

Review

Improving Explainability and Integrability of Medical AI to Promote Health Care Professional Acceptance and Use: Mixed Systematic Review

Yushu Liu^{1,2}, MPA; Chenxi Liu^{1,2}, PhD; Jianing Zheng^{1,2}, MPA; Chang Xu³, PhD; Dan Wang^{4,5,6}, PhD

¹School of Medicine and Health Management, Huazhong University of Science and Technology, Wuhan, China

²Major Disciplinary Platform under Double First-Class Initiative for Liberal Arts at Huazhong University of Science and Technology (Research Center for High-Quality Development of Hospitals), Wuhan, China

³Smart Hospital Research Institute, Peking University Shenzhen Hospital, Shenzhen, China

⁴School of Management, Hubei University of Chinese Medicine, Wuhan, China

⁵Hubei Shizhen Laboratory, Hubei University of Chinese Medicine, Wuhan, China

⁶Research Center for Traditional Chinese Medicine Development, Hubei University of Chinese Medicine, Wuhan, China

Corresponding Author:

Chenxi Liu, PhD

School of Medicine and Health Management, Huazhong University of Science and Technology

Hangkong Road No. 13

Qiaokou District

Wuhan, 430030

China

Phone: 86 13476237004

Email: liu_chenxi@hust.edu.cn

Abstract

Background: The integration of artificial intelligence (AI) in health care has significant potential, yet its acceptance by health care professionals (HCPs) is essential for successful implementation. Understanding HCPs' perspectives on the explainability and integrability of medical AI is crucial, as these factors influence their willingness to adopt and effectively use such technologies.

Objective: This study aims to improve the acceptance and use of medical AI. From a user perspective, it explores HCPs' understanding of the explainability and integrability of medical AI.

Methods: We performed a mixed systematic review by conducting a comprehensive search in the PubMed, Web of Science, Scopus, IEEE Xplore, and ACM Digital Library and arXiv databases for studies published between 2014 and 2024. Studies concerning an explanation or the integrability of medical AI were included. Study quality was assessed using the Joanna Briggs Institute critical appraisal checklist and Mixed Methods Appraisal Tool, with only medium- or high-quality studies included. Qualitative data were analyzed via thematic analysis, while quantitative findings were synthesized narratively.

Results: Out of 11,888 records initially retrieved, 22 (0.19%) studies met the inclusion criteria. All selected studies were published from 2020 onward, reflecting the recency and relevance of the topic. The majority (18/22, 82%) originated from high-income countries, and most (17/22, 77%) adopted qualitative methodologies, with the remainder (5/22, 23%) using quantitative or mixed method approaches. From the included studies, a conceptual framework was developed that delineates HCPs' perceptions of explainability and integrability. Regarding explainability, HCPs predominantly emphasized postprocessing explanations, particularly aspects of local explainability such as feature relevance and case-specific outputs. Visual tools that enhance the explainability of AI decisions (eg, heat maps and feature attribution) were frequently mentioned as important enablers of trust and acceptance. For integrability, key concerns included workflow adaptation, system compatibility with electronic health records, and overall ease of use. These aspects were consistently identified as primary conditions for real-world adoption.

Conclusions: To foster wider adoption of AI in clinical settings, future system designs must center on the needs of HCPs. Enhancing post hoc explainability and ensuring seamless integration into existing workflows are critical to building trust and promoting sustained use. The proposed conceptual framework can serve as a practical guide for developers, researchers, and policy makers in aligning AI solutions with frontline user expectations.

Trial Registration: PROSPERO CRD420250652253; <https://www.crd.york.ac.uk/PROSPERO/view/CRD420250652253>

KEYWORDS

artificial intelligence; explainability; integrability; healthcare professionals; systematic review

Introduction

Background

The rapid development of artificial intelligence (AI) has demonstrated profound impacts across various industries, particularly in the health care sector. The application of AI has shown significant potential and is widely used in areas such as disease diagnosis, patient monitoring, robotic surgery, and clinical decision-making [1]. However, with the increasing prevalence of AI technology, issues concerning doctors' acceptance, trust, and willingness to use AI have garnered widespread attention. A study revealed that only 10% to 30% of doctors use AI in real-world scenarios [2]. The poor acceptance and low use of AI systems by users are influenced by various factors, including the technical characteristics of AI itself, individual factors (eg, users' AI literacy), organizational factors (eg, advocacy by management), and policy-related issues (eg, responsibility attribution in the use of AI). These challenges hinder the widespread adoption of AI technology [3-7].

Among the technical characteristics of AI, explainability and integrability are considered 2 key factors influencing doctors' acceptance and use of AI [8,9]. While other elements, such as security and social influence, also play important roles in clinicians' trust in AI [10,11], their practical impact on daily clinical adoption differs. Security concerns are essential for AI implementation, but are most often managed at the technical or regulatory level [10]. Social influence, encompassing peer and organizational advocacy, tends to influence adoption at the institutional level [12]. In contrast, multiple systematic reviews and clinician surveys have identified explainability and integrability as the most immediate and actionable factors influencing real-world AI adoption in health care [8-10,13-16]. Accordingly, focusing on explainability and integrability provides practical, user-centered insights for promoting effective integration of AI into clinical practice. In this study, *explainability* refers to the extent to which an AI system provides human-understandable and faithful representations of its decision-making process [13,14], while *integrability* refers to the extent to which AI systems can be embedded into clinical workflows with minimal disruption, ensuring usability, interoperability, and alignment with routine practices [10,15,16]. Importantly, integrability is a broader concept that encompasses interoperability. While interoperability ensures the technical capacity for systems to exchange and interpret data using standardized formats, integrability goes further to consider how AI systems align with clinical roles, decision-making contexts, workflow timing, and user experience [17,18]. A system may be technically interoperable but still lack integrability if it fails to deliver value in practice or imposes additional burdens on clinicians.

