 (180)	USA	Cancer: Colorectal	Endoscopy Unit	Procedure	Real-time AI-based polyp detection system to reduce noise and increase video quality during colonoscopy.	Observational, multi-centre. *Description:* Prospective diagnostic validation study. 300 patients at two centres underwent colonoscopy with CAD system. Their results were compared to 300 historical controls performed by the same endoscopists 12 months prior to the CAD system being piloted.	Their study found that AI assistance increased detection of adenomas and serrated polyps during colonoscopy in comparison to historical controls without AI, the findings were not statistically significant.
Alessandro et al. 2022 (131)	Italy and Switzerland	Colorectal Cancer Screening	Gastro-enterology Department	Procedure	AI-enabled CADe was active for both insertion and withdrawal phases of the procedure, providing as output a bounding box any time a lesion suspected to be a polyp was recognised by CADe.	Interventional, multi-centre. *Description:* Reviewed the reported RCT method and findings, not the pooled findings: non-expert endocrinologists performed these colonoscopies. Prior to the procedure, subjects were randomised 1:1 between colonoscopy with or without CADe. Randomisation was stratified by gender, age, and personal history of adenomas. The operator was not blinded to the study arm assigned to the patient before colonoscopy treatment.	In the CADe group, 176/330 patients were diagnosed with at least one adenoma or CRC at colonoscopy as compared with 147/330 patients in the control group, corresponding to an ADR of 53.3% and 44.5%, respectively. Compared with the standard colonoscopy, CADe was associated with a difference in proportion of detected adenomas of 8.8% (95% CI: 2% to 17.9%). This means that ADR in the CADe group was non-inferior to the control group. Overall, 430/660 (65.2%) patients had polyp resections. Of these, 79/430 (18.4%) did not have histologically proven adenomas, SSLs or CRCs. These non-neoplastic polyp rates, representing 'unnecessary' polypectomies, were 12.1% and 11.8% in CADe and control group, respectively (RR: 1.03; 95% CI: 0.67 to 1.53).

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Wong et al. 2021 (181)	Canada	Cancer: central nervous system, head & neck, prostate	Cancer Centres	Procedure	Automated segmentation of OARs and clinical target volumes in CT-based radiotherapy planning for central nervous system (CNS), head and neck (H&N), or prostate cancer.	**Observational, multi-centre**. *Description:* Multi-centre survey study. Surveys were issued immediately after radiotherapy plans issued - asking for the number of edits needed to be made over the top of the AI OAR DC. Satisfaction with the contouring for OAR and clinical target volumes was also reported.	Wong et al. 2021 evaluated implementation of an AI-based auto-segmentation for CNS, H&N, and prostate radiotherapy planning at two Canadian cancer centres. AI generated plans for 551 cases RT planning required minimal edits and resulted in a positive user experience Radiation Therapists/Dosimetrists and Radiation Oncologists.
Wu et al. 2021 (154)	China	Cancer: Gastric	Endoscopy Centres	Procedure	ENDOANGEL: detect, score and grade upper gastrointestinal lesions from esophago-gastroduodenoscopy videos.	**Interventional, multi-centre**. *Description:* Patients were randomised (computerised) and assigned to ENDOANGEL assisted or control EGD. Examination protocol was the same.	Wu at al. 2021 conducted an RCT to demonstrate the effectiveness of a real-time AI assistance system for the detection of early gastric cancer involving 1050 patients at 5 five hospitals in China. Compared with the control group, the AI group had fewer blind spots (mean 5.4 vs. 9.8) and longer inspection time (5.4 vs. 4.4). The AI system correctly predicted all three early gastric cancer (one mucosal carcinoma and two high grade neoplasias) and two advanced gastric cancers, with a per-lesion accuracy of 85%, sensitivity of 100%, and specificity of 84%.
Xu et al. 2021 (168)	China	Colorectal Cancer Screening	Colonoscopy	Procedure	AI system alerts the endoscopist in real time to detect polyps visually with a green indicator box and audible sound	**Interventional, multi-centre.** *Description:* Eligible patients were randomly assigned to conventional colonoscopy (control group) or AI-assisted colonoscopy (AI group). AI assistance was our newly developed AI system for real-time colonoscopic polyp detection. Switched off for the control group. Primary outcome is polyp detection rate (PDR). Secondary outcomes include polyps per positive patient (PPP), polyps per colonoscopy (PPC), and non-first polyps per colonoscopy (PPC-Plus).	1175 patients control group and 1177 in AI group. the overall PDR of control group and AI group were 36.2% and 38.8%, respectively, and there was no significant difference between two groups. A total of 930 polyps were detected in control group, and 1042 polyps were detected in AI group. Among them, there were 505 non-first polyps (polyps detected after the first one during colonoscopy) in control group and 585 in AI group. For secondary outcomes, the PPP (2.3 vs. 2.2, $p = 0.113$) and PPC (0.9 vs. 0.8, $p = 0.092$) showed no statistical difference between the control and AI groups, while the PPC-Plus of AI-assisted colonoscopy was significantly higher than conventional colonoscopy (0.5 vs. 0.4, $p < 0.05$).

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Cardiovascular (n=5)							
Cheema et al. 2021 (123)	USA	Point of Care Cardiac Ultrasounds for COVID-19 patients	COVID-19 Intensive Care Unit ward	Procedure	CNN enabled device providing real-time prescriptive guidance to steer the user's transducer position and hand movements to acquire cardiac ultrasound images while displaying the current image quality. Automatic capture of the optimal image. It also automatically detects ejection fraction independent of chamber volumes.	**Observational, single centre.** *Description:* Five patient cases described where the AI-guided POCUS was used to obtain images that were then uploaded to the picture archiving and communication system for the attending cardiologist to over read.	Multiple diagnoses and measurements were made in all five cases. Treatment management adjustments, decision changes and clinical outcomes summarised in the context of each case but causality not established.
Chen J, Gao Y. 2021 (139)	China	Cardiovascular disease	Cardiovascular Imaging Department	Diagnosis	CNN that improves image quality and reduces noise in echocardiography images to diagnose and evaluate the effect of anti-heart failure western medicines in elderly patients.	**Interventional, single centre.** *Description:* 80 patients with Acute Left Heart Failure were divided randomly to control group (standard echocardiography) and observation group (CNN-echo). Control group treated with western medications (carvedilol, valsartan, hydrochlorothiazide) whilst the observation group received Chinese medicine (shengmai injection). Mortality rate, rehospitalisation rate, length of stay and hospitalisation expenses were observed over 5 months.	Demonstrated potential of echocardiography based on a deep learning algorithm to improve diagnosis of cardiovascular events in patients with heart failure. Rehospitalisation rate and mortality rate of patients from two groups were not statistically significant. Diagnostic accuracy of control group coincidence rate 74.29% compared to 93.94% for observation group. QoL measured by SF-36 scale showed both groups improved after treatment, with observation group statistically significant.

References

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Edalti et al. 2022 (127)	USA	Cardiovascular disease - MRI	Cardiovascular MRI department	Procedure	Two deep neural networks developed to reduce technical complexity and time execution of Cardiovascular Magnetic Resonance (CMR) imaging: EasyScan: Automatic slice planning AI-Shim: A generalised shimming tool using a mask-based artificial intelligence segmentation technique that assists the operator to attain the best scanner frequency.	**Observational, unknown number of centres.** *Description:* Pilot study for clinical validation - Two prospective studies performed. For the EasyScan validation, 10 healthy subjects underwent two identical CMR protocols: with manual cardiac planning and with AI-based EasyScan to assess protocol scan time difference and accuracy of cardiac planner prescriptions on a 1.5 T clinical MRI scanner. For the AI-Shim validation, 10 healthy and 10 cardio-oncology patients with referrals for a CMR examination were recruited. Images were obtained with standard cardiac volume shim and with AI-Shim. Signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR),overall IQ (sharpness and MR image degradation), ejection fraction (EF), and absolute wall thickening parameters compared.	Edalti et al. 2022 evaluated the performance of two AI algorithms to improve image quality and reduce noise in MRI images. Automate image acquisition was shown to minimise operator dependence and was 13% faster compared to manual planning of cardiac scans. Mean time difference manual cardiac planning vs Easyscan: 2.57 minutes faster with EasyScan. AI-Shim was more robust (higher signal-to-noise ratio).

References

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Liu et al. 2021 (144)	Taiwan	Acute Myocardial Infarction	Emergency Department	Diagnosis	AI-S: artificial intelligence alarm strategy for AMI detection. All obtained ECGs were uploaded in real-time to the AI-S platform to perform AMI auto diagnosis. Each ECG obtained a STEMI and a NSTEMI score (0-1) within 10 seconds and stored in the electronic medical record. Meanwhile, triage provided the symptom assessment, and the lab immediately uploaded the lab data. The AI-S incorporates chest pain symptoms, 12 lead ECG and hsTnI to produce a prediction score for AMI diagnosis. Once the AI-S indicated STEMI or NSTEMI, warning message is triggered to ED on-duty cardiologist.	**Observational, single centre.** *Description:* Prospective validation before and after study. The primary analysis was model performance. Secondary analysis evaluated each component of Door to Balloon time before and after AI-S implantation. One-year major adverse cardiac events (MACEs) after Primary Percutaneous Coronary Intervention (PPCI) including all-cause mortality, heart failure hospitalisation, and non-fatal MI after PPCI, before and after AI-S implementation were evaluated.	Patients with STEMI were alerted precisely by AI-S (F-measure=0.932, precision of 93.2%, recall of 93.2%). Compared with pre-AI-S(N=57) and post-AI-S (N=32) implantation in STEMI protocol, the median ECG-to-cardiac catheterisation laboratory activation (E to CCLA) time was significantly reduced from 6.0 (IQR,5.0–8.0min)to 4.0min (IQR,3.0–5.0min) (p<0.01). The median D to B time was shortened from 69 (IQR,61.0–82.0min) to 61min (IQR,56.8–73.2min) (p=0.037).
Zhou et al. 2021 (167)	China	Venous thromboembolism (VTE)	Hospital Wide	Monitoring	AI-CDSS embedded in EHR that analyses patient information, scored VTE, and bleeding risk 6-hourly. Notifies clinicians about patients at risk of VTE.	**Observational, single centre.** *Description:* A pre-and post implementation study design, January-July 2019 is pre AI-CDSS deployment. January-July 2020 is deployed. The primary endpoint of the study was diagnosed as a hospital-acquired VTE.	AI-enabled automated assessment of VTE risk every 6 hours or whenever new information was entered in the EHR was found to reduce the rate of VTE during hospitalisation by 19% and increased anticoagulant drug use by 14%.

