*Journal of the American Medical Informatics Association*, 2025, **32(11)**, 1767–1777
https://doi.org/10.1093/jamia/ocaf167
Advance access publication 15 October 2025
**Review**

Review

# The real-world impact of artificial intelligence ethics frameworks across a decade in healthcare: a scoping review

**Anastasia Chan, MRes[1], Hania Rahimi-Ardabilli, PhD[1], Wendy A. Rogers, PhD[2], Enrico Coiera, PhD*[,1]**

[1]Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Sydney, NSW 2109, Australia, [2]School of Humanities and School of Medicine, Macquarie University, Sydney, NSW 2109, Australia

*Corresponding author: Enrico Coiera, PhD, Centre for Health Informatics, Australian Institute of Health Innovation, Level 6, 75 Talavera Road, Macquarie University, Sydney, NSW 2109, Australia (enrico.coiera@mq.edu.au)

## Abstract

**Objectives:** The number of ethical frameworks designed to guide artificial intelligence (AI) use has grown substantially over the past decade, yet their real-world effect remains unclear. We aimed to synthesize existing evidence to analyze the practical impact of AI ethics frameworks (AIEFs) operationalized in healthcare.

**Materials and Methods:** We conducted a scoping review across 4 academic databases (Ovid MEDLINE, Ovid Embase, Scopus, and Web of Science), Google, and Google Scholar from January 2014 to January 2025. Eligible studies reported primary research on the qualitative or quantitative impacts of AIEFs implemented in healthcare. Data synthesis was conducted via narrative review.

**Results:** Of 1807 records identified, 16 studies met inclusion criteria. These comprised 5 preliminary initiatives testing guidelines in practice, 5 case studies, 5 implementation studies, and a comparative case study. AIEFs were implemented: (1) to develop new AI governance structures and guidelines, (2) as ethical review assessment systems for adopting clinical AI technologies, and (3) as ethical "audit" tools for identifying ethical risks. Impact was reported through qualitative improvements to process measures such as improved trust in AI. No studies demonstrated a direct link between AIEFs and health-related outcome measures such as patient safety.

**Discussion:** AIEFs led to changes in organizational or clinical processes, including increased compliance with ethical standards. When embedded in governance, AIEFs improved oversight and evaluation, but audits were constrained by their reliance on organizational cooperation.

**Conclusion:** Despite the proliferation of AIEFs over the past decade, their implementation in healthcare remains limited and impact on health outcomes unmeasured or underreported.

**Key words:** ethics; artificial intelligence; healthcare; frameworks; implementation; governance.

## Introduction

Advances in artificial intelligence (AI) systems have led to a boom in AI ethics frameworks (AIEFs) and guidelines over the last decade. As AI becomes more pervasive, so does its ability to negatively impact lives and livelihoods; for instance, by leading to unfair or discriminatory outcomes or by compromising confidential patient data.[1,2] In this context, AIEFs have emerged to ensure that high-level ethical principles—such as transparency, fairness, and beneficence—are adhered to across the AI lifecycle and translated into real-life practice. This increased attention on AI ethics has been called "something of a gold rush," with frameworks appearing from companies, professional organizations, academia, governments, and international bodies.[2] A meta-analysis[3] in 2023 found 200 guidelines and recommendations for AI governance published worldwide (see also Figure 1 and Appendix S1 where we conducted a preliminary search for the keywords "artificial intelligence," "ethics," and "frameworks," "guidelines," or "principles" contained in titles across 4 medical and general research databases [Ovid MEDLINE, Ovid Embase, Scopus, and Web of Science] and found 136 new frameworks from 2014 to 2024—with an upwards trend from 2018). To promote ethical practice, AIEFs have been used to structure new AI governance and review boards[4] and assess AI models before and after deployment.[5,6]

Unsurprisingly, AIEFs have also proliferated in healthcare, given the potential for AI use to impact patient welfare. An ethically justifiable AI application might, for example, provide substantial benefits, such as an AI model for identifying sepsis that provides more accurate diagnoses than human-only interpretations. In contrast, some AI applications have the potential to disadvantage particular groups, increase healthcare costs for patients, or cause serious physical harm by failing to perform as expected.[5]

**Received:** May 15, 2025; ; **Accepted:** September 9, 2025

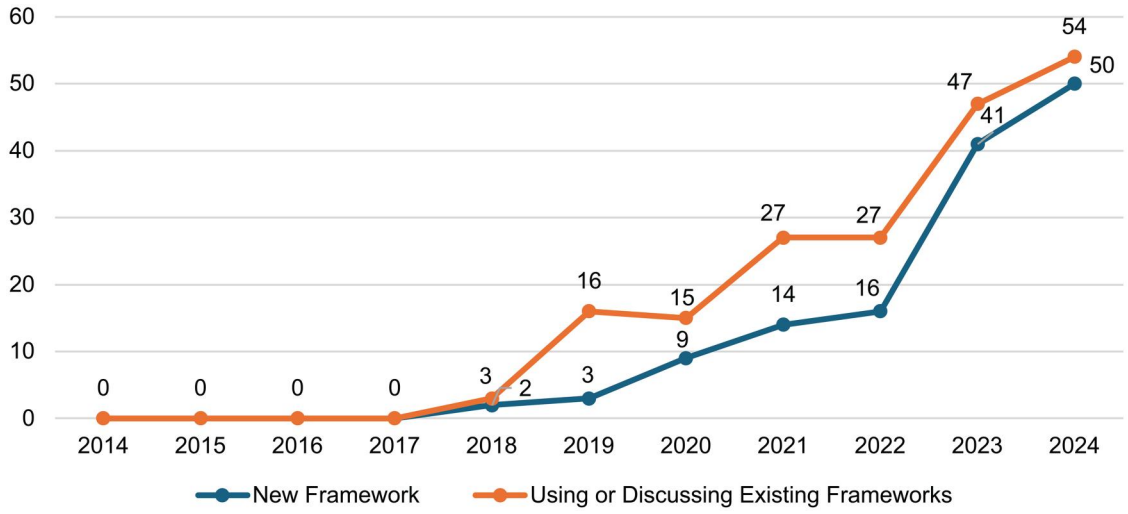


**Figure 1.** AI ethics frameworks indexed by title in Scopus, Medline, Embase, and Web of Science (2014 to 2024).

Several authors[2,7,8] have identified distinct shifts in the AI ethics literature, from the profusion of high-level ethics frameworks to efforts to develop a normative consensus on ethical principles. The current landscape is witnessing a practical shift from "what to how,"[8] exploring how AIEFs and principles can be translated into practice, with proposals for impact assessments, auditing tools, and governance mechanisms.[2] However, in both healthcare and AI ethics generally, a large gap exists—with limited review of AIEFs already operationalized in practice.[9]