Explainability is crucial in the medical field, where decision-making is highly complex and involves significant risks. Clinicians need to ensure the accuracy and safety of AI

outputs before they can trust and rely on AI. A lack of explainability in AI clinical decision support systems (AI-CDSS) may lead to distrust among decision makers and reduce their willingness to use these technologies [19]. In addition, integrability ensures that AI-CDSS can seamlessly integrate into existing clinical workflows. When AI-CDSS are effectively embedded into doctors' daily routines, clinicians can more efficiently access and use patient data, receive recommendations that better meet practical needs, and reduce the time spent on redundant data entry, allowing them to focus on clinical decision-making [20]. Conversely, a lack of integrability in AI-CDSS may negatively impact clinician work by increasing operational complexity, workload, and time costs. Such systems, which fail to align with practical requirements, can reduce doctors' willingness to adopt them [21-23].

However, on the one hand, most existing studies primarily neglect the end users, namely, physicians' understanding of and need for explainable AI [24-27]. Studies have mainly concentrated on the perspectives of developers and researchers and have developed methods to present technical descriptions of model processes in AI [27], although such approaches seem meaningless if they are not aligned with physicians' understanding of and requirement for explainability in real-world application in medical settings [26]. On the other hand, a few studies have explored how system compatibility, user-friendly design, and workflow adaptability [24-26] may contribute to the seamless integration of AI into the clinical workflow. However, what AI integrability entails and what its underlying components are from the physicians' perspective remains unclear.

Objectives

Therefore, this study aims to systematically review existing literature on explainability and integrability of AI systems in health care from the perspective of health care professionals (HCPs). Specifically, it seeks to identify how these 2 factors—explainability and integrability—influence clinicians' acceptance and use of AI-based decision support systems. On the basis of the findings, the study proposes a conceptual framework that synthesizes key user-centered concerns, with the goal of informing the design of more acceptable and adoptable clinical AI tools.

Methods

Literature Search

We conducted this systematic review of systematic reviews according to a protocol registered in PROSPERO (CRD420250652253), and the databases we searched included PubMed, Web of Science, Scopus, IEEE Xplore, ACM Digital Library, and arXiv. We did not explicitly search proceedings from major machine learning and AI conferences such as Association for Computational Linguistics, Neural Information Processing Systems, International Conference on Learning

Representations, or International Conference on Machine Learning, as many papers from these venues are concurrently available on arXiv. We used keywords such as “Explainable AI,” “explainability,” “XAI,” “AI integrability,” “usability,” and “human-centered AI” (the detailed search strategy is provided in [Multimedia Appendix 1](#)) and covered publications from January 2014 to July 2024. In addition, references cited in the retrieved articles and reviews were manually screened.

Inclusion and Exclusion Criteria

Studies were eligible for inclusion if they addressed either the explainability or the integrability of AI in health care (ie, meeting at least one of the following two sets of topic-specific criteria): (1) explainability, the study must address at least 1 of the following aspects—how AI processes input information; how conclusions are reached; the rationale behind these

conclusions; how explainability fosters user trust and their use; or the knowledge, attitudes, and perceptions of HCPs regarding AI explainability—and (2) integrability, the study must address at least 1 of the following aspects—how AI integrates with hospital information systems and clinical workflows; how integrability fosters HCPs’ trust and use; or the knowledge, attitudes, and perceptions of HCPs regarding AI integrability.

In addition, studies were required to (3) focus on the perspective of AI end users, specifically HCPs such as doctors, nurses, and medical laboratory staff and (4) be original research, including original quantitative, qualitative, or mixed methods studies.

Editorials, reviews, conference abstracts, narrative studies, and studies focusing on the perspectives of AI developers, engineers, or technical professionals were excluded ([Textbox 1](#)).

Textbox 1. Inclusion and exclusion criteria for the literature.

Inclusion criteria
<ul style="list-style-type: none">• Time: published between 2014 and 2024• Population: artificial intelligence (AI) users, focusing on health care professionals (HCPs)• Field: related to the medical field or broadly relevant to health care• Outcome: the study must address at least 1 of the following aspects: explaining how AI processes input information; how AI reaches conclusions; the rationale behind AI conclusions; how AI integrates with hospital information systems or clinical workflows; AI usability; how explainability or integrability fosters user trust and use, HCPs (eg, doctors and nurses) knowledge, attitudes, or perceptions of AI explainability or integrability• Study design: original qualitative, quantitative, or mixed method research• Other: published in English
Exclusion criteria
<ul style="list-style-type: none">• Time: published at other periods• Population: AI experts, engineers, or other technical professionals• Field: other specific industries• Study design: abstracts, editorials, commentaries, reviews, and narrative studies• Other: full text unavailable

To maximize comprehensiveness, we initially applied a broad search strategy, but studies were then screened against strict inclusion and exclusion criteria based on the study objectives. YL initially screened titles and abstracts to assess eligibility, with all decisions systematically recorded in a structured Microsoft Excel spreadsheet to ensure transparency. Owing to resource constraints, this stage was conducted by a single reviewer following strictly defined inclusion and exclusion criteria to minimize potential bias. For full-text screening, 2 reviewers (YS and CL) evaluated each study using a standardized Excel form designed to promote consistency in assessment. Any disagreements or uncertainties were discussed with a third author (JZ) until a consensus was reached, further strengthening the reliability of the screening process. Data extraction was conducted by a single reviewer but performed twice to ensure accuracy and completeness. Extracted items included study characteristics, AI system details, target health care settings, participant types, and key findings related to explainability and integrability. Any uncertainties during extraction were resolved by referring back to the full text.