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Clinical Deterioration (n=3)							
Martinez et al. 2023 (113)	USA	Clinical Deterioration	Hospital Wide	Monitoring	Advance Alert Monitor (AAM), to improve early detection and intervention for in-hospital deterioration. The AAM predictive model is designed to give clinicians 12 hours of lead time before clinical deterioration, permitting early detection and a patient goal–concordant response to prevent worsening.	**Observational, multi-centre.** *Description:* This literature is a case summary describing successful deployment and implementation across 21 hospitals.	The AAM program is associated with statistically significant decreases in mortality (between 550 and 3,020 over four years), hospital length of stay, and ICU length of stay. In the intervention cohort, there was a 3.8% absolute decrease in mortality within 30 days after an event reaching the alert threshold. This difference translated into 3.0 deaths (95% confidence interval [CI] = 1.2–4.8) avoided per 1,000 eligible patients, or 520 deaths avoided (95% CI = 209–831) per year over the 3.5-year study period.
Schwartz et al. 2022 (161)	USA	Clinical Deterioration	Hospital Wide	Monitoring	CONCERN is a predictive CDSS implemented at two hospitals and currently under investigation for its ability to predict in-hospital deterioration.	**Observational, multi-centre.** *Description:* Interview data analysis of clinicians from 24 acute and intensive care units in two hospitals. who used the CDSS, guided by a conceptual framework called the 'human-computer trust framework.	Study confirmed that trust is influenced by clinician perceptions about being able to form a mental model and predict future system behaviour as well as the system's technical capabilities to perform tasks accurately and correctly based on the information that is input. Perceptions about system accuracy were found to be influenced by the concordance between clinician impressions of patient clinical status and system predictions and understandability was influenced by system explanations. Trust was also influenced by actionability of system recommendations, scientific and anecdotal evidence as well as fairness in system predictions. The findings were largely similar between nurses and prescribing providers.

References

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Winslow et al. 2022 (160)	USA	Clinical Deterioration	Medical-Surgical Ward	Monitoring	eCART: electronic Cardiac Arrest Risk Triage score - an ML algorithm that identifies patients at risk of death in the next 24 hours	**Observational, multi-centre.** *Description:* Before and after study - measure the real-world impact on provider behaviour and patient outcomes of prospectively integrating an ML early warning analytic into clinical workflows at four hospitals. The primary outcome was all-cause hospital mortality among patients who ever had an elevated eCART score.	Deployment of the system across a multicentre health system in the US over 10-months was associated with a decrease in hospital mortality (9% vs 14%). Compared with the baseline, hospital mortality was significantly lower during the intervention period (8.8% vs 13.9%; $p < 0.01$) for the main cohort. This represented a relative risk reduction for death of 36.7%. This decrease in mortality was seen in both the high-risk (17.9% vs 23.9%; $p = 0.001$) and intermediate-risk subgroups (2.0% vs 4.0%; $p = 0.001$). Being in the intervention period was associated with an adjusted odds ratio (aOR) for death of 0.60 (95% CI, 0.52–0.71) across the main study population, with similar benefits across the two risk subgroups. The only patients that did not appear to benefit from the intervention were those whose first eCART elevation occurred after admission to the ICU (aOR, 1.04; 95% CI, 0.75–1.43)

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Clinical Trial Eligibility Screener (n=1)							
Kanbar et al. 2022 (116)	USA	Clinical Trial Eligibility Screener	Paediatric Emergency Department	Triage	ACTES: an NLP based Automated Clinical Trial Eligibility Screener for real-time identification of patients for research studies in a paediatric emergency department. a step-by-step process that extracts data from the eHR, processes it, and provides a recommendation in the form of automated alerts that could be sent from the research system to the eHR in real time.	**Observational, single centre.** *Description:* ACTES was prospectively evaluated using a time-and-motion study, quantitative assessments of enrolment, and post-evaluation usability surveys collected from the CRCs. During the time-and-motion study, an observer monitored the activities a CRC was engaged in at 30-second increments for two hours. The time spent per activity was compared to that prior to the use of ACTES. This study was repeated monthly for four months, and it was distributed among CRCs and shifts.	After the implementation of ACTES, the CRCs spent 12.9% (P<.001) less time on electronic screening. The quantitative assessments of enrolment evaluated the number of patients screened, the number of patients approached, and the number of patients enrolled. The use of ACTES significantly improved the number of screened patients for the majority of trials and improved the number of approached patients and enrolled patients, with statistical significance in two of seven trials [52]. Finally, results from the System Usability Survey and additional open-ended questions were analysed on a monthly basis to improve ACTES.
COVID-19 (n=4)							
Alrajhi et al. 2022 (146)	Saudi Arabia	COVID-19	Admissions Ward	Diagnosis	After trialling four ML classifiers a Random Forest model with feature selection breakdown was the superior model to predict severity of COVID-19 infections using eHR data, matching patients with appropriate levels of needed care, improving resource management.	**Observational, single centre.** *Description:* An initial cohort study was performed to provide training data sets for the models (March 2020 – April 2021 COVID-19 cases). Model was then implemented and underwent validation via prospective cases (April - May 2021)	Alrajhi et al. demonstrated performance of a home-grown AI to predict the severity of COVID-19 infection for patients at the time of hospital admission (recall: 78–90; precision: 75–98% for different severity classes of COVID-19 (Asymptomatic, Mild, Moderate and Severe). Precision was highest (98%) for severe COVID-19.

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Garzon-Chavez et al. 2021 (150)	Ecuador	COVID-19	Radiology Department	Triage	AI analyses chest CT scan to stratify COVID-19 suspected patients (non-severe 0-30%, moderately severe 30-70% and severe >70%). Does this by comparing predicted lesions from a trained AI with the actual lesions from the CT scan. Depending on the score, patients get distributed to different 'score rooms'. across three hospital towers.	**Observational, single centre.** *Description:* Retrospective review of the first seventy-five patients triaged by this AI.	Reported the severity scores for 37/75 laboratory-tested patients (49.3%). Sensitivity corresponded to 21.4% and specificity to 66.7% when considering the likelihood to classify a patient as COVID-19 positive with a score >70%. Thus, 7/28 positive and 3/9 negative laboratory-tested cases (n = 10) were allocated in 70% score rooms; 10/20 positive and 1/9 negative laboratory-tested cases (n = 11), in score rooms for the 30–70% category; and 11/28 positive and 5/9 negative laboratory-tested cases (n = 16) were allocated in rooms for scores less 30%.
Hinson et al. 2022 (151)	USA	COVID-19	Emergency Department	Triage	Developed, implemented, and evaluated an electronic health record (eHR) embedded clinical decision support (CDS) system that leverages ML to estimate short-term risk (scoring 0-10) for clinical deterioration in patients with or under investigation for COVID-19.	**Observational, multi-centre.** *Description:* Conducted across five sites, this prospective validation study had two cohorts - silent CDS deployment group and Visible CDS deployment group.	Hinson et al. undertook a staged evaluation to assess the performance of an AI system that provides a COVID-19 Clinical Deterioration Risk Level (1–10) for each ED encounter in real-time based on EHR data. Prospective validation over 18-months at five emergency departments including an initial silent deployment showed ML system performance with AUC ranging from 0.85 to 0.91 for prediction of critical care needs and 0.80–0.90 for inpatient care needs. Total mortality was reduced among high-risk patients after AI implementation.
Maheshwarappa et al. 2021 (125)	India	COVID-19	Intensive Care Unit	Diagnosis	Vscan Extend™ is a handheld ultrasound device with a dual probe and an artificial intelligence application to detect ejection fraction. The application automatically traces the endocardial border of the left ventricle, deriving the left ventricular end-diastolic volume (LEDV) and left ventricular end-systolic volume LVESV. The ejection fraction of the left ventricle is calculated from these two values.	**Observational, single centre.** *Description:* This is a prospective observational study (Vscan extend vs conventional ultrasound machine). Pair wise approach. Intensivist A used Vscan Extend device to assess cardiac function, lung fields, diaphragm, deep veins, and abdomen. Intensivist B used clinical examination, x ray chest, ECG and ECHO.	96 paired readings. The median duration of examination using handheld ultrasound was 9 (8.0–11.0) minutes, compared to 20 (17–22) minutes with the conventional method ($P < 0.001$). The agreement between the intensivists diagnostic findings were compared statistically using Cohen's kappa coefficient: 1.0 for left ventricular systolic function (perfect agreement), and 1.0 for most of the lung parameter fields., There was poor agreement for right ventricular systolic function and pericardial effusion (0.07 and -0.01 respectively).