## Objective

This review aims to understand the impact of AIEFs operationalized in healthcare. Assessing AIEF impact is crucial for improving the effectiveness of future clinical and organizational interventions and strengthening reporting quality, especially in settings where decisions affect patient safety, public trust, and the responsible adoption of innovation. A key objective of this study is thus to drive a renewed focus on the real-world implementation of AIEFs and the measurement of their impact. Without close examination of AIEF impact, there is a real risk of effort being wasted, misplaced confidence, or missed opportunities to improve ethical oversight in practice. To the best of our knowledge, no scoping reviews exist on this topic. Several general[10–13] and healthcare-specific[14] reviews synthesize existing AIEFs and key ethical principles. Others focus primarily on the AI design phase[1,15] or on tools available for implementing AIEFs—such as publicly available educational tools and practical methods.[9,16–18] One study[9] examines how AIEFs have been implemented or recommended for use in AI-based healthcare applications but has no analysis of the qualitative or quantitative impact of included frameworks.

To address this knowledge gap, the research questions addressed by this review are:

1) What types of AI ethics frameworks have been used in healthcare?
2) How have AI ethics frameworks been operationalized in healthcare (eg, on hospital governance boards) and how have they been evaluated?
3) What is the practical impact of AI ethics frameworks in healthcare?
4) What are the challenges in implementing AI ethics frameworks in healthcare?

## Materials and methods

As the literature around AIEFs is rapidly evolving and heterogenous, we conducted a scoping review to map and give an indication of the volume and focus of the literature.[19] We examined peer-reviewed and grey literature published on the topic of AIEFs implemented in healthcare settings between 2014 and 2025. The development of AIEFs frameworks has only truly gained momentum in academia, policy, and industry in the last decade[10]; hence, our search was limited to the last 10 years to capture the most relevant instances of AIEF operationalization (see also Figure 1). This period captures both the impact of large language models (LLMs) and generative AI on healthcare[20] as well as widely known frameworks from before the transformer-driven boom.[3] Our review follows the Joanna Briggs Institute's (JBI) guidelines for conducting scoping reviews[21,22] and the Preferred Reporting Items for Systematic Reviews Extension for Scoping Reviews (PRISMA-ScR) reporting standards.[23]

### Search strategy

We searched 4 traditional medical and multidisciplinary databases, Google, and Google Scholar to identify peer-reviewed articles and grey literature on the topic. The search strategy was developed in consultation with a research librarian (see Appendix S1). The selection of literature was undertaken in 3 phases. First, we searched Ovid MEDLINE, Ovid Embase, Scopus, and Web of Science using a combination of MeSH terms and text words contained in titles, abstracts, and keywords pertaining to the 4 key areas of artificial intelligence, ethics frameworks, healthcare, and operationalization. These databases were queried on January 16, 2025. Second, we conducted a supplementary Google search using 16 search strings on January 29-30, 2025, screening the first 3 pages (30 results) for each string based on title and preview text, with full content examined when potentially relevant. Third, we searched the first 20 pages (200 results) of Google Scholar

sorted by relevance, following Bramer et al,[24] on February 4, 2025 for supplementary sources and to capture preprints.

### Study selection

We defined "AI ethics frameworks" as documents containing a structured set of normative principles, processes, or guidelines designed to guide or inform ethical decision-making AI within the AI lifecycle.[17,25] To cover the heterogenous array of AIEFs, we included both principle-based tools and technical, step-by-step tools. These frameworks could be general-purpose or healthcare-specific. We defined AIEF "operationalization" as implementation in clinical or organizational settings, including frameworks used to structure clinical governance processes, inform decision-making, assess AI compliance with established ethical standards, or to guide the deployment of AI technologies. As we focused on healthcare contexts rather than AI interventions, we excluded AIEFs used solely in the design phase, ie, by AI developers, but AIEFs across the full AI lifecycle were included. Our scope encompassed technologies used in both clinical and broader organizational healthcare settings, like predictive AI for disease diagnosis as well as AI workflow optimization and digital scribes.

We included only operationalized AIEFs. Studies were included if (1) the study design was primary research, (2) the study reported qualitative or quantitative effects of an AIEF (general or healthcare-specific) operationalized in a healthcare setting at a deployment level (eg, hospitals, doctors, users), (3) the study was published between January 2014 and January 2025, and (4) the publication was in English (the language of the researchers).

Studies were excluded if they focused on non-healthcare settings, discussed frameworks not implemented in practice (eg, protocols), or examined frameworks operationalized at a design level. Study selection was carried out with a 2-step screening process using Covidence.[26] One reviewer (A.C.) screened titles/abstracts for all retrieved articles and a second reviewer (H.R.-A.) independently screened 10% of these. Any disagreements were resolved by reviewing the full text and discussion among reviewers. This process was repeated for full-text review, and those meeting eligibility criteria were selected for data extraction. Eligible articles identified by hand search were also included.

### Data extraction and synthesis

One author (A.C.) conducted data extraction using standardized forms in Excel (Microsoft) under 5 overarching categories: Study Identification, Study Characteristics, AI System Details, Ethics Framework Information, and Operationalization of Framework. A second reviewer (H.R.-A.) independently reviewed 30% of studies to ensure consistency of interpretation and no disagreements between the 2 reviewers (A.C. and H.R.-A.) were noted. Information extracted from studies included: author, study design, healthcare setting, deployment stage, key framework characteristics, challenges in operationalization, reported benefits, reported limitations, and evidence of impact. Given the qualitative nature of the data, a narrative synthesis was performed for this review. The characteristics of each study were first analyzed to provide an overview of the data, following which key features of the operationalization of the AIEFs were analyzed, compared, and synthesized.

## Results

The database searches identified 1807 unique records; 1660 of these did not meet the eligibility criteria and were excluded at the title/abstract screening stage. No documents identified through our Google grey literature search met the inclusion criteria. 147 full-text articles were reviewed; 16 studies met inclusion criteria (Figure 2).