To capture users’ perceptions of explainability and integrability across both health care and general domains, we initially adopted a broad search strategy without restricting to medical-specific terms. However, as 21 (95%) of the 22 included studies were from the health care domain, the final analysis focused on medical contexts.

Quality Assessment

All included studies underwent quality assessment based on their respective study designs. Quantitative and qualitative studies were evaluated using the Joanna Briggs Institute critical appraisal checklists, while mixed methods studies were assessed using the Mixed Methods Appraisal Tool (version 2018). For the JBI checklists, studies scoring ≥80% (ie, meeting at least 8 of 10 items) were considered high quality, 60% to 79% as medium quality, and <60% as low quality. For the Mixed Methods Appraisal Tool, studies that met ≥4 of 5 criteria were rated high quality, while those meeting 3 were considered medium quality. Only studies rated as medium or high quality (22/26, 85%) were included in the final synthesis.

Data Extraction and Evidence Synthesis

The data extraction process included the following items for each eligible article—(1) basic information: authors, year, study region, content, and participants; (2) methodology: study design, study population, data collection methods, and data analysis methods; (3) results: HCPs' understanding and needs regarding AI explainability or integrability.

This study analyzed quantitative and qualitative results separately and integrated them through narrative synthesis, using a descriptive method to present the findings [28]. For qualitative data, we followed the guidelines of Braun and Clarke for thematic analysis. An inductive approach was applied to code data, covering 4 stages: familiarization with the data, initial coding, identification of themes, and review of themes. Thematic analysis was conducted using both deductive and inductive approaches. For the analysis of explainability, we adopted a deductive coding structure based on a preexisting conceptual framework, which includes 3 first-order dimensions: preprocessing, model-level, and postprocessing explainability [25]. Specifically, preprocessing explainability refers to enhancing transparency during data preparation and feature engineering before model training. Model explainability focuses on understanding and interpreting the inner mechanisms, parameters, or representations of the model itself during training. Postprocessing explainability involves applying interpretability techniques after model predictions, which can provide both local (instance-level) and global (model-level) explanations of the model's behavior. These categories guided our initial coding and interpretation. Within each of these dimensions, we

generated subthemes inductively from the data to capture participants' nuanced perspectives. For integrability, no prior framework was applied; all themes were developed inductively based on participants' responses. Because there is a lack of theory guiding the coding for integrability, themes regarding AI integrability emerged based on the coding of extracted text from retrieved studies. NVivo 10 was used for thematic analysis. YL conducted the initial coding individually, and the coded data were then reviewed by a second person to ensure the validity of the themes. Any discrepancies in coding were discussed in group meetings, where consensus was reached.

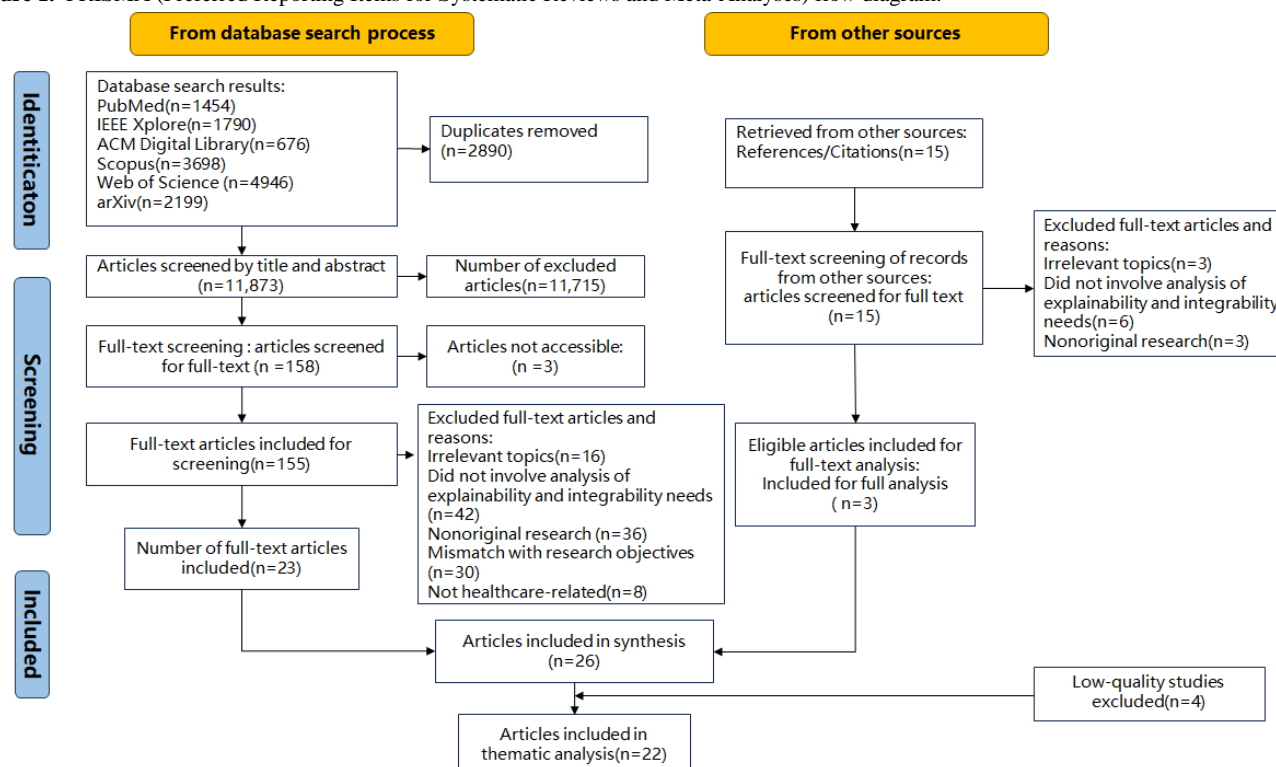
In addition, quantitative data related to the explainability and integrability of medical AI from the user's perspective were also extracted. Owing to substantial heterogeneity in study outcomes and exposure measures, a meta-analysis was inappropriate, and a narrative synthesis was used to analyze the collated studies, including methods, sample size, participants, outcome variables, and dimensions of explainability and integrability.