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Dermatology (n=1)							
Pangti et al. 2021 (141)	India	Dermatological Diseases (Multiple)	Outpatient Departments	Diagnosis	A CNN enabled decision support mobile phone application to predict top three skin conditions with probabilities.	**Observational, multi-centre.** *Description*: Part 1: in silico model validation on 41 skin conditions. Part 2: Multi-centre, observational, prospective diagnostic study including 3699 patients from tertiary hospitals.	Pangti el al. 2021 undertook a large-scale study involving 5014 patients across a wide variety of clinical settings in India to demonstrate the utility of a smartphone mobile app as a point-of-care tool for diagnosis of 41 skin conditions in people of colour (overall accuracy 75%, top 3 accuracy 90%).
Gastroenterology (n=1)							
Maeda et al. 2022 (166)	Japan	Chronic Inflammatory Bowel Disease	Endoscopy Unit	Monitoring	ML enabled prediction and categorisation (Healing or Active) of histologic disease activity of ulcerative colitis. After the endoscopist presses the capture button on the endoscope to acquire an image, the endoscopy monitor displays a 2-category prediction output with the probability of the prediction. When the probabilities of both categories are <70% the AI systems outputs "low confidence" instead of showing a specific prediction.	**Observational, single centre.** *Description:* This open-label, prospective, cohort study was conducted at a single centre. UC Patients recruited May - Dec 2019 and followed up for 12 months after AI-assisted colonoscopy. Immediately after each colonoscopy the endoscopist completed the case report form by inputting Mayo Endoscopic subscore MES for each segment and any adverse events. The AI prediction was automatically recorded in a csv file.	The relapse rate was significantly higher in the AI-Active group (28.4% [21/74]; 95% confidence interval, 18.5%-40.1%) than in the AI-Healing group (4.9% [3/61]; 95% confidence interval, 1.0%-13.7%; P < .001). We obtained biopsy samples from 810 segments in 135 patients. The overall diagnostic sensitivity, specificity, and accuracy of the AI output for predicting persistent histologic inflammation were 82.5%, 95.4%, and 93.8%, respectively.

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Hematology (n=1)							
Choudhury 2022 (158)	USA	Blood Transfusion for any indication	Hospital Wide	Treatment	This is an AI-based Blood Utilisation Calculator (113), a module of an electronic decision support program known as the Digital Intern (iVMD). Proprietary computer-based algorithm that retrieves patient information from the electronic medical record and delivers data-driven personalised recommendations for the number of packed red blood cells to transfuse for a given patient.	**Observational, single centre.** *Description:* Survey analysis study with quantitative and descriptive variables. Mass email with description of study delivered to medical professionals who use BUC. The email described the purpose of the study between February - July 2021. We used RedCap to collect survey responses. The survey contained a screening question asking whether they have ever used the BUC system (with an explanation and picture of BUC). Only BUC users were asked to complete the survey. We discarded incomplete and duplicate responses.	119 survey responses analysed. Clinicians agreed that AI systems could improve patient outcomes (mean 3.97, max 5) and disagreed that the use of BUC can put them or their patients at risk (mean 1.95 and 1.83, respectively). Clinicians also perceived BUC as an easy to use AI system (mean3.76); they agreed that learning how to use BUC and becoming skilful at it was easy (mean 3.81 and 3.82, respectively). Most of the clinicians neither agreed nor disagreed with the question asking if the BUC increased their chances of achieving/fulfilling important clinical tasks (mean 3.33). However, most of them agreed that BUC improved their pace (mean 3.36) and effectiveness at blood transfusion (mean 3.64).
Infection (antimicrobial) (n=1)							
Rawson et al. 2021 (121)	UK	Infection (Antimicrobial)	Hospital	Treatment	Case Based Reasoning (43) algorithm underpins an antibiotic prescribing CDSS. A supervised ML tool also provides support on the likelihood of infection being present.	**Observational, single centre.** *Description:* Real-world evaluation of the CDSS using two patient population study. Escherichia coli patients and ward-based patients presenting with a range of potential infections ("ward patients"). The CDSS was deployed and used by six members of specialist medical staff.	Of the 224 individual patients included, 202 (90%) of the CBR recommendations were deemed appropriate based on the spectrum of antimicrobial activity required. This was compared to 186/224 (83%) of prescriptions made by physicians. There was no statistical difference between physicians and CBR recommendations.
Mental Health – Suicide (n=1)							
Wilmitis et al. 2022 (164)	USA	Mental Health – Suicide risk	Multiple acute settings	Monitoring	VSAIL: is as real-time suicide risk prediction: Suicide Attempt (SA) and Suicidal Ideation (SI)	**Observational, single centre.** *Description:* Comparing the VSAIL prediction model ability to predict suicide attempt and suicidal ideation to the standard C-SSRS. Then combined both to see if that had a synergistic effect on performance. The primary outcomes were SA and SI occurring within 7, 30, 60, 90 and 180 days after the discharge date of each documented visit during the time period.	Combined models outperformed either model alone for risks of suicide attempt and suicidal ideation in a cohort study of 120,398 adult patient encounters in the USA. In the highest risk-decile, the combined methods had PPV of 1.3% to 1.4% for SA and 8.3% to 8.7% for SI and sensitivity of 77.6% to 79.5% for SA and 67.4% to 70.1% for SI, outperforming VSAIL alone and C-SSRS alone.

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Mortality Risk (n=2)							
Park et al. 2023 (162)	USA	Trauma Mortality Risk	Emergency Department	Monitoring	A validated mortality Risk Calculator (Parkland Trauma Index of Mortality PTIM) embedded in the EHR that calculates hourly prediction of mortality. The clinician can then utilise this mortality prediction for planning (e.g. operative intervention time, goals of care).	**Observational, single centre.** *Description:* Acceptability and usability analysis. If the PTIM score was utilised in medical decision making, an anonymous survey could be completed via REDCap site. The survey first queried what top three measurable factors (Glasgow Coma Score (GCS), age, creatinine, and hemoglobin) the clinician felt contributed most to the patient's PTIM score. Next, the survey queried whether the score assisted in guiding the clinician's recommended treatment plan. Finally, the survey addressed the clinician's perceived ease of use, current utility, and future use.	35/40 surveyed said they used the PTIM score in medical decision making. Top three predictors of mortality align with the algorithms calculated actual most significant predictors of mortality (GCS, age, and max pulse rate). 27/35 reported that the PTIM score assisted in determining the course of the treatment plan and surgical intervention timing. 22/36 thought it was easily integrated into ward rounds and patient assessments. 21/36 said it improved efficiency in assessing mortality. 21/36 said they would continue to use it, 15/36 were neutral on that.
Kermani et al. 2023(163)	Iran	Neonatal Mortality	Neonatal Intensive Care Unit	Monitoring	Case-Based Reasoning (43) system web-based deployment, predicting neonatal survival (mortality risk score) and Length of Stay (LOS). After a user enters a new case, the CBR module retrieves similar cases based on the previous cases in the case base (search process in the case base) using the weighted Euclidean distance similarity function and KNN algorithm (K is determined by the user).	**Observational, single centre.** *Description:* Multi-stage research: Development phase (CBR model development) then Evaluation phase made up of : • 1: Retrospective evaluation prior to web-based system launch • 2: Prospective evaluation and external validation based on 3 months deployment in NICU - 92 neonates followed until discharge. Compared model prediction against ground truth. • 3: Acceptability and confidence evaluation by Likert Questionnaire. Usability evaluation by 'think-aloud' method and System Usability Scale Questionnaire (5-Likert scale ranging from one to five). Neilsen severity scale to classify the usability problems.	During the implementation period for the external validation, 92 neonates were admitted and included in the analysis. 74 (80.43%) neonates were finally alive, and 18 (19.57%) were dead. The average LOS was 11.39 days (1–90 days). External validation on the unbalanced case base showed the accuracy and specificity measures were 97.82% and 88.88%, respectively. Furthermore, the kappa coefficient was 0.928 which indicated a very good agreement between the system predictions and the real outcome. The physicians' acceptance of survival prediction system outputs was higher than the LOS prediction system. For the survival prediction system, the mean score for acceptability and confidence were 4.88 and 4.25, respectively. Furthermore, the physicians' acceptance and confidence in LOS prediction system responses were 4.96 and 3.96, respectively.

References

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Neurology (n=1)							
Kanbar et al. 2022 (116)	USA	Epilepsy	Ambulatory Neurology Clinic	Diagnosis	EPILEPSY ID: generates surgical candidacy score for each patient using NLP.	**Observational, single centre** *Description:* Case summary - implementation and key learning	EPILEPSY ID: The epilepsy ID system performed as well as board-certified neurologists in identifying surgical candidates (with a sensitivity of 71% and positive predictive value of 77%
Ophthalmology (n=1)							
Hao et al. 2022 (142)	China	Diabetic retinopathy	Local Community Hospital	Diagnosis	EyeWisdom: AI analysis software that detect and grade Diabetic Retinopathy (DR) severity from fundus images.	**Observational, single centre** *Description:* Diagnostic Accuracy Study: The AI based diagnostic system and ophthalmologists were tasked with screening for diabetic retinopathy in 7824 eye-fundus photos independently, and the consistency rate, sensitivity, and specificity of the two methods in diagnosing DR were calculated and compared.	Hao et al. 2022 evaluated the performance of an AI system for diabetic retinopathy screening involving 3933 patients in a community hospital in rural China. The AI was demonstrated to have a sensitivity of 81% and specificity of 94% and was consistent with screening by ophthalmologists.
Orthopaedics (n=4)							
Li et al. 2022 (147)	Taiwan	Surgery: Hip repair	Surgical Ward	Diagnosis	An ML-based application that can assist anaesthesiologists in assessing specific adverse outcomes for patients required to undergo hip repair surgery.	**Observational, single centre** *Description:* Retrospective, model validation. After ML model training and performance testing, the optimal models were deployed into the existing IT infrastructure to assist anaesthesiologists in performing preoperative risk assessment for patients with hip fractures. Study's primary outcome was a composite of postoperative adverse events, ICU admissions prolonged length of stay (PLOS)	The AI was demonstrated higher sensitivity, specificity, accuracy, and performance than that of the American Society of Anaesthesiologist-Physical Status (ASA-PS), the traditional risk stratification method: primary composite outcomes (0.810 VS 0.629, P<0.01), ICU admissions (0.835 VS 0.692, P<0.01), and PLOS (0.832 vs 0.618, p<0.01). Demographics and incidences of adverse outcomes in 545 and 500 patients before and after implementing the online we-based application. There was no statistically significant decrease in the incidence of primary composite adverse events (3.3 vs 1.6%, p=0.117) or ICU admission (4.4 vs 2.4%, p=0109) after the application was initially employed for clinical use. Clinician satisfaction score increased from 3.21 (1st month) to 4.70 (month 10). The score was significantly higher starting in the 4th month after the application was launched (p<0.01).