### Study characteristics

Studies were conducted across 9 countries, predominantly in the United States ($n = 7$) and Germany ($n = 2$). Remaining studies were conducted in Canada, Copenhagen, Finland, Ireland, Israel, Italy, and the United Kingdom (each $n = 1$). Of these, one study was a European collaborative study[27] and one had global researcher participation.[28] Although our search spanned 10 years, all papers were published in the last 5 years, with the first dating back to June 2021.[28]

All 16 included studies were cross-sectional in design, collecting data from one time point with no comparison to data from before AIEF implementation. These comprised: 5 pilot studies (preliminary initiatives testing AI guidelines in practice),[6,27,29–31] 5 case studies (an in-depth examination of the use of an AIEF typically in one setting or for one application),[28,32–35] 5 implementation studies (describing the integration of an AIEF into practice, eg, the creation of an AI ethics committee),[4,5,36–38] and a comparative case study (comparing the use of different AIEFs).[25] Settings included hospitals ($n = 7$), academic and research institutions ($n = 4$), academic hospitals ($n = 3$), and private sector ($n = 2$).

### What types of AI ethics frameworks have been used in healthcare?

Over half of the studies ($n = 9$) used a healthcare-specific AIEF.[4,5,29,33–38] These included frameworks developed by academic hospitals ($n = 5$),[5,33,35–37] such as University of Wisconsin Health or Duke University Health, ranging from high-level governance principles[37] to more comprehensive checklists and assessments covering desirable AI characteristics.[5,33]

Intergovernmental AIEFs were also commonly used ($n = 7$),[4,6,25,27,28,30,32] particularly the European Commission High-Level Expert Group on AI's (AI-HLEG) 3 guidance documents: the "Ethics Guidelines for Trustworthy AI,"[40] "Assessment List for Trustworthy AI,"[41] and "Policy and Investment Recommendations for Trustworthy AI."[42] However, these frameworks are not healthcare-specific and do not account for changes in AI systems over time. To address these gaps, a small subset of studies ($n = 3$)[6,28,30] explored the use of Z-Inspection, a process-based ethical assessment tool that tailors the AI-HLEG principles to a variety of practical domains such as healthcare and business.

### How have AI ethics frameworks been implemented in healthcare?

We found that AIEFs have been implemented in healthcare using 3 distinct approaches:

- Establishing AI governance structures and ethics committee guidelines ($n = 4$).[4,35,37,38]
- As ethical review assessment systems for adopting or rejecting AI tools in clinical settings based on standardized evaluation criteria ($n = 8$).[4,5,31,33,35–38]

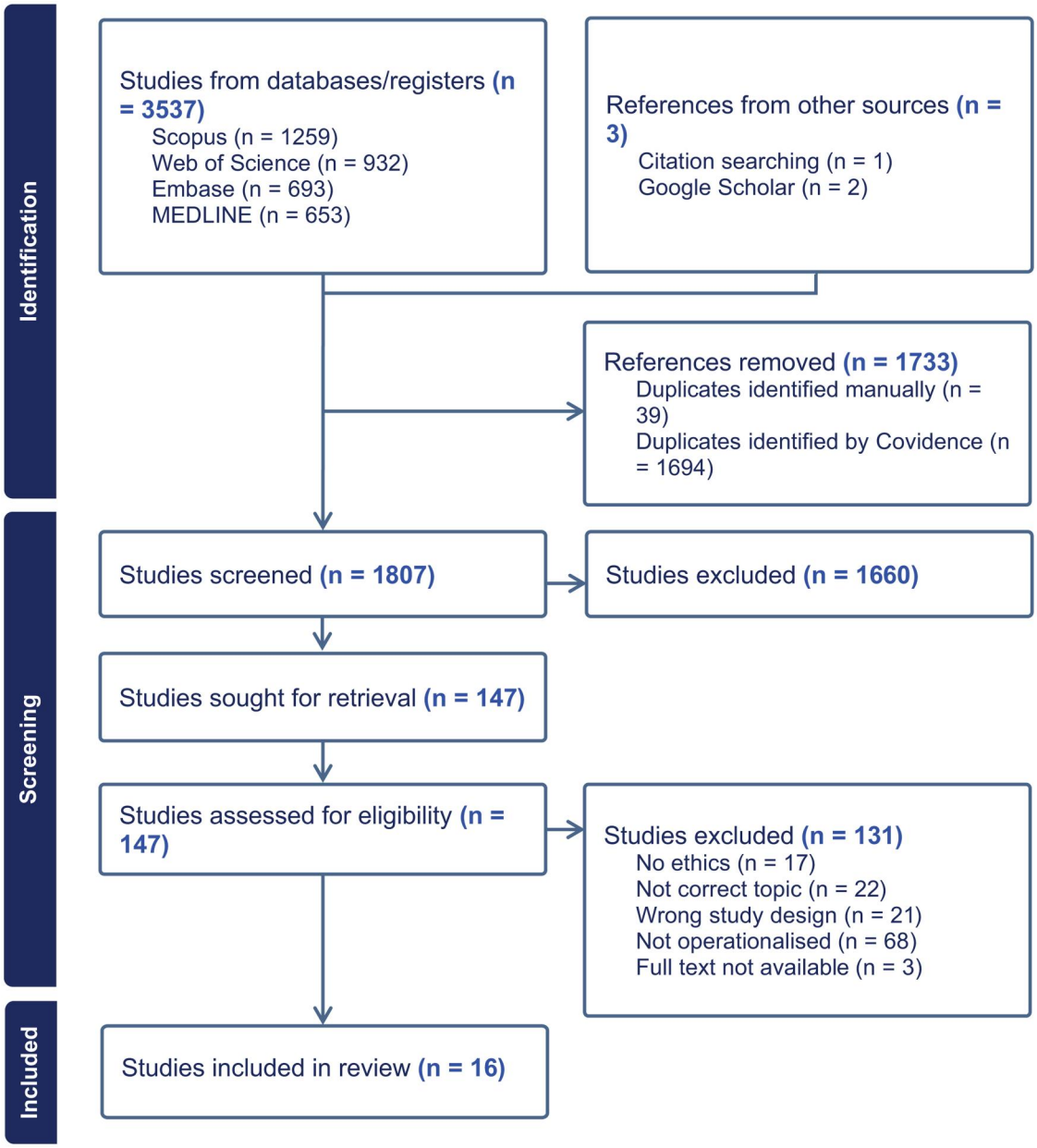


**Figure 2.** PRISMA flow diagram.

- As ethics "audits" to identify ethical risks in AI tools and recommend improvements ($n = 8$)[6,25,27–30,32,34] (see Tables 1 and 2).