Results

Literature Screening Results

A total of 11,888 articles were retrieved through the search and their references. Among them, 26 (0.22%) articles met the inclusion criteria for this study. After quality evaluation, 4 (15%) low-quality articles were excluded, and a total of 22 (85%) articles were included in the analysis. The study selection process is shown in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.



Characteristics of the Included Studies

All the included studies were published in 2020 or later. Most of the studies were conducted in developed countries (18/22, 82%), including 9 (41%) from the United States, and only 2

(9%) studies originated from developing countries (China and Brazil). Regarding study type, most studies (17/22, 77%) were qualitative in design, and 5 (23%) studies adopted quantitative or mixed method approaches. The basic characteristics of all the included studies are presented in [Table 1](#).

Table 1. Characteristics of the included studies.

Study	Region	Study design	Content	Participants	Main findings
Graziani et al [13], 2023	Worldwide	Qualitative	Explainability	Health care professionals, industry practitioners, and academic researchers	<ul style="list-style-type: none"> Explainability in AI^a refers to the ability to understand and interpret how models make decisions. It is typically categorized into 3 types: intrinsic explainability, where models are inherently interpretable or designed to be transparent; global explainability, which focuses on understanding the overall behavior of a model through methods such as feature importance and visualization; and local explainability, which aims to explain individual predictions using techniques such as local approximations, counterfactual explanations, or adversarial examples.
Marco-Ruiz et al [29], 2024	Europe and America	Qualitative	Integrability	AI technology developers in hospitals, clinicians using AI, and clinical managers involved in adopting AI, among others	<ul style="list-style-type: none"> Interviewees highlighted that varying protocols across health care organizations can affect AI system effectiveness, emphasizing the need for local validation to assess performance and workflow impact. They also stressed the importance of tools such as process mining and visualizations to analyze patient pathways and optimize complex workflows using real-world data.
Liaw et al [23], 2023	Multicountry	Mixed method	Integrability	Clinicians managing diabetes	<ul style="list-style-type: none"> Interviewees expressed concerns that AI tools might affect patient outcomes and clinical workflows, leading to overdiagnosis, increased costs, and exacerbating health disparities and alert fatigue. The tool's utility is limited for doctors familiar with their patients. Inaccurate data can cause false alarms or missed diagnoses, with a lack of evidence supporting its accuracy. A user-centered design is recommended to improve the system.
Panagoulas et al [30], 2023	Greece	Quantitative	Explainability	Medical personnel (including medical students and medical practitioners)	<ul style="list-style-type: none"> Clinicians rated diagnostic information, certainty, and related reasoning as very important, particularly when their diagnoses conflicted with AI recommendations
Wang et al [22], 2021	China	Qualitative	Integrability	Clinicians in rural clinics	<ul style="list-style-type: none"> Heavy workload: rural doctors handle many patients daily with frequent interruptions, leaving little time for detailed communication or documentation, hindering AI-CDSS^b use. Resource limitations: lack of necessary equipment and medications in clinics makes many AI-CDSS recommendations impractical, reducing their utility. Design mismatch: AI-CDSS are designed for time-consuming, standardized processes that do not fit the fast-paced rural practice, and poor integration with other systems leads to data and recommendation issues.
Zheng et al [31], 2024	America	Qualitative	Explainability and integrability	Pediatric asthma clinicians	<ul style="list-style-type: none"> User needs and challenges in using ML^c systems fell into 3 main areas: how well the system fits into daily workflows (eg, avoiding alert fatigue), the need for clear explanations behind system decisions, and difficulties adapting the tool to real-world settings.
Wolf and Ringland [32], 2020	America	Qualitative	Explainability	Users and developers involved in the design and use of XAI ^d systems	<ul style="list-style-type: none"> Nonexpert users preferred simple, intuitive explanations of AI decisions, using clear language, visual tools, and real-life examples. They focused on fairness, transparency, and the impact on daily life, needing straightforward charts and simplified explanations to build trust.
Morais et al [33], 2023	Brazil	Qualitative	Explainability	Oncologists	<ul style="list-style-type: none"> Visualization helps: experts found visual elements useful for identifying major and minor influencing features. They also expressed a need for more detail: participants wanted more traceability to see how results are generated, enhancing confidence in decisions.