References

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Liu et al. 2021 (172)	China	Imaging: CT (rib fracture)	Radiology Department	Diagnosis	Detect rib fractures on CT images	**Observational, multi-centre.** *Description:* Detection of fractures with and without AI by Junior Radiologists. All the CT images were randomly divided into two sets of images at each institution, with each set assigned to one of the two radiologists. In a routine manner, the readers went through all cases over two sessions. Each radiologist read the same CT image twice, with and without AI software (uAI-BoneCare) assistance, with a one-month washout period between the second read of the same CT image.	Use of AI improved the sensitivity of rib fracture detection on CT images for junior radiologists and reduced the reading time by ~1 min per patient without decreasing the specificity.
Liu, Cheng. 2021 (157)	China	Surgery: Scapular fracture	Operating Theatres	Procedure	CNN applied to ultrasound images to improve accuracy and thereby correct location of the anaesthesia point during scapular fracture surgery.	**Interventional, single centre** *Description:* Multi-stage research with quasi-randomised controlled study. Part 1: measure the difference in image accuracy between deep learning ultrasound images and manual/ordinary; Part 2: observe the adoption of AI-ultrasound to optimise anaesthesia puncture path. Part 3: determine the effectiveness of IS-imaging guided scapular regional nerve-block in the treatment of surgical pain of fracture. 100 patients were randomly assigned AI or ordinary.	It was found that the adoption of deep learning greatly improved the accuracy of the image. It took an average of 7.5±2.07 minutes from the time the puncture needle touched the skin to the completion of the in the AI group. The operation time of the control group (anatomical positioning) averaged 10.2±2.62min. The effects of the motion block between the two groups showed the block effect to be statistically different and improved in for the AI group. For adverse events: the number of needle tracks needed to be adjusted during puncture in the control group was 3.25±1.36 times, AI group was 2.11±1.31 times, P=0.009. The times of encountering bone during puncture were 1.91±1.34 times and 0.68±0.73 times in the two groups. Evaluation of anaesthetic effect was better in the AI group (as evaluated by a second anaesthesiologist within 30 minutes of injection).

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Zhang et al. 2021 (169)	China	Imaging: CT (rib fracture)	Radiology Department	Diagnosis	Deep learning (DL) algorithm to identify and highlight rib fractures during the review of CT images.	**Observational, single centre.** *Description:* Retrospective validation examining the rib fracture detection accuracy by asking radiologists to interpret images unassisted, AI assisted and with the AI, either as a second reader or concurrent reader assistance. All images were from January to June 2019 and blunt chest trauma patients. Analysed at an independent workstation with the prototype DL software in place.	Zhang et al. 2021 investigated the impact of an AI system on detection accuracy and reading efficiency of rib fractures on CT by asking radiologists to interpret images unassisted, assisted and with the AI as a second reader. Use of AI as a second reader was found to improve detection accuracy (5–6% more rib fractures were found by the readers with AI assistance than without AI) and reading efficiency for rib fracture (reading time reduced by 34-36%).
Paediatric Medicine (n=1)							
Eng et al. 2021 (118)	USA	Growth disorders and scoliosis	Radiology Department	Diagnosis	A Deep Neural Network algorithm to analyse hand radiographs rapidly and accurately diagnose skeletal maturity of paediatric participants.	**Interventional, multi-centre.** *Description:* Prospective, RCT, 792 with AI enabled hand radiograph examination vs 739 without AI. Multicentre (superiority diagnostic study). The primary efficacy outcome was the mean absolute difference between the skeletal age dictated into the radiologists' signed report and the average interpretation of a panel of four radiologists not using a diagnostic aid. The secondary outcome was the interpretation time.	Overall mean absolute difference in skeletal age was lower when radiologists used the AI algorithm compared with when they did not (5.36 months vs 5.95 months; P = .04). The proportions at which the absolute difference exceeded 12 months (9.3% vs 13.0%, P = .02) and 24 months (0.5% vs 1.8%, P = .02) were lower with the AI algorithm than without it. Median radiologist interpretation time was lower with the AI algorithm than without it (102 seconds vs 142 seconds, P = .001).
Renal (n=1)							
Chen et al. 2022 (165)	China	Chronic Kidney Disease	CT Imaging Department	Monitoring	The AI (wavelet transform de-noising) optimises the CT images of kidneys of patients with chronic kidney disease who can't have high dose CT contrast fluid, enhancing the accurate measurement of renal perfusion.	**Interventional, single centre.** *Description:* Quasi-randomised controlled study. A random table method was used to divide the patients 60:60 (n=120) to normal nutritional nursing model 'control group' vs the 'Internet+H2H" group. The IWT algorithm was used for CT images for both groups and was compared with two 'traditional' algorithms Mean Filter De-noising (MFD) and Orthogonal Wavelet De-noising (162) algorithms.	Whilst the primary focus of this study was the nutritional intervention - the AI component of this research was an IWT algorithm. The clarity of the IWT algorithm enhanced kidney CTs were compared to the images obtained by more traditional algorithms MFD and OWD and based on their MSE values and SNR values, the de-noising effect of the IWT was superior. (MSE 40.0781,45.2891, and 59.2123, SNR values 20.0122, 18.2311, and 15.7812).

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Respiratory (n=8)							
Carvallo et al. 2023 (117)	USA	Imaging: Pulmonary Nodules	Emergency Department: Radiology	Diagnosis	The QA program uses a vision-based CNN algorithm to analyse images (AIDOC- cloud based), a natural language processing (NLP) tool to analyse radiology reports (classified to +ive or -ive result) and a semi-automated e-mail notification system to alert physicians and track relevant studies.	**Observational, single centre** *Description:* Retrospective analysis of the performance of this deployed QA program from October 2021 to June 2022.	19k+ CT scans of which 15k+ were categorised negative based on the NLP read of the final radiology report. Those 15k+ CTs were then pushed through the imaging AI which identified 50 suspected discrepancies. A radiologist reviewed these 50 and found 34 to be warranted for addenda to be issued to the original report. Median time from original report to 2nd report was eleven hours (facilitated by auto-email). 20 resulted in a recommendation to get more images. Of the 16 CTs that were not addended, most were due to false-positives by the AI nodule detection software, 1 was a false-negative by the NLP system.
Dean et al. 2022(130)	USA	Pneumonia	Emergency Department	Diagnosis	ePNa: a CDSS extracting real-time and historical data to guide diagnosis, risk stratification, microbiological studies, site of care and antibiotic therapy. Specific ML features in ePNa are NLP to identify information in free-text radiology reports to determine radiographic pneumonia. A Bayesian probabilistic algorithm calculates and displays percent likelihood of pneumonia and the pertinent data elements directly to ED clinicians. ePNa alerts clinicians when pneumonia probability is ≥ 40%. The clinician chooses either to launch ePNa or not.	**Interventional, multi-centre**. *Description:* Stepped-wedge, cluster-controlled trial. Deployed ePNa into six geographic clusters of 16 Intermountain hospital ESs at 2-month intervals between December 2017 and November 2018 according to a pre specified plan (www. clinicaltrials.govidentifier:NCT03358342). Mortality and processes of care were the primary and secondary outcomes of this study. 4500 pts formed the pre-AI cohort, 2300 post-AI cohort.	Observed 30-day all-cause mortality, including both outpatients and inpatients, was 8.6% before deployment versus 4.8% after deployment of ePNa. Mortality reduction was greatest inpatients directly admitted to ICUs from the ED (OR,0.32; P=0.01) compared with those admitted to the medical floor (OR,0.53; P=0.09) and with outpatient disposition. Amon g patients admitted to the hospital, guideline-/ePNa-concordant antibiotic prescribing increased from 79.5% to 87.9%. Use of broad spectrum antibiotics did not change pre-and post deployment. Mean time from ED admission to first antibiotic use was 159.4mins and went down to 150.9mins after deployment. Overall, ePNa was used by the ED clinician in 67% of eligible patients with pneumonia after deployment. Use was 69%in the 6 larger hospitals but 36% in the 10 smaller rural hospitals.

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Knighton et al. 2022 (136)	USA	Acute Respiratory Distress Syndrome	Intensive Care Unit	Monitoring	A CDS synchronous alert tool associated with existing computerised ventilator protocols and targeted patients with possible Acute Respiratory Distress Syndrome (190) not receiving Lung Protective Ventilation (LPV). Specifically, a natural language program looks for trigger words in chest X-ray reports and feeds that into the CDS.	Observational, multi-centre *Description:* Explanatory mixed methods study (quantitative methods to measure service outcomes and qualitative methods to understand attitudes/ appropriateness / acceptability). Across 13 ICUs in a healthcare system. Implementation outcomes included appropriateness, discriminatory power, and acceptability of the CDS alert tool and its accuracy. Service outcomes: increased visibility of non-adherent practices, clinician behaviour changes and minimising unnecessary alerts.	1553 trigger events: 775 events where possible ARDS was detected, 455 events where possible ARDS was detected and LPV treatment was not detected during study time frame. 38% had at least one episode of initial guideline nonadherence. Overall, 48% of recommendations were followed within the defined adherence timeframe. There was a 34% survey response rate. 57% of survey respondents identified one or more potential benefits associated with use or potential use of the alert. 68% strongly agreed/agreed that generally using an automated alert in the EHR fits with the say they like to work. Among 73 intubated patients, the AUROC of the CDS alert tool was 0.62 (95%CI:0.47–0.74), with a sensitivity of 0.87 (95%CI:0.73–0.96), a false positive rate of 0.66 (95%CI:0.450.80) and a positive predictive value of 0.62 (95%CI:0.48–0.75).
Lee et al. 2022 (124)	South Korea	Imaging: Chest X rays	Radiology Department	Diagnosis	Lunit Insight CXR MCA: Identifies suspected findings of lung nodules, consolidation and pneumothorax, mark regions of interest and provide abnormality scores from chest x-rays.	Observational, single centre. *Description:* Case study - implementation at a general hospital, describing benefits that can be gained in daily practice and the factors needed for successful implementation.	Lee et al. 2022 describe their experiences in setting up and operating AI interpretation of chest x-rays in a hospital setting in South Korea. Both accuracy and immediate availability of AI results was reported to be necessary and critical, along with explainable visualisation of disease-specific results and improved medical software platforms providing data presentation that were configurable by users.