Included studies were further categorized into 3 AI lifecycle stages, using stages adapted from the Australian Government's Digital Transformation Agency[42] relevant to operationalization: evaluation of an existing AI tool before implementation; implementation of an AI tool for use; and continuous monitoring of AI performance post-implementation. Most studies focused on the pre-implementation evaluation of AI systems ($n = 7$),[4,25,30–33,36] or all 3 operational lifecycle stages ($n = 5$).[5,29,35,37,38]

## What is the practical impact of AI ethics frameworks in healthcare?

AIEFs' impact was largely reported via process outcomes[43,44] such as improved trust in AI systems, increased transparency, and improved oversight from healthcare professionals. No studies reported on health-related outcomes such as increased quality of care. All 16 studies reported qualitative rather than quantitative results. Furthermore, no studies provided evidence of comparative benefit between AIEF implementation and a historical control, although one study assessed the comparative benefit between different AIEFs in terms of time and expertise required for implemention.[25]

Key reported benefits from AIEF implementation in healthcare included: improved tracking of AI projects as they move through the implementation process,[5] the identification and resolution of technical error (eg, incorrect predicted risk score for unplanned cancer patient hospital admissions),[36] and recommendations to correct identified compliance gaps (eg, insufficient details on AI model development and source code).[29]

The studies identified various positive comments on the use of AIEFs, such as adding value by identifying unconsidered

**Table 1.** AI ethics framework uses in healthcare and AI characteristics.

| Purpose of AI ethics framework | Source | Framework name | Type of assessment | AI system lifecycle | AI application |
|---|---|---|---|---|---|
| 1. Establishing AI governance structures and ethics committee guidelines; **and**<br>2. Ethical review assessment systems for adopting or rejecting AI in clinical settings | Liao et al., 2022 | Guiding Principles | Evaluation of third-party AI tools | Evaluation; Implementation; Operation and monitoring | General |
| | Loufek et al., 2024 | FDA Guiding Principles "Predetermined Change Control Plans for Machine Learning-Enabled Medical Devices" and "Good Machine Learning Practice for Medical Device Development" | | | |
| | Saenz et al., 2025 | Official Guidelines | | | |
| | Borkowski et al., 2022 | World Health Organization "Guidance on Ethics and Governance of AI for Health" | | Evaluation | Pathology and radiology |
| 2. Ethical review assessment systems for adopting or rejecting AI in clinical settings | Dagan et al., 2024 | OPTICA (Organizational PerspecTIve Checklist for AI solutions adoption) | Evaluation of third-party AI tools | Evaluation | General |
| | Economou-Zavlanos et al., 2024 | Implementation Guide | | | |
| | Makridis et al., 2023 | AI Institutional Review (IRB) Supplement | | | All (Lung cancer case study) |
| | Callahan et al., 2024 | Fair, Useful, and Reliable AI Model (FURM) assessments | | Evaluation; Implementation; Operation and monitoring | General |
| 3. Ethics "audits" for identifying ethical risks in AI tools and recommendations | Treacy et al., 2022 | European Commission AI-HLEG "Ethics Guidelines for Trustworthy AI," "ALTAI," and "Policy and investment recommendations for trustworthy AI." Med-I's "Legal, Privacy, Social and Ethical Requirements and Impact Assessment" | Self-assessment | Evaluation | AI for medical imaging |
| | Rajamaki et al., 2023 | European Commission AI-HLEG "Assessment List for Trustworthy AI" (ALTAI) | | Implementation | ML for aged care and remote monitoring |
| | Allahbadi et al., 2022 | Z-Inspection | | Operation and monitoring | Deep-learning decision support |
| | Kaas et al., 2023 | The Principles-based Ethics Assurance Argument Pattern | | Operation and monitoring | Autonomous natural language clinical telephone assistant |
| | Qiang et al., 2023 | European Commission AI-HLEG "Ethics Guidelines for Trustworthy AI," Open Roboethics Institute "Foresight into AI Ethics Toolkit," County of San Francisco "Ethics and Algorithms Toolkit," and Treasury Board of Canada "Algorithmic Impact Assessment" | Evaluation of third-party AI tools | Evaluation | AI recommender system for patients with clinical depression |
| | Zicari et al., 2021a | Z-Inspection | | Evaluation | AI for predicting cardiovascular disease risk |
| | Zicari et al., 2021b | Z-Inspection | | Implementation | AI for detecting early cardiac arrest in emergency calls |
| | Fehr et al., 2022 | Transparency and Trustworthiness Assessment | | Evaluation; Implementation; Operation and monitoring | Clinical ML prediction models |

**Table 2.** Frequency of AI ethics frameworks used in governance versus audit.

| AI governance versus audit | Frequency (n) | Studies |
|---|---|---|
| Frameworks used for AI governance and/or ethical reviews before AI adoption in clinical settings: | 8 | [4,5,31,33,35–38] |
| Frameworks for AI governance *and* ethical review | 4 | |
| Frameworks for ethical reviews only | 4 | |
| Frameworks used for ethics "audits" | 8 | [7,26,28–31,33,35] |

ethical risks[30]; assisting Institutional Review Board (IRB) and Research and Development committee members through more standardized reviews of AI research proposals[31]; and helping an AI governance committee by broadening its "perspective on patient equity and fairness" for AI-related technologies and other potentially biased or inequitable aspects of healthcare provision.[35]

Overall, implementation of AIEFs in healthcare led to reported improvements in AI governance, ethical oversight, and actionable recommendations for improving AI systems. Several studies highlighted the value of these frameworks in revealing ethical gaps—such as fairness concerns, consent procedures, and explainability issues. The reported impacts are listed in Table 3, with key outcomes in bold to distinguish tangible effects from the results of the implementation process.

When examined collectively, the studies suggest that AIEFs have been most effective when integrated into hospital AI governance structures where they support the review and assessment of AI products prior to implementation.[4,5,31,33,36–39] Through this approach, AIEFs facilitated informed decisions on AI adoption and ensured a standardized review process. By comparison, ethics audit systems were primarily used to review technologies *already* implemented in hospitals or after deployment to market. The audits identified ethical gaps, but little evidence was provided on whether companies or healthcare organizations enacted these recommendations.