Study	Region	Study design	Content	Participants	Main findings
Helman et al [34], 2023	America	Qualitative	Explainability and integrability	Doctors, nurse practitioners, and physician assistants	<ul style="list-style-type: none"> Doctors primarily focused on several key aspects when using AI tools: analytic transparency, graphical explainability, the impact on clinical practice, the value of integrating dynamic patient data trends, decision weighting (how much to trust and balance AI outputs in real decisions), and display location—including usability and how the interface is viewed by patients and families.
Ghanvatkar and Rajan [11], 2024	Singapore	Quantitative	Explainability	Clinicians	<ul style="list-style-type: none"> The integration of XGB^e with SHAP^f, as well as the combination of LR^g with SHAP, showed high usefulness because of their strong conceptual explanations, with the XGB and SHAP combination performing best in prediction but lowest in fidelity. Usefulness scores also improved during neural network training, indicating better alignment between explanation importance and predictive power over time.
Kinney et al [35], 2024	Portugal	Qualitative	Explainability and integrability	Doctors, educators, and students	<ul style="list-style-type: none"> Transparency in sources: users need to know where AI obtains its information to trust it, similar to how students cite sources. Impact on doctor-patient relationships: doctors fear AI could reduce personal interaction, mirroring issues with EHRsh. Increased burnout: additional AI-driven tasks may increase physician stress and lead to uncritical reliance on AI suggestions.
Burgess et al [36], 2023	America	Qualitative	Integrability	Endocrinology clinicians	<ul style="list-style-type: none"> The study proposes these design principles: (1) ensure algorithms are practical for clinical settings to avoid unrealistic insights; (2) allow clinicians to consider patient-specific factors and maintain control over model outputs; (3) avoid adding “research” tasks to patient visits; and (4) focus on aiding complex decisions in the workflow, not repeating known information.
Yoo et al [37], 2023	South Korea	Qualitative	Integrability	Medical and nursing staff in emergency departments and intensive care units of tertiary care hospitals	<ul style="list-style-type: none"> Anticipated benefits: most participants believe medical AI can reduce decision-making time and handle repetitive tasks, easing workloads and improving efficiency. Main concerns: worries include workflow disruptions, added tasks, reduced clinical autonomy, overreliance on algorithms, skill decline, alert fatigue, and the inability to integrate information beyond electronic records.
Schoonderwoerd et al [38], 2021	Netherlands	Quantitative	Explainability	Pediatrician clinicians	<ul style="list-style-type: none"> Diagnosis explanations are essential: clinicians agree that understanding how CDSSⁱ arrives at a diagnosis is important for trust and decision-making. Need for personalized explanations: while most information elements are seen as valuable, preferences vary, suggesting that explanations should be tailored to individual needs. Balance detail and overreliance: key explanation elements include evidence used, supporting or contradicting data, certainty level, missing information, alternative diagnoses, and past performance—but too much detail may lead to blind trust in the system.
Hong et al [39], 2020	America	Qualitative	Explainability	Practitioners in various industries, such as health care, software companies, and social media	<ul style="list-style-type: none"> Understanding model behavior: during validation, builders need to know why a model produces a specific output for a given case, especially when it performs unexpectedly. Methods such as LIME^j and SHAP help provide these insights. Feature importance analysis: builders assess model logic by examining feature importance, focusing not only on key features but also on less important ones to gain a complete understanding of decision-making.
Gu et al [16], 2023	America	Qualitative	Integrability	Medical professionals in pathology	

Study	Region	Study design	Content	Participants	Main findings
					<ul style="list-style-type: none">Enhanced accuracy and efficiency: the xPath system improved diagnostic accuracy and efficiency, reducing workload and boosting confidence.Traceable evidence for transparency: it provides a layered evidence chain (eg, heat maps and confidence scores), making AI diagnoses transparent and verifiable.User-friendly design: the system aligns with pathologists' workflow, supporting easy verification and adjustment of AI recommendations, thus enhancing usability and adoption.
Wenderott et al [40], 2024	Germany	Qualitative	Integrability	Radiologists	<ul style="list-style-type: none">The key barriers to AI adoption are (1) workflow delays, (2) extra steps, and (3) inconsistent AI-CAD^k performance. The key facilitators are (1) good self-organization and (2) software usability.
Verma et al [41], 2023	Switzerland	Qualitative	Explainability and integrability	Clinicians involved in cancer care (large health care organizations)	<ul style="list-style-type: none">Integration challenges: integrating AI into clinical practice is difficult because of issues with data integration, ontologies, and generating actionable insights.Trust and generalization: clinicians distrust "black-box" models, and AI performance varies across different populations, limiting widespread use.
Tonekaboni et al [42], 2019	Canada	Qualitative	Explainability	Clinicians in intensive care units and emergency departments	<ul style="list-style-type: none">Transparency: doctors need to know the model's context and limitations, such as missing patient information, to trust it even if accuracy is not perfect.Feature explanation: clearly explaining the features used in decisions helps build trust and guides appropriate use in different patient groups.Visualization: well-designed visualizations enhance understanding and support clinical reasoning.
Brennen [43], 2020	America	Qualitative	Explainability	End users and policy makers	<ul style="list-style-type: none">Model debugging and understanding: XAI tools should help users understand model behavior (eg, using LIME or SHAP), but they often require advanced knowledge of ML.Bias detection: tools should identify and explain systemic bias in models and provide context to assess fairness and reliability.Building trust: clearly presenting the data and logic behind decisions helps users understand and trust AI systems.
Fogliato et al [44], 2022	America	Quantitative	Integrability	Radiologists	<ul style="list-style-type: none">One-stage workflow boosts AI reliance: participants more closely followed AI suggestions, especially on noncritical points.AI outperforms but risks overtrust: AI performed better, but improvements were largely because of reliance on AI, even when incorrect.Workflow impacts experience: 1-stage users felt increased confidence and speed; 2-stage users found it more complex and burdensome.
Salwei et al [45], 2021	America	Qualitative	Integrability	Emergency physicians	<ul style="list-style-type: none">The study identified 25 components for integrating a human factors-based CDSS^l into emergency departments, organized into 4 dimensions: time (when the CDSS is used), flow (how it integrates into workflows), patient journey scope (which care stages it covers), and level (integration at individual, team, and organizational levels).

^aAI: artificial intelligence.
^bAI-CDSS: artificial intelligence clinical decision support systems.
^cML: machine learning.
^dXAI: explainable artificial intelligence.
^eXGB: extreme gradient boosting.
^fSHAP: Shapley additive explanations.
^gLR: logistic regression.
^hEHR: electronic health record.
ⁱCDSS: clinical decision support systems.

^jLIME: local interpretable model-agnostic explanations.

^kAI-CAD: artificial intelligence–based computer-aided detection.

^lCDSS: clinical decision support system.