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Nhat et al. 2023 (43)	Vietnam	Imaging: Point of Care Lung Ultrasounds for Acute Respiratory Distress Syndrome	Intensive Care Unit & Emergency Department	Diagnosis	RAILUS (Real-time AI-assisted LUS) consists of an AI model integrated into the PRETUS platform (a plug-in-based platform for real-time ultrasound imaging research). RAILUS provides continuous real-time prediction of ARDS through a laptop and can be used in both pre-recorded LUS clips and real-time clinical environment with clinicians carrying out the LUS examination, with the ultrasound machine's video output connected to the laptop. RAILUS also captures the user prediction, model prediction and time-to-interpret.	Observational, single centre. *Description:* This was a three-phase prospective study. In the first phase, the performance of four different clinical user groups in interpreting LUS clips was assessed. In the second phase, the performance of 57 non-expert clinicians with and without RAILUS for LUS interpretation was assessed in retrospective offline clips (workshop type setting). In the third phase, a prospective study was initiated in the ICU where 14 non-expert clinicians were asked to carry out LUS examinations in seven patients with and without our AI too. Clinicians were interviewed regarding the usability of the AI system. Ground truth was an expert who performed the LUS within two hours of the non-expert clinician.	Seven patients recruited for real-time testing of RAILUS software. 168 videos performed with the AI system, 144 without the tool. Accuracy of image identification was higher in those using the RAILUS AI system than those using the standard LUS technique: 93.4% (95% CI 89.0–97.8%) compared to 68.1% (95% CI 57.9–78.2%), ($p < 0.001$). Performance was better in all classes for clinicians using our AI system compared to those without AI assistance. The time taken to interpret one LUS clip was shorter when using the RAILUS software compared to the standard LUS technique: a median of 5.0 s (IQR 3.5–8.8) compared to 12.1 s (IQR 8.5–20.6) ($p < 0.001$). In addition, the median confidence level of clinicians improved from 3 out of 4 to 4 out of 4 when scanning patients using the AI system. 13/14 (93%) found the AI-assisted tool useful in the clinical context and wanted to use the tool in the future (12/14, 86%). 64% (9/14) of clinicians thought the tool was useful for both real-time and post-exam evaluation of LUS imaging.
Rabinovich et al. 2022 (128)	Argentina	Imaging: Chest X rays	Radiology Department	Diagnosis	TRx: is an AI application that assists users in chest x-ray interpretation. It combines four deep learning models that were trained for the detection of four critical findings: pneumothorax, rib fracture, pleural effusion, and lung opacities. In the interface, user feedback can be optionally completed at the time of image evaluation.	Observational, single centre. *Description:* Observational, mixed methodology user experience study. User satisfaction questionnaires based on the four factors of the Technology Acceptance Model and System Usability Scale. Qualitative feedback via six physician interviews.	Rabinovich et al. 2022 used the Technology Acceptance Model to evaluate actual use and satisfaction with an AI system for the automated detection of findings in chest x-rays, after 5-months of use at an Argentinian Emergency Department. The system was used for 15% of studies (n=1186), with an average of 8 accesses per day. Emergency physicians and radiology residents shared perceptions about the usability of the system, while differing on output quality and usefulness for their work.

References

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Sarti et al. 2021 (137)	Canada	Respiratory - Extubation assessment	Intensive Care Unit	Treatment	The Extubation Advisor (EA) tool standardises and optimises the Spontaneous Breathing Trial (SBT), one of many methods to assess extubation readiness. The web-based is a predictive model of the risk of extubation failure, RSBI, clinical impression of extubation failure risk and standardised extubation readiness checklist to generate a report to assist extubation decision making.	Observational, single centre. *Description:* Model performance, feasibility evaluation, and qualitative feedback (interview of clinicians and questionnaires). Sarti et al. 2021 enrolled 117 patients, totalling 151 SBTs and 80 extubations.	The incidence of extubation failure was 11% in low-risk patients and 21% in high-risk patients stratified by the predictive model; 38% failed extubation when both the model and clinical impression were at high risk. The tool was well rated: 94% and 75% rated the data entry and EA report as average or better, respectively. Interviews (n=15) revealed favourable impressions regarding its user interface and functionality, but unexpectedly, also concerns regarding EA's potential impact on respiratory therapists' job security.
Schmuelling et al. 2021 (138)	Switzerland	Imaging: CT (pulmonary embolism)	Radiology and Emergency Department	Triage	Aidoc DL-powered algorithm: Detect and alert radiologists about cases with suspected pulmonary embolism on CT pulmonary angiograms (CTPA) along with annotated images, using Electronic Notification System (ENS).	Observational, single centre. *Description:* Observational single site implementation case study. Study team extracted all CTPAs between April 2018 and June 2020 to establish how each exam was communicated back to the referring physician. Primary outcome was Report Communication Time. There were three distinct time periods: 'baseline' i.e. pre-AI, then 'Light Messenger only' then 'LM+DL Algorithm' all about nine months each in duration. Other outcomes were 'Turnaround Time' and 'Time to anticoagulation'.	Schmuelling et al. 2021 assessed the impact of the implementation of an electronic notification system and AI algorithm for automated detection of on CT pulmonary angiograms. While the study demonstrated good diagnostic accuracy of the AI after clinical implementation (sensitivity 80%, specificity 95%, PPV 82%, and NPV 94%) there was no statistically significant effect on report communication times and patient turnaround in a Swiss emergency department nine-months after technical implementation.

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Sepsis (n=5)							
Adams et al. 2022 (173)	USA	Sepsis	Embedded in EHR	Diagnosis	Targeted Real-Time Early Warning System (TREWS) identify and notify clinicians about patients at risk of sepsis using eHR data.	**Observational, multi-centre.** *Description:* Conducted at five centres, this is a prospective two arm cohort study from a population of 0.5mil patient encounters, where there were 6877 patients with sepsis identified by the alert before initiation of antibiotic treatment. The study group contained patients who had the alert confirmed by a clinician within three hours (n=4220, of which 1430 were high risk), vs the comparison group who did not have the alert confirmed within three hours (n=2657, of which 935 were high risk).	Adjusting for patient presentation and severity, the study group patients had a reduced in-hospital mortality rate (3.3% CI 1.7%, 5.1% adjusted absolute reduction, and 18&%, CI, 9.4, 27% adjusted relative reduction), organ failure and length of stay compare with patients in the comparison group (who's alert was not confirmed within 3 hours). Improvements in mortality rate (4.5%, CI 0.8, 8.3%, adjusted absolute reduction) and organ failure were larger among those patients who were additionally flagged as high risk.
Boussina et al. 2023 (175)	USA	Sepsis	Hospital Wide	Diagnosis	Predictive analytics platform: purpose of the cloud-based platform is to process eHR data on any patient and provide real-time recommendations to clinicians natively within the eHR.	**Observational, site number unspecified.** *Description:* This is a use case of leveraging this platform. The research team deployed a Deep Learning Model for the early prediction of sepsis onto the platform and into clinical practice.	Largely a summary of their experience: explained the platform architecture, cloud implementation, data pipelines, and then described the sepsis model COMPOSER that was deployed silently for six months (alerts not displayed to clinicians) to determine indications for use of the algorithm. This was followed by a three month period of design sessions with nursing teams to build a display. Prospective validation of the model followed. Apart from model performance (e.g. positive predictive values) they also described how the model's performance would be tracked weekly for model drift, metrics for system downtime and uptime, and the number of patient hours processed in a seven month period (1.3 mil)

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Ericson et al. 2022 (119)	Sweden	Sepsis	Intensive Care Unit	Diagnosis	NAVOY Sepsis prediction algorithm bases its prediction on variables routinely collected at ICUs. Validated in a prospective RCT and is CE marked as a SaMD.	**Health Economics Research, unknown number of sites.** *Description:* Health economics study on short term (<1yr) and long-term effects of sepsis, looking at NOVAY cost-saving potential.	Under these assumptions, an ML algorithm that can detect sepsis three hours before current practice will reduce the cost per ICU patient by 0.5%. The total cost per patient with such an algorithm is €16 436, and the cost per patient for current practice is €16 512. The potential cost savings per patient is thus €76, and the aggregated yearly cost saving for the Swedish healthcare system is €2 798 915. The largest cost savings are due to a shorter average length of stay in the ICU (0.16 days shorter for an algorithm like NAVOY® Sepsis compared with current practice), resulting in a cost saving of 8.9% (€10 322 vs €11 331) per patient related to ICU hospitalisation. The shorter length of stay in ICU for the sepsis prediction algorithm compared with current practice results in 5860 fewer ICU days per year on an aggregated national level. In addition to the reductions in resources used, faster detection also implies reduced in-hospital mortality, resulting in 356 lives saved per year in Sweden alone.
King et al. 2022 (129)	USA	Sepsis	Neonatal Intensive Care Unit	Diagnosis	HeRO: monitoring during NICU stay predicting mortality or neurodevelopmental impairment. Feature of interest in the Heart Rate Characteristics (SA node changes) which change in the presence of cytokines - in the lead up to infection.	**Interventional, multi-centre.** *Description:* 3,003 very low birthweight infants at eight study centres were randomised to receive either standard of care monitoring, or standard of care monitoring plus HeRO. It was a pragmatic study design, meaning that there were no mandatory interventions based on the HeRO Score. Instead, the HeRO Scores were displayed to the clinicians for half the patients, and then outcomes were tracked.	Among all patients in the RCT, those randomised to HeRO display experienced a 22% reduction in all-cause mortality (Number Needed to Treat to save one life: 48). There was no significant increase in testing or antibiotic usage. While clinicians were only able to see and act upon HeRO Scores for half the patients in the RCT, HeRO Scores were generated but not displayed for the other half. From this, we can see that HeRO Scores were significantly higher in patients randomised to non-display for a full week prior to the overt clinical deterioration that prompted the blood culture.