### What are the challenges in implementing AI ethics frameworks in healthcare?

We identified several challenges relating to AIEF implementation. First, embedding AIEFs into the internal processes of a healthcare institution requires local skill development and cross-disciplinary collaboration. Multidisciplinary teams were predominantly used for structuring a new AI governance board or review process.[5,6,28,30,37,35,37,38]

Next, different AIEFs require varying amounts of time and expertise to implement. In a comparative study of 4 different AIEFs, the time it took for framework application ranged from 1.5 hours for checklist-based frameworks to 20 hours for process-based frameworks.[25] While checklist-based frameworks (such as Government of Canada's *Algorithmic Impact Assessment*) follow a structured list of close-ended questions, process-based frameworks (such as Open Roboethics Institute's *Foresight into AI Ethics*) require multiple internal and external stakeholders to answer open-ended questions. Some studies found that when implementing checklist-based frameworks, there was a strong need for technical guidance and expertise.[25,33] Contrastingly, process-oriented frameworks—which can explore a wider range of AI system impacts—require greater resources (eg, time to completion, stakeholder consultation) and can be unsatisfying for organizations searching for measurable benchmarks.[25]

Operationsalizing a framework in clinical governance settings also presents challenges due to existing workflows, systems, and business practices. Liao et al, for example, experienced resistance to the centralization of a proscribed pathway for AI model evaluation[37] and at Duke University's AI oversight board, the review process was initially perceived as burdensome.[36] Barriers to acquiring data on AI health technologies were also raised: in one study, the non-disclosure of product information meant that the audited companies received low transparency scores.[29]

Furthermore, not all AIEF recommendations are suitable for use in healthcare, nor will non-mandatory recommendations necessarily be implemented by companies and health services. When using the European Commission's "Assessment List for Trustworthy AI" (ALTAI), participants found the recommendations extensive, difficult to use, and were unsure how applicable they were to their AI solution.[27]

## Discussion

This scoping review reports findings of a comprehensive search on the use and impact of AIEFs in healthcare settings. Overall, 16 studies were identified and all reported on qualitative impacts. Despite the publication of at least 173 AIEFs in the last decade,[3] our review reveals that few studies have reported on the implementation impacts of AIEFs.

### Ethics framework impact in governance versus audit

When integrated into hospital AI governance structures, AIEFs achieved their impact through improved oversight and evaluation of AI tools against predefined metrics. A future question to answer is whether the impact of this approach is simply due to more available resources, an institutional prioritization of ethical values, or executive leadership backing.

By comparison, an inherent limitation of post-implementation ethics audits is that while recommendations can be made to organizations such as companies, AI vendors, and healthcare organizations, they may choose to ignore assessment results and withhold information to avoid negative results.[45] This reflects a broader limitation of audit-based AIEFs: recommendations must be understood and then implemented. The success of the ethics audits depends on good-faith cooperation[45]; but given that these assessments are voluntary, organizations typically come with a high openness for proposed changes.[28]

### Challenges in measuring and reporting impact

Determining the clinical impact of AIEFs is challenging due to the gap between an operationalized framework and downstream health and healthcare outcomes. While AIEF implementation may contribute to improved patient safety, clinical effectiveness, enhanced quality of care, or patient satisfaction, current studies have not established direct evidence linking AIEFs to these health outcomes. For this field to advance, studies will need to establish a clearer link with better health outcomes through analyses of patient and system-level impacts over time (eg, reduced patient harms related to AI systems) and evidence of comparative benefit (eg, before/after studies).[43] The collection of baseline data on patient

**Table 3.** Types of AI ethics frameworks and evidence of impact.

| Ethics framework and source | Place of implementation | AIEF use | Reported AIEF impact |
|---|---|---|---|
| Z-Inspection based on European Commission AI-HLEG "Ethics Guidelines for Trustworthy AI" (Academic; Intergovernmental) | Radiology, Brescia, Italy[6] | Audit | Ethics assessment for an AI system resulted in 10 recommendations (including need for a clinical trial, need for a larger dataset with different geographic areas, need for radiologists to report results before reviewing AI system's output to reduce bias). |
| | Copenhagen, Denmark[30] | Audit | Ethics assessment for an AI system resulted in 5 recommendations to address age bias and higher AI accuracy for male than female patients. Qualitative feedback from assessor: the process added value by identifying unconsidered ethical risks. |
| | Emergency Medical Dispatch Centre, Copenhagen[28] | Audit | Ethics assessment for an AI system resulted in 5 recommendations (eg, adding interpretable local approximations for dispatchers to understand AI prediction, involving stakeholders in AI redesign). Potential impacts: improved comprehensibility, public trust, and transparency allowing governance teams to explain their funding, decisions, and system operation. |
| World Health Organization "Guidance on Ethics and Governance of AI for Health" (Intergovernmental) | James A. Haley Veterans' Hospital, Florida[4] | Governance and pre-implementation ethical review | Framework used to create Ethics Subcommittee Guidelines. Two AI radiology tools evaluated and implemented. Potential impacts: Overcoming lack of clinical buy-in and trust in AI implementation. |
| European Commission AI-HLEG "Assessment List for Trustworthy AI" (ALTAI) (Intergovernmental) | Laurea University, Finland[27] | Audit | Resulted in high-level recommendations following ethics self-assessment (eg, providing in-the-loop training, surveying users about their understanding of AI systems). |
| Fair, Useful, and Reliable AI Models (FURM) assessments (Developed by academic hospital) | Stanford Health Care[5] | Pre-implementation ethical review | Integration of ethics assessments into strategic decision-making. Endorsement by executive leadership. Resulted in improved tracking of AI project status across implementation process, ability to prioritize and triage proposed deployments, and better responses to regulatory requests for information. |
| OPTICA (Organizational Perspective Checklist for AI solutions adoption) (Developed by academic hospital) | Clalit Health Services, Israel[33] | Pre-implementation ethical review | OPTICA applied to 18 AI solutions. Resulted in improved oversight and accountability for AI use. More informed decisions on AI solution deployment against predefined metrics. |
| Implementation Guide (Developed by academic hospital) | Duke University Health System[36] | Pre-implementation ethical review | Organizing Committee used guide to conduct 36 reviews of 32 AI technologies against predefined metrics. From review, error in algorithm's clinical risk score revealed, and development team implemented a technical solution. |
| Guiding Principles (Developed by academic hospital) | University of Wisconsin Health[37] | Governance and pre-implementation ethical review | Development of AI governance structure. Favorable feedback: consistent, supervised process for model implementation. |
| Official guidelines (Developed by academic hospital) | Mass General Brigham, Massachusetts[35] | Governance and pre-implementation ethical review | AI governance framework resulted in improved oversight and accountability. High-risk models monitored more frequently. Qualitative feedback: "broadened... perspective on patient equity and fairness." |
| Transparency and Trustworthiness Assessment (Academic) | Online survey and teleconference, Germany[29] | Audit | Resulted in transparency and compliance gaps identified following ethics assessment (eg, insufficient details on AI model development and source code, availability of datasets). Scores and recommendations provided to participants. |
| The Principles-based Ethics Assurance Argument Pattern (Academic) | National Health Service, United Kingdom[34] | Audit | Ethics audit revealed the AI system was safe and appreciated by patients. Ethics gaps identified (eg, need for more consideration of risk borne by clinicians as AI-system may over-refer patients) and solutions constructed. |
| FDA Guiding Principles "Predetermined Change Control Plans for Machine Learning-Enabled Medical Devices" and "Good Machine Learning Practice for Medical Device Development" (Governmental) | Mayo Clinic, Rochester[38] | Governance and pre-implementation ethical review | Board received over 300 requests for review, around half received a review. Resulted in AI models being assessed based on predefined metrics and internal innovators being enabled. Potential impact: mitigating risk of harm. |