Qualitative Research

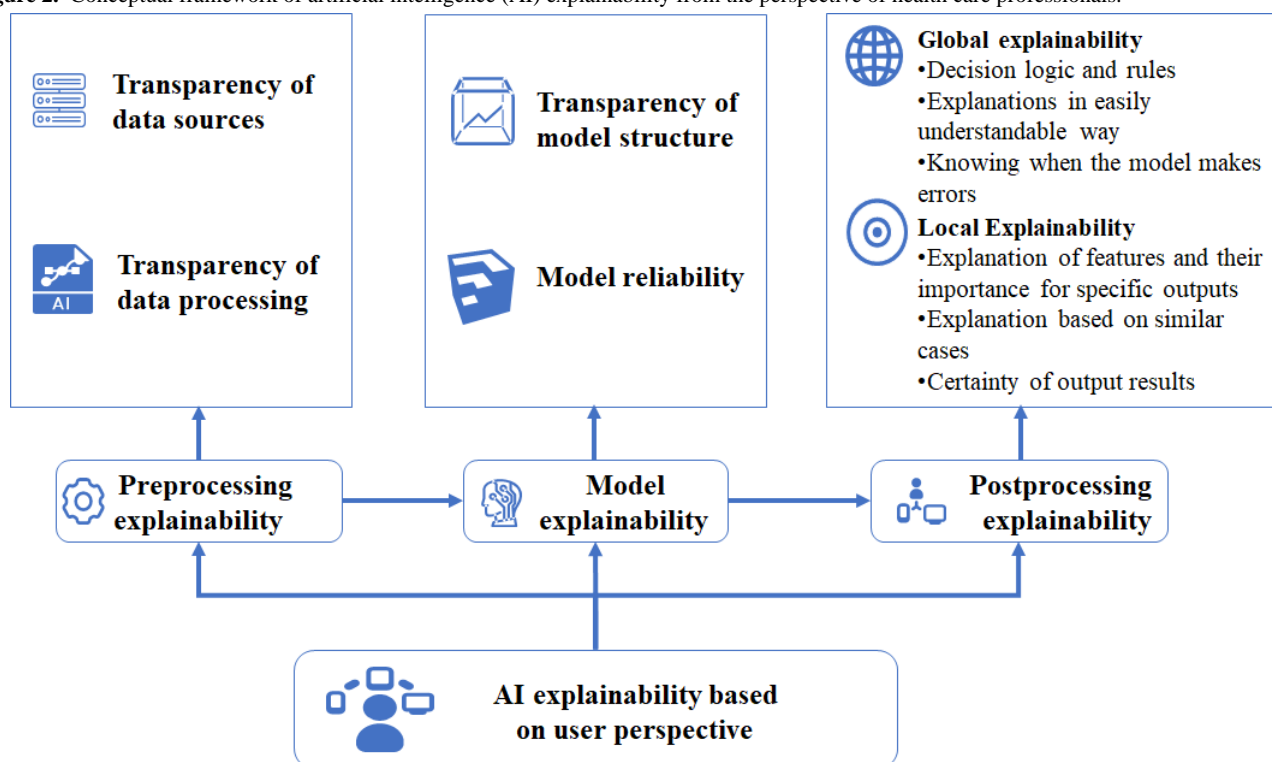
Dimensions of AI Explainability From the User's Perspective

Overview

A total of 16 articles focusing on the explainability of AI in the medical field were included [11,13,16,23,30-36,38,39,41,42,46]. According to the results of thematic analysis, HCPs are most

concerned with postprocessing explainability, with 14 articles highlighting the necessity and importance of explanations provided in the postprocessing stage for HCPs [11,13,16,23,30-34,38,39,41,42,46]. The second most discussed aspect is the doctors' concern regarding model explainability [23,31,42,46]. HCPs showed the least interest in preprocessing explainability [23,35,36]. The themes and subthemes regarding AI explainability from the user perspective are shown in Figure 2.

Figure 2. Conceptual framework of artificial intelligence (AI) explainability from the perspective of health care professionals.



Postprocessing Explainability

Overview

Postprocessing explainability refers to the explanations provided after the AI system has made a decision or prediction. This stage focuses on clarifying the model's output, helping HCPs understand how specific features or data points contributed to a particular decision. HCPs demonstrate the strongest interest in postprocessing explainability of AI, which could be divided into 2 dimensions: local explainability and global explainability. *Local explainability* refers to explanations for individual decisions or specific instances, helping HCPs understand how the model reaches a particular conclusion in a given situation. By contrast, *global explainability* refers to the HCPs' understanding of how the model functions and its underlying decision logic. Comparatively, HCPs focus more on local explainability [11,13,16,23,31,33,34,38,39,41,42].

Local Explainability

On the basis of the current synthesis, local explainability is the most important aspect of explainability for HCPs, as highlighted by 12 studies, including (1) explanation of features and their importance for specific outputs, (2) certainty of output results, and (3) explanation based on similar cases.