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Lipatov et al. 2022 (148)	USA	Sepsis	Intensive Care Unit and Emergency Department	Diagnosis	EMR based sepsis surveillance system (Sepsis Sniffer) augmented with CDS and completion feedback. The algorithm included a treatment failure component aimed at recognising patients lacking appropriate management and computerised sepsis treatment support system.	**Observational, single centre.** *Description:* Retrospective observational before and after study. The primary outcomes were sepsis care bundle compliance (all or none) and completion of individual components of sepsis management. Secondary outcomes included mortality as well as hospital and ICU length of stay.	Assuming concordance of the positive alerts and severe sepsis recognition, the performance of the sepsis alert could be calculated with sensitivity of 79.9% (95% CI 77.5% to 82.2%) and specificity of 80% (95% CI 76.1% to 77.8%) and the positive and negative predictive values 28% (95% CI 27.0% to 28.9%) and 97.2% (95% CI 96.8% to 97.5%), respectively. There were 3424 unique alerts and 1131 confirmed sepsis patients after sniffer implementation. Average care bundle compliance was higher; however, after taking into account improvements in compliance leading up to the intervention, there was no association between intervention and improved care bundle compliance. Similarly, the intervention was not associated with improvement in hospital mortality (odds ratio: 1.55; 95% CI: 0.95 to 2.52; p-value: 0.078).
Sleep Disorder (n=1)							
Hwang et al. 2022 (145)	South Korea	Sleep Disorders	Neurology Department	Diagnosis	A CDSS that automatically score sleep studies from EEG patterns and other physiological data collected during sleep studies (Polysomnography).	**Observational, multi-centre.** *Description:* User-centred design phase. Then assessed sleep staging performance under two settings. The first was sleep scoring using the CDSS against the baseline AI, where technicians scored stages with AI systems that included only AI predictions provided without any explanation. The second was sleep scoring using our CDSS versus a conventional setting, where technicians need to score each epoch without the predictions by AI. Configured the baseline AI and conventional settings to compare sleep staging settings for our CDSS.	Hwang et al. 2022 describe their experience in using an iterative, user-centred design process with sleep technicians (9) to develop clinical sound explanations for AI that automatically scores sleep studies. Evaluation study on nine polysomnographic technicians quantitatively and qualitatively investigated the helpfulness of the tool. For technicians with <5 years of work experience, their quantitative sleep staging performance improved significantly from 56.75 to 60.59 with a P value of .05. Qualitatively, participants reported that the information provided effectively supported them, and they could develop notable adoption strategies for the tool.

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Stroke (n=11)							
Chien et al. 2022 (174)	Taiwan	Imaging: CT (Intracranial Haemorrhage)	Emergency Department	Diagnosis	Deep CT: A deep CNN which can identify haemorrhagic lesions from non-contrast head CTs (NCCT).	**Observational, single centre** *Description:* Before and After, Retrospective, Observational study. Non-controlled pilot trial: January to April 2020 physicians read NCCTs without Deep CT. May to August 2020 NCCTs were read with DeepCT assistance. 2999 patients in total.	DeepCT diagnosed ICH significantly shortened the Length of Stay (560.67 ± 604.93 min with DeepCT vs. 780.83 ± 710.27 min without DeepCT; p = 0.0232). When the diagnosis was not intracranial haemorrhage (ICH), the LOS did not significantly differ before and after implementing the DeepCT system (705.90 ± 760.86 min with DeepCT vs. 679.45 ± 681.97 min without DeepCT; p = 0.3362). Reported back on model performance (e.g. specificity, sensitivity, accuracy etc.). The calculated personnel costs per patient bed-hour were calculated to be $58.20 for an urban academic medical centre ED in eastern US. Accordingly, the LOS shortened by approximately 3.67 hours after implementing the DeepCT system in the ED for ICH patients, suggesting an approximate cost saving of $210 per patient for the hospital.
Elijovich et al. 2022 (183)	USA	Large Vessel Occlusion Acute Ischaemic Stroke	Stroke Centre	Diagnosis	VizAI: cloud-based technology post-processing of DICOM images. Runs in parallel with PACS, images automatically transferred to Cloud based AI algorithm trained to detect LVO AIS from CT Angiography and generates automated alerts via it's secure messaging platform allowed for communication by the entire care team.	**Observational, multi-centre.** *Description:* Retrospective chart reviews of ELVO patients, either AI detected or detected by usual care, at a comprehensive stroke centre and two of its spoke hospitals, impact on stroke workflow metrics. Primary outcome of the study was effect of VizAI on stroke workflow: 1) comparison of treating team notification by SoC compared with VizAI notification. 2) CTA2AP (CTA to arterial puncture) comparison between SoC and VizAI 3) Door to Arterial Puncture (DAP) time comparison between SoC and VizAI.	Results: 45 pts with ELVO identified by AI, 59 by usual care. The CTA to treatment team notification times were significantly faster for AI than with usual care notification by the neuroradiologist for all ELVOs (7 min vs 26 min; p<0.001). For patients presenting to the hub hospital, AI notification would be expected to be 10 min faster (β coefficient, −10.7; 95% CI −21.4 to −0.088; p=0.048) than traditional notification. DAP (141 vs 185 min; p=0.027) and CTA2AP (101 vs 164 min; p=0.009) were both significantly shorter for patients transferred from a spoke hospital when the ELVO was detected by AI and would yield an approximate 23 (β coefficient −23.1; 95% CI −40.7 to −0.001; p=0.049) and 33 (β coefficient, −32.7; 95% CI −51.6 to −6.68; p=0.019) minute time savings per patient.

References

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Gunda et al. 2022 (120)	Hungary	Stroke	Stroke Centre	Diagnosis	e-Stroke Suite - characterises ischemic regions from no-contrast CT and identify occlusions on CT angiography. It is CE-marked software package.	**Observational, single centre.** *Description:* Observational Retrospective chart review. Gunda et al.'s examined automated analysis of CT angiography at a primary stroke centre in Hungary. Study outcomes included: number of patients receiving intravenous thrombolysis and/or thrombectomy, the time to treatment; and outcome at 90 days for thrombectomy.	Use of the system over a 7-month period with 399 patients was reported to increase thrombolysis rates 11% to 18% and thrombectomy (2.8–4.8%). There was a trend towards shorter door-to-needle times (44–42 min) and CT-to-groin puncture times (174–145 min). There was a non-significant trend towards improved outcomes with thrombectomy. Among physicians the system was perceived to increase decision-making confidence and improved patient flow.
Hu et al. 2022 (135)	China	Imaging: CT (Cerebral Infarct)	Radiology Department	Treatment	The Deep CNN algorithm DLR was used to process CT perfusion images (de-noise) thereby improving the effectiveness and safety of the treatment for acute cerebral infarction.	**Observational, unknown number of sites.** *Description:* Prospective, diagnostic study. 100 patients divided to Algorithm group vs conventional group. Unspecified numbers of study sites involved. Unspecified follow up period. Efficacy evaluation: At one day, one week, half a month, and one month after the thrombolytic therapy, the NIHSS score was recorded, which was used as the evaluation standard of thrombolytic effect, and patients were evaluated for the brain nerve defects. Safety Evaluation. After thrombolytic therapy combined with the NIHSS score (above 4 points), as well as CT or MRI examination, whether there is symptomatic cerebral haemorrhage is used as a safety evaluation index. In addition, the intracranial haemorrhage rates before and after thrombolytic therapy were compared.	The study reported improvements in image quality. The differences in the National Institute of Health stroke scale (NIHSS) scores for the two groups indicated that the thrombolytic effect on the algorithm group was superior to that on the control group. Thrombolytic therapy for the algorithm group showed therapeutic effects on neurologic impairment. The symptomatic intracranial haemorrhage rate of the algorithm group within 24 hours was lower than the haemorrhage conversion rate of the control group, and the difference between the two groups was 14%. The data differences between the two groups showed statistical significance ($P < 0.05$).