(continued)

**Table 3.** (continued)

| Ethics framework and source | Place of implementation | AIEF use | Reported AIEF impact |
|---|---|---|---|
| AI Institutional Review Board (IRB) Supplement (builds on Executive Order (EO) 13960: Promoting the Use of Trustworthy Artificial Intelligence in the U.S. Federal Government) (Institutional; Governmental) | Department of Veteran Affairs[31] | Pre-implementation ethical review | Resulted in positive effect on IRB reviewer's attitudes and ease of review. Highly positive qualitative feedback: standardization of reviews, unnecessary back-and-forth delays between investigators and reviewers avoided. |
| European Commission AI-HLEG "Ethics Guidelines for Trustworthy AI," Open Roboethics Institute "Foresight into AI Ethics Toolkit," County of San Francisco "Ethics and Algorithms Toolkit," and Treasury Board of Canada "Algorithmic Impact Assessment" (Non-profit think tank; Intergovernmental; Governmental) | AI healthcare startup, Canada[26] | Audit | Resulted in increased satisfaction from companies branding as "responsible innovators." Increase in organizational clarity around ethical action items and a roadmap to follow for design and policy decisions. Increase in number of key stakeholders engaged in AI ethics discussions. |
| European Commission AI-HLEG "Ethics Guidelines for Trustworthy AI," "ALTAI," and "Policy and investment recommendations for trustworthy AI." Med-I's "Legal, Privacy, Social and Ethical Requirements and Impact Assessment" (Intergovernmental; Academic) | Med-I Project, Ireland[32] | Audit | AI audits resulted in several ethical gaps being identified and documented (eg, the right to withdraw consent, to object, and to be forgotten). |

outcomes and organizational practices will be essential for quantifying AIEF impacts, while retrospective studies remain valuable particularly when using routinely collected health system data to examine changes following AIEF implementation. An additional way forward is through the development of standardized guidelines that require users to identify key qualitative or quantitative measures for ensuring impact across key timepoints.

### How can we better evaluate of the impact of AI ethics frameworks?

Health technology interventions can have various impacts, from operational changes to health outcomes. The Information Value Chain Theory[46] is a health informatics framework specifying the benefits of such interventions, at different stages of a value chain. It helps identify where an intervention fails and, for our purposes, can assist in evaluating ways that AIEF impact can be improved. The 5 key stages are as follows: (1) an *interaction* between a user and an intervention or technological system occurs; (2) some interactions result in *information received* by that user; (3) some information will lead to a *decision* change; (4) some decision changes will result in a *care process altered*; (5) due to process changes, an *outcome change* may occur for patients or end users.

Table 4 details identified measures and potential measures for tracking impact across 2 AIEF uses: establishing new AI governance structures and conducting ethics audits. The left-hand side of the Information Value Chain captures changes to organizational or clinical processes. Process measures here could include a quantifiable increase in conducted ethical reviews or increased compliance with ethical standards. The right-hand side (Health-related Outcome Changed) captures health outcomes-based measures, such as a reduced number of data breaches or reduced patient harms related to AI systems. The identified measures for framework impact emerged from our review, and the potential measures were developed through critical analysis against expected framework outcomes (eg, addressing key ethical issues in healthcare such as data and privacy) or gaps identified by included studies. For instance, some studies[34,40] noted qualitative feedback provided by clinicians and review boards but highlighted that patient and consumer perspectives should be included in the future—this was then added as a potential measure for determining AIEF impact. All measures were then mapped onto the Information Value Chain to identify gaps in AIEF implementation.

While we would expect to see empirical evidence of AIEF impact on health outcomes a decade on from the inception of the AI ethics boom, most studies have described process impacts clustering around the left-hand side of the value chain, such as a higher occurrence of audits and more structured information. As health outcomes serve as indicators of whether the intended benefits of AIEFs are realized, systematically capturing this data is necessary for improving reporting quality and strengthening future impact of AI ethics initiatives in healthcare. Further research is also needed to develop and validate AIEF evaluation measures.

## Limitations

This scoping review used a comprehensive search strategy covering 4 academic databases, Google Scholar, and Google. However, many private sector and hospital initiatives are not publicly available through a public literature search, thus may be underrepresented in this review. Additionally, while we would expect organizations using AIEFs to report upon operationalized practices through public releases or reports, our Google search did not find any eligible documents. This absence may reflect both a lack of AIEF impact assessment for operationalized frameworks[47] and underdeveloped public

**Table 4.** Examples of identified and potential outcomes at different stages of AIEF operationalization, using the information value chain.