First, the features of AI are the most critical component of local explainability, with 11 studies [11,16,23,31-34,38,39,41,42] highlighting that HCPs are concerned with which features were included and their importance in contributing to a specific AI-generated output. On the one hand, identifying which features were used in decision-making is fundamental for clinicians to build trust. By clearly showing the features used by the model, HCPs can verify their relevance, which enhances confidence in the model's decisions. On the other hand, HCPs are also concerned with the extent to which a feature would impact an AI decision, and HCPs would construct trust if their perception of the importance of features was consistent with the results showing which features had the greatest impact on

AI decisions [8,20], and visualization methods play a critical role in enabling HCPs to quickly grasp and understand how these features influence the predictions [31-33,39,42]. By providing clear visual cues such as color-coded severity indicators (eg, red, yellow, and green for high-, medium-, and low-risk categories, respectively) alongside numerical data, these visualizations allow HCPs to assess risk levels at a glance [31]. In addition, the ease of interpreting visual elements helps in distinguishing major and minor influencing factors [33], which enhances the accuracy and speed of decision-making in clinical settings. Moreover, personalized visualization designs are considered an effective way to enhance explainability [32]. Visualization schemes customized to meet HCPs' specific needs can further improve their understanding of the model's decision-making process. See illustrative descriptions from Zheng et al [31] and Hong et al [39]:

Visual indications of severity, such as red, yellow, and green to define high, medium, and low-risk categories paired with a numerical indication were required. [Zheng et al, 2024]

Visual elements are easy to interpret and identification of minor influencing features...most domain experts acknowledged that the visual elements are easy to interpret and were able to perform the identification of major/minor influencing features. [Hong et al, 2020]

Second, the certainty of output results also plays an important role in local explainability. Displaying the confidence or certainty of the model's predictions provides HCPs with additional reference information, enabling them to better understand the model's outputs and thereby increasing their trust in the model [38,42]. For example, providing CIs can help HCPs more effectively assess the reliability of the predictions [39]. See illustrative descriptions from Schoonderwoerd et al [38] and Hong et al [39]:

More specifically, clinicians stated that supporting- and counterevidence, and the certainty of the system will likely remain important in explanations, while information that is used to make the diagnosis, and the diagnosis in similar cases is likely to become less important over time. [Schoonderwoerd et al, 2021]

Presenting certainty score on model performance or predictions is perceived by clinicians as a sort of explanation that complements the output result. [Hong et al, 2020]

Third, AI local explainability means the model's ability to explain its decisions by offering examples of previous instances that are similar to the current case [42]. This allows clinicians to interpret and understand how the model arrived at a decision based on prior cases with comparable features. It essentially helps in making the model's predictions more transparent and interpretable through analogies to similar real-world examples. See illustrative description from Tonekaboni et al [42]:

For example, in cases where an ML model is helping clinicians find a diagnosis for a patient, it is valuable to know the samples the model has previously seen. Clinicians view this as finding similar patients and

believe that this kind of explanation can be only helpful in specific applications. [Tonekaboni et al, 2019]

In addition, the timing of explanations is a key consideration. Providing HCPs with excessive information may lead to information overload, potentially hindering their understanding of the system. For example, some participants noted that they do not want to verify every AI-generated result each time they use a clinical decision support tool. Instead, they only wish to delve into the underlying logic when the results are unexpected [36]. This indicates that HCPs prefer the ability to choose to access more information as needed, rather than being overwhelmed by excessive or redundant details during the explanation process.

A mixed method study confirmed the aforementioned results [38]. Schoonderwoerd et al [38] explored physician requirements for clinical support system explainability and surveyed 6 pediatricians, in which clinicians rated the following aspects as very important in any scenario involving AI-CDSS use, namely, features and their importance, results certainty, features increasing results certainty, and the ability of AI to generalize results to similar situations (median importance ratings were rated as highly important by doctors whether the diagnosis of the doctor and the computerized decision support systems is consistent).

Global Explainability

Six studies highlighted the importance of global explainability for HCPs [13,30,39,41,42,46], including (1) decision logic and rules, (2) explanations in an easily understandable way, and (3) knowing when the model makes errors.

First, explaining the decision logic and rules is a core part of understanding the overall functioning mechanism of AI models. This involves helping HCPs understand the entire decision-making process of the model, from input features to final outputs, including feature combinations, trade-offs, and the determination of decision boundaries (the dividing criteria set by medical AI). Such explanations are key to enhancing model transparency and enabling HCPs to grasp the overarching decision logic of the system [13,39]. See illustrative description from Hong et al [39]:

The majority of our participants desired better tools to help them understand the mechanism by which a model makes predictions; in particular regarding root cause analysis (P1), identification of decision boundaries (P3), and identification of a global structure to describe how a model works (P13). [Hong et al, 2020]

Second, explaining the model in a way that is easy for HCPs to understand is also an important approach to enhancing overall explainability [39,42]. When the model's explanations align with the HCPs' cognitive models and logic, they serve as evidence to support decisions, which can greatly enhance the HCPs' understanding and acceptance of AI model outputs. See illustrative description from Hong et al [39]:

Explanations for model predictions can be used as evidence (P16, P18, P20) to corroborate a decision,

when ML model and user's mental model agree.
[Hong et al, 2020]

Third, understanding when the model might make mistakes is another key aspect of global explainability. HCPs need to not only understand the normal decision-making logic of the model but also recognize the conditions under which the model may fail or generate incorrect decisions [13,39,42]. Only when HCPs are fully aware of the model's limitations can they use AI cautiously in practical applications, thereby improving decision accuracy and safety. For example, HCPs should be informed of potential risks, such as when the model fails to account for specific historical data or lacks certain critical information [42].

Model Explainability

From the user's perspective, the explainability of the model itself can be discussed regarding model reliability and the structural explainability of the model.

Model Reliability

Model reliability refers to the ability of an AI tool to consistently produce accurate and dependable results over time. It is typically assessed using performance metrics, such as accuracy, specificity, and sensitivity, which evaluate how well the model performs in predicting outcomes [23,31,42]. These metrics significantly influence clinicians' initial adoption of the tools [42].