References

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Kotovich et al. 2023 (182)	Israel	Imaging: CT (Intracranial Haemorrhage)	Emergency Department	Diagnosis	AI based computer-aided triage and prioritisation solution for the detection of all types of ICH in radiology. All relevant CT studies are automatically sent for AI analysis with no manual trigger. Upon detection of suspected positive ICH findings, the AI solution delivers notifications directly to the radiologist workstation.	Observational, single centre.	A significant decrease in the 30-day mortality rate was observed in the post-AI group compared to the pre-AI group (pre-AI 27.7% vs post-AI 17.5%, odds ratio = 0.48, CI of odd 0.29 to 0.79, p = 0.004), and a significant decrease in the 120-day mortality in the post-AI group in comparison to the pre-AI group (pre-AI 31.8% vs. post-AI 21.7%, odds ratio 0.58, CI of odds 0.37 to 0.91, p = 0.017) was observed for the ICH dataset. A sub-analysis of mortality rates in the ICH dataset with respect to anticoag and anti-aggregation treatment revealed a decrease in mortality in either at 30 days and 120 days
Martinez-Gutierrez et al. 2023 (114)	USA	Large Vessel Occlusion Ischaemic Stroke	Stroke Centre	Diagnosis	Cloud-based AI-algorithm (Viz.AI) trained to detect Large Vessel Occlusion, Acute Ischaemic Stroke. Analyses CTA images and decides presence or absence of LVO within minutes. The decision is transmitted to a mobile phone application, which the clinical care team was required to download on to their phones and arrived in the form of a pushed alert notification. Within the application, a mobile picture archiving and communication system (PACS) allowed users to verify imaging findings and a secure messaging platform allowed for communication by the entire care team.	Interventional, multi-centre. *Description:* Randomised, stepped wedge clinical study design overcomes the impracticality of randomising at the individual patient level but retaining a robust means to evaluate this intervention. Randomised four comprehensive stroke centre (CSC) hospitals to initiate LVO detection software in pre-determined stepped-time intervals and hypothesised that initiation of this intervention would result in a decrease in Door To Groin (D2G) time in patients with LVO AIS. LVO-AIS numbers: 131 patients pre-AI, nine in transition period and 103 post AI.	D2G time (time it takes to go from hospital arrival to initiating endovascular thrombectomy) was reduced by 11.2 minutes in the post AI cohort. Time from arrival to IV tPA bolus did not change between the cohorts. Time from CT to start of EVT was reduced (9.8 mins.) LOS did not change, neither did the safety outcomes other than mortality, which decreased post-AI. In exploratory analyses on the impact of the software intervention on clinical outcomes, rates of functional independence at 90 days (mRS,0-2) were similar in univariable comparisons of the pre-AI and post-AI cohorts (32% vs 42%, pre-AI vs post-AI; P=.47). In multivariable logistic regression adjusted for age, NIHSS and ASPECTS, there was no observable difference in likelihood of 90-day disability (mRS,0-2) in the post-AI cohort relative to pre-AI (oddsratio,1.3; 95%CI,0.424.0). Similarly, there were no differences in rates of good functional outcomes at discharge defined as mRS0-2 (28% vs 41%, pre-AI vs post-AI; P=.17).

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Seyam et al. 2022 (177)	Switzerland	Imaging: CT (Intracranial Haemorrhage)	Intensive Care Unit	Diagnosis	Aidoc: AI triage of urgent cases via flags and widgets, annotates CT images and communicates in PACS, and is an automated second-read QA checker to avoid diagnostic misses	**Observational, single centre**. *Description:* Prospective diagnostic study. Tested diagnostic performance of the AI-based tool for ICH on prospectively acquired CT images, compared clinical workflow metrics pre and post AI implementation.	3017/4450 patients CT scanned after AI implementation. F1 score of 0.78, accuracy of 93.0%, sensitivity of 87.2%, specificity of 93.9%, positive predictive value of 70.5%, and negative predictive value of 97.8%. We observed high overall detection rates for intraventricular haemorrhage (97.1%, 34 of 35) but lower rates for subarachnoid haemorrhage (173) (80.0%, 36 of 45) and subdural haemorrhage (69.2%, 74 of 107). Workflow metrics as pre-AI versus post-AI implementation, respectively: overall communication time of ICH (70 minutes [95% CI: 54, 85] vs 63 minutes [95% CI: 55, 71]), during regular working hours (96 minutes [95% CI: 68, 123] vs 78 minutes [95% CI: 63, 93]), communication time of acute ICH (73 minutes [95% CI: 49, 97] vs 58 minutes [95% CI: 48, 68]), and overall consultation time (166 minutes [95% CI: 98, 233] vs 163 minutes [95% CI: 55, 272]). ED turnaround time for ICH exclusion, particularly during regular working hours (205 minutes [95% CI: 180, 230] vs 167 minutes [95% CI: 154, 181]), is expedited.
Van Leeuwen et al. 2021 (122)	UK	Large Vessel Occlusion Ischaemic Stroke	Health economics study but based on available literature	Diagnosis	Any AI (such as Aidoc) that assist with LVO detection.	**Health Economics Research, unknown number of sites.** *Description:* Health economics study but based on available literature. Cohort predominantly UK stroke registry data. Modified Rankin Score data came from five RCT studies, costs from another study, health outcomes from a RCT.	No costs calculated for the innovation of the AI. For the projected lifetime per ischemic stroke patient, the incremental costs and incremental efficacy were − $156 (− 0.23%) and + 0.0095 QALYs (+ 0.07%) respectively. Using the reference value of $25,662 per QALY, 0.0095 QALY would translate to $244. For each yearly cohort of patients in the UK this translates to a total cost saving of $11 million and QALY gain of 682 ($17.5 million).

References

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Wang et al. 2023 (184)	Australia	Imaging: CT (Intracranial Haemorrhage)	Emergency Department	Diagnosis	Veriscout: an artificial intelligence-based CT haemorrhage detection and triage tool. Triages head CT scans acquired in ED that have a high likelihood of haemorrhage and flags the scans for expedited reporting, via its integration with the Radiology Information System (RIS) and notification in existing clinical system.	**Observational, single centre.** *Description:* Observational, Retrospective cross sectional study analysis of Veriscout performance reading 527 CT head scans. Ground truth was by expert consensus.	527 CT scans read by Veriscout and Expert consensus panel. They found 79 scans with evidence of haemorrhage. For all scans, VeriScout™ detected haemorrhage with a sensitivity of 0.92 (CI 0.84–0.96) and a specificity of 0.96 (CI 0.94–0.98) using the expert consensus as ground truth. VeriScout™ returned a result to the RIS within 10 min in 100% of cases analysed; and appropriately flagged all positive cases as determined by the algorithm. Upload speed from the hospital network to the cloud analysis server was the primary determinant, with inference completed in less than 1 min in all cases. Integration with the PACS was confirmed by the presence of an appropriate VeriScout™ image(s) in the relevant scan session; a technical misconfiguration prevented initial processing in 11 cases, but this was detected in real-time by the informatics platform and all cases were subsequently re-triggered successfully.
Yahav-Dovrat et al. 2021 (212)	Israel	Large Vessel Occlusion Ischaemic Stroke	Stroke Centre	Diagnosis	Viz LVO: Analyse computed tomography angiograms (CTAs) images, notifying cases with suspected positive findings of Large Vessel Occlusion (99).	**Observational, single centre.** *Description:* Observational, retrospective model accuracy study. Viz LVO deployed January 2018. All head and neck CTAs were scanned by the algorithm. System results compared to the formal reports (ground truth) for presence of LVO.	Yahav-Dovrat et al. 2021 evaluated the detection accuracy of an AI algorithm to detect large-vessel occlusions on CTA's and notify the treatment team in real-time via a dedicated mobile application at a stroke centre in Israel. The system was found to be highly accurate when used to scan all head and neck CTAs over a 15-month period. 75 LVOs ground truth vs 61 detected by Viz AVO. Model performance parameters (specificity etc) reported.

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Zia et al. 2022 (213)	Australia	Imaging: CT (Intracranial Haemorrhage)	Radiology Department	Diagnosis	Aidoc: AI triage of urgent cases via flags and widgets, annotates CT images and communicates in PACS, and is an automated second-read QA checker to avoid diagnostic misses	**Observational, single centre.** *Description:* Mixed methods: retrospective application of Aidoc and then Prospective diagnostic study. study aims to first apply the Aidoc ICH detection algorithm retrospectively, on ICH-negative studies, to assess the pre-implementation radiologist's miss-rate, and to determine whether these were clinically significant to patient care. Second, to evaluate the prospective diagnostic accuracy of Aidoc's ICH detection algorithm at the same tertiary hospital, in terms of diagnostic accuracy and changes in turn-around-time (133).	Looking at the prospective validation only: 212/1446 head CTs had ICH. Aidoc flagged 220, of which 180 were TP, 30 were false neg. The diagnostic accuracy of the software for all cases was as follows: sensitivity 85.7% (95% CI 80.3–90.2%); specificity 96.8% (95% CI 95.6–97.6%); PPV 81.8% (95% CI 76.8–86.0%), NPV 97.6% (95% CI 96.6–98.2%). For all ICH-positive scans, the mean pre-implementation TAT was 66.7 (SD 41.5) minutes, and the post-implementation TAT was 80.0 (SD 54.25) minutes. There was a decrease in TAT for ICH-positive scans in the emergency and outpatient cohorts by 3.7 min (– 5.1%) and 9.9 min (– 14.2%), respectively, not statistically significant. Out of 49 consultant radiologists and registrars, 26 responded to the survey. Three radiologists used Aidoc 100% of their reporting time; three 75%; four 50%; seven 25% and nine 0%.
Triage (n=4)							
Ivanov et al. 2021 (149)	USA	Triage	Emergency Department	Triage	KATE analyses EHR data to estimate patient acuity - ESI scale (1-5)	**Observational, multi-centre.** *Description:* Observational, retrospective validation study. The purpose of this retrospective study was to determine whether historical eHR data can be used with clinical NLP and ML algorithms (KATE) to produce accurate emergency severity index predictive models. Two hospitals involved.	Impact of this program on emergency nurse triage decisions reported an overall improvement in triage accuracy from 54% to 67% for paediatric patients and from 62% to 78% for adult patients.