| | i. Interaction | ii. Information Received | iii. Decision Changed | iv. Care Process Altered | v. Health-related Outcome Changed |
|---|---|---|---|---|---|
| 1. *New AI governance structures and ethical review systems developed* | Increased frequency or regularity of working groups meetings to discuss AI use cases[35]<br>• Establishment of a dedicated committee that oversees AI adoption and monitoring[4,37,38]<br>• Increased number of ethical reviews conducted[35,36,38] | • AI governance boards report on more structured and more detailed information from AI vendors including technical data and performance[38]<br>• Reviewers report on more consistent criteria for evaluating AI performance and acceptability[37]<br>• Information consistently received according to AI monitoring plan | • Endorsement by executive leadership and integration of ethics assessments into strategic decision-making[5]<br>• High-risk models are monitored more frequently[35]<br>• Reviewers report greater confidence in their assessments[35]<br>• Reported improvements in status tracking for AI projects and better responses to regulatory requests for information[5]<br>• Increased number of AI tools rejected for not meeting ethical criteria | • AI tools are modified by clinical and patient feedback before adoption[36]<br>• Increased number of AI models updated or retired following performance and impact monitoring | • Qualitative clinical feedback about improved trust, transparency, and comprehensibility of AI systems[35]<br>• Qualitative patient feedback about improved trust, transparency, and comprehensibility of AI systems[34]<br>• Reduced number of clinical or patient complaints about AI systems<br>• Reduced number of data breaches<br>• Reduced or maintained number of patient harms related to AI systems |
| 2. *Ethics assurance or audits for ethical risks and solutions conducted* | • Occurrence of self-assessment/independent audits of AI system's risks and benefits[6,25,28,30,34,34]<br>• Increase in number of key stakeholders engaged in AI ethics discussions[25,28,30]<br>• Increased satisfaction from companies branding as "responsible innovators"[25] | • Vendors or companies receive increased number of ethical recommendations and gaps identified[6,27–30,32,34]<br>• More comprehensive information from stakeholders collected for more accurate assessments of value trade-offs[25] | • Increased compliance with ethical standards[6]<br>• Percentage of vendors or companies that have constructed and implemented solutions in line with ethical recommendations | • Proportion of healthcare providers undergoing training on AI ethics principles and risk mitigation[25]<br>• Reported improvements in patient-clinician communication about AI | |

Items with citations were identified through the review, uncited items represent potential measures developed through analysis and synthesis.

reporting practices in this area. There is a further incentive for practitioners or researchers not to report results when an implementation is unsuccessful, leading to a potentially unbalanced dataset. In our findings, included studies typically provided a measured perspective on the challenges, benefits, and limitations of AIEF implementation, but several noted that future research could include broader stakeholder perspectives.[34,38] Due to the English language limitation in our search, prominent actors in AI development and use such as China and Korea[9] are likely underrepresented.

Lastly, 3 included studies drew from the same AIEF process. One pair[28,30] had the same first author, with an AIEF process called "Z-Inspection" applied to 2 different cardiovascular disease AI tools. A third paper[6] applied Z-Inspection to an AI system for diagnosing COVID-19 damage from chest X-rays and featured the researcher as the last author. There is potential for the impact of the AIEF to be overrepresented.

## Conclusion

Despite the profusion of AI ethics frameworks and guidelines in the last decade, evidence of their impact in healthcare remains surprisingly limited. To bolster the impact of AIEFs, researchers, companies, and clinical governance entities must explicitly capture and report on the impacts of operationalized AIEFs, including clinically significant ones such as changes in health outcomes. Only then can evidence-based ethical decision-making truly occur in the implementation, review, and monitoring of AI systems.

## Acknowledgments

The authors would like to thank Mary Simons who contributed to the development of the search strategy.

## Author contributions

Anastasia Chan (Conceptualization, Investigation, Methodology, Writing—original draft, Writing—review & editing), Hania Rahimi-Ardabili (Investigation, Methodology, Supervision, Writing—review & editing), Wendy A. Rogers (Writing—review & editing), and Enrico Coiera (Conceptualization, Investigation, Supervision, Writing—review & editing)

## Supplementary material

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

## Funding

NHMRC Investigator Grant 2008645.

## Conflicts of interest

There are no conflicts of interest in this project.

## Data availability

The data underlying this article are available from the corresponding author upon reasonable request.

## References

1. Solanki P, Grundy J, Hussain W. Operationalising ethics in artificial intelligence for healthcare: a framework for AI developers. *AI Ethics*. 2023;3:223-240.
2. Ayling J, Chapman A. Putting AI ethics to work: are the tools fit for purpose? *AI Ethics*. 2022;2:405-429.
3. Corrêa NK, Galvão C, Santos JW, et al. Worldwide AI ethics: a review of 200 guidelines and recommendations for AI governance. *Patterns*. 2023;4:100857.
4. Borkowski AA, Jakey CE, Thomas LB, Viswanadhan N, Mastorides SM. Establishing a hospital artificial intelligence committee to improve patient care. *Fed Pract*. 2022;39:334-336.
5. Callahan A, McElfresh D, Banda JM, et al. Standing on FURM ground: a framework for evaluating fair, useful, and reliable AI models in health care systems. *NEJM Catal Innov Care Deliv*. 2024;5:CAT. 24.0131.
6. Allahabadi H, Amann J, Balot I, et al. Assessing trustworthy AI in times of COVID-19: deep learning for predicting a multiregional score conveying the degree of lung compromise in COVID-19 patients. *IEEE Trans Technol Soc*. 2022;3:272-289. https://doi.org/10.1109/TTS.2022.3195114
7. Georgieva I, Lazo C, Timan T, Van Veenstra AF. From AI ethics principles to data science practice: a reflection and a gap analysis based on recent frameworks and practical experience. *AI Ethics*. 2022;2:697-711.
8. Morley J, Floridi L, Kinsey L, Elhalal A. From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Sci Eng Ethics*. 2020;26:2141-2168.
9. Goirand M, Austin E, Clay-Williams R. Implementing ethics in healthcare AI-based applications: a scoping review. *Sci Eng Ethics*. 2021;27:61. https://doi.org/10.1007/s11948-021-00336-3
10. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. *Nat Mach Intell*. 2019;1:389-399.
11. Fjeld J, Achten N, Hilligoss H, Nagy A, Srikumar M. Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center Research Publication. 2020. Accessed January 21, 2025. http://nrs.harvard.edu/urn-3:HUL.InstRepos:42160420
12. Hagendorff T. The ethics of AI ethics: an evaluation of guidelines. *Minds Mach*. 2020;30:99-120.
13. Floridi L, Cowls J. A unified framework of five principles for AI in society. In: Carta S, ed. *Machine Learning and the City: Applications in architecture and urban design* [Internet]. John Wiley & Sons Ltd. 2022:535-545. Accessed February 11, 2025. https://doi.org/10.1002/9781119815075.ch45
14. Marwood T, Boyd J, Khan UR, Barclay SJ, Jackson K. The ethical application of AI in Health: a desktop review. Digital Health CRC; 2022. Accessed January 21, 2025. https://digitalhealthcrc.com/wp-content/uploads/2022/12/Ethical-Application-of-AI-in-Health.pdf
15. Li F, Ruijs N, Lu Y. Ethics & AI: a systematic review on ethical concerns and related strategies for designing with AI in healthcare. *AI*. 2022;4:28-53. https://doi.org/10.3390/ai4010003
16. Kijewski S, Ronchi E, Vayena E. The rise of checkbox AI ethics: a review. *AI Ethics*. 2025;5:1931-1940.
17. Ortega-Bolaños R, Bernal-Salcedo J, Germán Ortiz M, Galeano Sarmiento J, Ruz GA, Tabares-Soto R. Applying the ethics of AI: a systematic review of tools for developing and assessing AI-based systems. *Artif Intell Rev*. 2024;57:110.
18. Prem E. From ethical AI frameworks to tools: a review of approaches. *AI Ethics*. 2023;3:699-716.