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Jordan et al. 2023 (153)	USA	Triage: Emergency Medicine	Emergency Department	Triage	KATE is an AI that implements the most widely used triage system in the USA - the Emergency Severity Index - comprising of five mutually exclusive acuity levels. KATE offers a clinical decision on patient acuity that incorporates elements of the ESI, current presentation, and medical history. When the KATE acuity decision does not match the nurse's ESI determination, nurses have options to expand documentation, verify or change the acuity level, or ignore/bypass the suggestion.	**Observational, single centre.** *Description:* Exploratory qualitative study. The overarching research question was "What were the processes by which emergency department triage nurses understood, contextualised, and incorporated a new clinical decision support aid into their understanding and practice of triage?"	13 emergency department nurses participated in interviews that were underpinned by the Campinha-Bacotes model of competence in healthcare delivery. Thematic analysis using NVivo yielded the following. 1: the value of cultural embeddedness impacting the way a nurse triages "I am Hispanic, this hospital is in a latino community, cultural background impacts the way you triage. 2: Just another checkbox: scepticism because KATE 'cannot see' the patient. 3: Gut trumps Data. Many of the nurses believed having a computer program provide a recommended ESI level reduced the drive for nurses in the ED to refine their analytic skills and maintain cultural competence. 4: Higher acuity with no resources. Spike in sepsis alerts since the introduction of KATE. This aspect was generally positively received by the 13 participants. 5: Technology as a safety net. Many of the participants commented that the feedback from the AI program caused them to re-examine the thinking behind their initial triage and be more conscious of any assumptions that may have influenced their approach. Other issues around implementation were raised.
Soltan et al. 2022 (73)	UK	Triage: COVID-19	Emergency Department	Diagnosis	Identify patients attending ED with COVID-19 using routine blood test, blood gas, and vital signs collected within 1 hour of presentation to hospital.	**Observational, multi-centre.** *Description:* External validation of ML models. Aim to shorten the time between arriving at ED and receiving COVID-19 screening result generated by the AI model. Two-part study: External prospective validations of 3 AI models (multi-site across four UK NHS Trusts), compared against Lateral Flow Devices (LFDs). Best model was then deployed at one ED.	Automated identification using routinely collected clinical data was reported to detect COVID-19 in 45 min, 61 min sooner than a lateral flow device, and 6 h 52 min (90%) sooner than with PCR. Classification performance was high (sensitivity 87%; specificity of 85%, and negative predictive value 100%). The AI system correctly excluded infection for 31 (58%) of 53 patients who were triaged by a physician to a COVID-19 suspected area but went on to test negative by PCR.

Author and Year	Country	Clinical Area	Setting	Task supported	ML system task	Study design & Description	Key Findings
Wang et al. 2022 (152)	Taiwan	Triage: Chest pain	Emergency Department	Triage	AI based triage system: detect ST-elevation myocardial infarction (STEMI) on electrocardiography (ECG), and a computerised risk score provide a clinical risk score (ASAP) to prioritise patients for ECG examination.	**Observational, single centre.** *Description:* Observational, before and after study. The purpose of the study was to compare total Door 2 Balloon (D2B) times and individual components of D2B time between patients with STEMI enrolled before and after introducing the AI-based triage system.	Wang et al. demonstrated the impact of an AI system in improving clinical decision-making and triage of chest pain in in a Taiwanese emergency department. Automated detection of ST-elevation myocardial infarction (STEMI) on electrocardiography (ECG) and assessment of clinical risk (ASAP score) was reported to shorten the time to treatment (door-to-balloon time 64 min to 53 min). Among patients with ASAP score of 3 or higher, the median door-to-ECG time decreased from 30 min to 6 minutes.
Wound Management (n=1)							
Howell et al. 2021 (140)	USA	Wound Management	Wound Care Centres	Diagnosis	AI based wound assessment tool to enhance accuracy and consistency of wound area and percentage of granulation tissue from photographs taken by wound clinicians.	**Observational, multi-centre.** *Description:* Multi-centre, prospective diagnostic study to evaluate the AI-based would assessment tool. Statistical comparison of error measure distributions between AI traces and reference human traces (human vs AI), with error distributions between two humans (human vs human). For each photograph, each of the human tracings served as both reference and test, resulting in four test vs reference comparisons: AIvsH1, AIvsH2, H1vsH2, and H2vsH1. To quantify error measures, ROIs for each image were imported into ImageJ and the AND command was used to create a new ROI for the overlapping regions within the test and reference traces.	Howell et al. 2021 evaluated the performance of AI-based software for wound assessment against manual wound assessments performed by wound care clinicians. While AI-based wound annotation algorithms perform similarly to human wound specialists (the comparisons were found to not be statistically significant), the degree of agreement regarding wound features among expert physicians can vary substantially, presenting challenges for defining a criterion standard. False Negative area (FNA) was slightly elevated AI vs Human compared to human vs human FNA - AI slightly underestimate the wound boundary.

Appendix G: Studies reporting effects of AI problems on care delivery and patient outcomes

Authors (Year)	Study period (no. of months)	Study design and methods	Country	Sample	AI system/s	Setting	Key findings
Beede et al. 2020 (196)	2018–19 (3)	Prospective: descriptive study of model performance, observations, interviews (mixed-methods)	Thailand	50 patients, five nurses and one camera technician.	Deep learning algorithm	Diabetic retinopathy clinic	*Data input issues:* Out of 1838 fundus images that were entered into the system 393 (21%) were poor quality and did not meet the system's high standards for grading. Noticeable consequence but no patient harm: Ungradable images had to be re-taken, frustrating nurses and patients.
Eng et al. 2021 (118)	2018-19 (11)	Prospective: RCT (quantitative)	USA	93 radiologists at six centres without (n = 739 radiographs) and with (n = 792) an AI algorithm.	Deep learning model for assessment of skeletal age from hand x-rays	Six radiology departments	*Use error (automation bias):* The AI algorithm resulted in higher diagnostic error when inaccurate AI predictions were presented to radiologists in the AI-assisted group compared with when inaccurate predictions were not presented to them in the control group (absolute difference in skeletal age compared to gold standard 10.9 months [AI] vs 9.4 months [control]; P = .06).
Wong et al. 2021 (195)	2018–19 (11)	Retrospective: descriptive study of model performance (quantitative)	USA	27 697 patients	Epic Sepsis Model	Academic health system	*Algorithm issue (distributional shift):* The Epic Sepsis Model performed substantially worse in real-world use (AUC, 0.63) than claimed by the manufacturer (AUC, 0.73–0.83). *Disrupted care delivery:* Generated alerts for 18% of all 38 455 hospitalised patients. *Potential or actual harm to a patient:* Identified only 7% of 2552 patients with sepsis who were not treated with antibiotics in a timely fashion; failed to identify 1709 patients with sepsis that the hospital did identify.
Wong et al. 2021 (194)	2019–20 (5)	Retrospective: descriptive study of model performance (quantitative)	USA		Epic Sepsis Model	24 hospitals across 4 health systems	*Algorithm issue (distributional shift):* In the weeks following the first COVID-19 hospitalisations, sepsis alerts more than doubled from 9% (953 of 10,159) to 21% (1363 of 6634). Presence of the virus made it difficult for the algorithm to differentiate bacterial sepsis from COVID, thereby limiting the usefulness of alerts.
Daneshjou et al. 2022 (193)	2010–20 (120)	Retrospective: descriptive study of model performance (quantitative)	USA	656 images from Diverse Dermatology Images Dataset	Three algorithms: ModelDerm, DeepDerm and HAM10000	Dermatology clinic	*Algorithm issue (bias):* Limitations on detecting lesions on dark skin tones and uncommon diseases. Consequences not reported

Authors (Year)	Study period (no. of months)	Study design and methods	Country	Sample	AI system/s	Setting	Key findings
Glissen Brown et al. 2022 (170)	2019–2020 (18)	Prospective: single-blind RCT (quantitative)	USA	223 patients	Deep learning: computer aided lesion detection in colonoscopy	4 academic medical centres	*Algorithm issues:* 203 false positives and three false negatives i.e. polyps detected by the endoscopist that were not recognised by the AI. No immediate adverse events were reported. *Hazard*
Kanbar et al. 2022 (116)	2016–18 (34)	Prospective: descriptive study of model performance (mixed-methods)	USA	Paediatric epilepsy clinic and emergency department patients	1. Ensemble ML system to identify epilepsy patients for surgery 2. Machine learning system to screen emergency department patients for clinical trial eligibility	Children's hospital	*Date input issues* 1. Issues extracting patient notes from the electronic health record delayed running of the epilepsy system in 12 out of 150 (8%) weeks of operation. 2. Updates to the electronic health record and supporting IT infrastructure caused multiple breakdowns interrupting less than 2 out of 52 weeks of operation. *Noticeable consequence but no patient harm*
Lyell et al. 2023 (191)	2015–21 (82)	Retrospective: incident analysis (qualitative)	USA	266 safety events reported to the US Food and Drug Administration	25 ML-enabled medical devices	All	*AI safety problems:* Safety events involving ML-enabled medical devices arose from: data input issues: 82% (n=219) algorithm issues:11% (n=28) use errors: 4% (n=11) contraindicated use: 3% (n=7) data output issues: <1% (n=1) *Consequences:* Safety events included hazards with potential to harm (66%), actual harm (16%), consequences for healthcare delivery (9%), near misses (4%), no harm or consequences (3%), and complaints (2%). While most events involved device problems (93%), use problems (7%) were 4 times more likely to harm (relative risk 4.2; 95% CI 2.5–7).

Authors (Year)	Study period (no. of months)	Study design and methods	Country	Sample	AI system/s	Setting	Key findings
Tierney et al. 2024 (61)	2023–24 (2.3)	Prospective: pilot study (mixed-methods)	USA	35 AI-generated transcripts	Proprietary AI scribe	Diverse settings across health system	Pilot deployment of ambient AI scribe technology to >9,000 clinicians. AI-generated transcripts scored an average of 48 out of 50 in 10 key domains. *Algorithm issues (hallucination):* False information provided by AI without sound basis including: • Clinician mentioned scheduling a prostate examination for the patient and the AI scribe summarised that a prostate examination had been performed. • Clinician mentioned issues with the patient's hands, feet, and mouth and the AI summary recalled the patient being diagnosed with hand, foot, and mouth disease. • Summary was missing some details, such as missing chest pain and anxiety assessments. • Summarised clinical content was not consistent with pre-existing note templates, resulting in inconsistencies in summarisation. Consequences not reported. The study concluded that AI scribes were not a replacement for clinicians as they could produce inconsistencies that required review and editing to ensure that they remain aligned with the doctor–patient relationship.