19. Munn Z, Peters MD, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol*. 2018;18:143-147.
20. Manoj R, Nandhini G, eds. A comprehensive investigation on leveraging generative AI and large language models in the healthcare domain. In: *2024 IEEE 12th Region 10 Humanitarian Technology Conference (R10-HTC)*. IEEE. 2024:1–6.
21. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. *JBI Evid Implement*. 2015;13:141-146.
22. Pollock D, Peters MDJ, Khalil H, et al. Recommendations for the extraction, analysis, and presentation of results in scoping reviews. *JBI Evid Synth*. 2023;21:520-532.
23. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169:467-473.
24. Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Syst Rev*. 2017;6:245-212.
25. Qiang V, Rhim J, Moon A. No such thing as one-size-fits-all in AI ethics frameworks: a comparative case study. *AI & Soc*. 2024;39:1975-1994.
26. Kellermeyer L, Harnke B, Knight S. Covidence and Rayyan. *JMLA*. 2018;106:580.
27. Rajamäki J, Gioulekas F, Rocha PAL, Garcia XD, Ofem P, Tyni J. ALTAI tool for assessing AI-based technologies: lessons learned and recommendations from SHAPES pilots. *Healthcare*. 2023;11:20. https://doi.org/10.3390/healthcare11101454
28. Zicari RV, Brusseau J, Blomberg SN, et al. On assessing trustworthy AI in healthcare. Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. *Front Hum Dyn*. 2021;3:24. https://doi.org/10.3389/fhumd.2021.673104
29. Fehr J, Jaramillo-Gutierrez G, Oala L, et al. Piloting a survey-based assessment of transparency and trustworthiness with three medical AI tools. *Healthcare*. 2022;10:1923. https://doi.org/10.3390/healthcare10101923
30. Zicari RV, Brodersen J, Brusseau J, et al. Z-Inspection®: a process to assess trustworthy AI. *IEEE Trans Technol Soc*. 2021;2:83-97.
31. Makridis CA, Boese A, Fricks R, et al. Informing the ethical review of human subjects research utilizing artificial intelligence. *Front Comput Sci*. 2023;5:1235226.
32. Treacy C, Regan G, Shahid A, Maguire B. Legal, privacy, social and ethical requirements and impact assessment for an artificial intelligence based medical imaging project. In: Yilmaz M, Clarke P, Messnarz R, Wöran B, eds. *Communications in Computer and Information Science (CCIS 1646)*. Springer Science and Business Media Deutschland GmbH; 2022:29-44.
33. Dagan N, Devons-Sberro S, Paz Z, et al. Evaluation of AI solutions in health care organizations—the OPTICA tool. *NEJM AI*. 2024;1:AIcs2300269.
34. Kaas MHL, Porter Z, Lim E, Higham A, Khavandi S, Habli I. Ethics in conversation: building an ethics assurance case for autonomous AI-enabled voice agents in healthcare. *Assoc Comput Mach*. 2023:1-13. https://doi.org/10.1145/3597512.3599713
35. Saenz AD, Centi A, Ting D, You JG, Landman A, Mishuris RG; Mass General Brigham AI Governance Committee. Establishing responsible use of AI guidelines: a comprehensive case study for healthcare institutions. *NPJ Digit Med*. 2024;7:348. https://doi.org/10.1038/s41746-024-01300-8
36. Economou-Zavlanos NJ, Bessias S, Cary MP, et al. Translating ethical and quality principles for the effective, safe and fair development, deployment and use of artificial intelligence technologies in healthcare. *J Am Med Inf Assoc*. 2024;31:705-713. https://doi.org/10.1093/jamia/ocad221
37. Liao F, Adelaine S, Afshar M, Patterson BW. Governance of clinical AI applications to facilitate safe and equitable deployment in a large health system: key elements and early successes. *Front Digit Health*. 2022;4:931439. https://doi.org/10.3389/fdgth.2022.931439
38. Loufek B, Vidal D, McClintock DS, et al. Embedding internal accountability into health care institutions for safe, effective, and ethical implementation of artificial intelligence into medical practice: a Mayo clinic case study. *Mayo Clin Proc*. 2024;2:574-583. https://doi.org/10.1016/j.mcpdig.2024.08.008
39. European Commission High-Level Expert Group on Artificial Intelligence. *Ethics Guidelines* for Trustworthy AI. 2019. Accessed March 20, 2025. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
40. European Commission High-Level Expert Group on Artificial Intelligence. *The Assessment List* for Trustworthy Artificial Intelligence (ALTAI). 2020. Accessed March 20, 2025. https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment
41. European Commission High-Level Expert Group on Artificial Intelligence. Policy and Investment Recommendations for Trustworthy Artificial Intelligence. 2019. Accessed March 20, 2025. https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence
42. Australian Government. Pilot AI assurance framework guidance. Accessed January 21, 2025. https://www.digital.gov.au/policy/ai/pilot-ai-assurance-framework/guidance/step-1
43. Mant J. Process versus outcome indicators in the assessment of quality of health care. *Int J Qual Health Care*. 2001;13:475-480.
44. Crombie I, Davies HTO. Beyond health outcomes: the advantages of measuring process. *J Eval Clin Pract*. 1998;4:31-38.
45. Vetter D, Amann J, Bruneault F, et al.; Z-Inspection® Initiative (2022). Lessons learned from assessing trustworthy AI in practice. *Digital Society*. 2023;2:35.
46. Coiera E. Assessing technology success and failure using information value chain theory. In: Scott P, de Keizer N, Georgiou A, eds. *Applied Interdisciplinary Theory in Health Informatics*. IOS Press; 2019:35-48.
47. Washington State Health Care Authority. HCA Artificial Intelligence Ethics Framework. 2024. Accessed January 21, 2025. https://www.hca.wa.gov/assets/program/hca-artificial-intelligence-ethics-framework.pdf



Journal of the American Medical Informatics Association, 2025, 32, 1767–1777
https://doi.org/10.1093/jamia/ocaf167
Review