Transparency of Model Structure

Structural explainability, on the other hand, relates to how transparent and interpretable a model is in demonstrating the relationship between input features and the final output [42]. AI models based on algorithms, such as decision trees or logistic regression, can clearly demonstrate how input features influence the final output, helping clinicians understand the model's parameters and reasoning mechanisms [46]. In contrast, "black-box" models, such as deep learning, while demonstrating superior performance in certain scenarios, have a complexity that makes it challenging for HCPs to understand their internal workings. See illustrative descriptions from Tonekaboni et al [42] and Fischer et al [46]:

Familiar metrics such as reliability, specificity, and sensitivity were important to the initial uptake of an AI tool, a critical factor for continued usage was whether the tool was repeatedly successful in prognosticating their patient's condition in their personal experience. [Tonekaboniet al, 2019]

If you know what that model is based on, it is not some mysterious black box where something comes out, but we as doctors know what those models are based on and what parameters are included. Then I can live with it not seeing the parameters for each prediction.
[Fischer et al, 2023]

Preprocessing Explainability

Preprocessing explainability refers to the transparency of AI systems when processing input data, including data sources and data preprocessing methods. Clinicians often need to understand what types of data AI systems are based on for predictions or diagnoses to evaluate the model's applicability and reliability. This includes (1) transparency of data sources and (2) transparency of data processing.

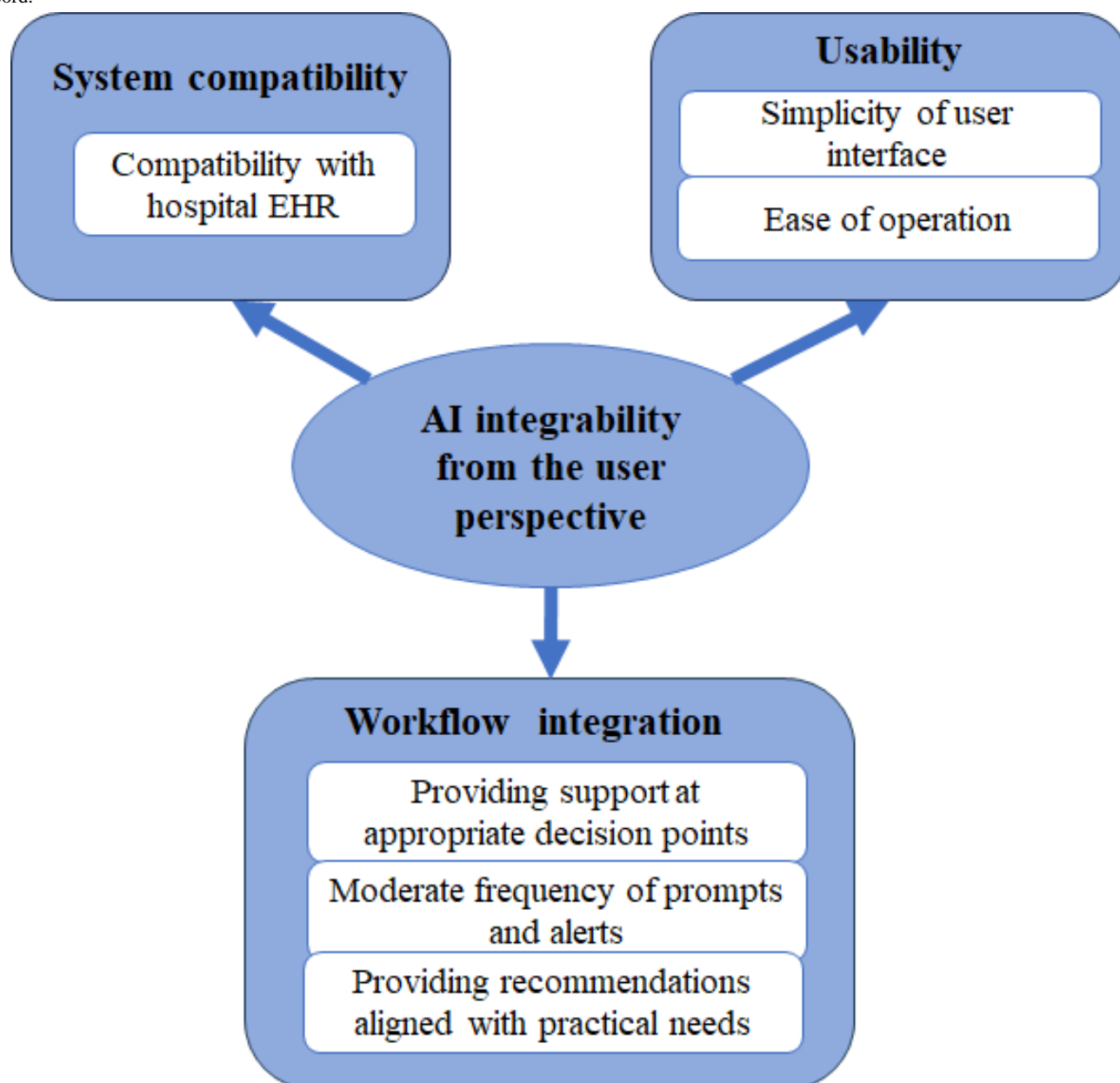
First, the transparency of data sources is the foundation of HCPs' trust. A total of 3 studies indicated that HCPs pay attention to the data sources of AI systems [23,35,36]. In the design of medical decision support systems, HCPs often want to know the origin of the data when they first encounter AI tools.

Second, the lack of transparency in data processing may undermine HCPs' trust in AI systems. When HCPs cannot clearly understand how input data are processed, they may become skeptical of the system. For example, some HCPs mentioned that if they are unaware of who processes the input data, how they are processed, or how they are stored, this uncertainty can lead to reduced trust in the AI system [36].

Dimensions of AI Integrability From the User Perspective

According to the thematic analysis, AI integrability can be understood from the following 3 dimensions: workflow adaptation, system compatibility, and usability (Figure 3).

Figure 3. Conceptual framework of artificial intelligence (AI) integrability from the perspective of health care professionals. EHR: electronic health record.



Workflow Adaptation

Workflow adaptation is a critical dimension of AI integrability, referring to the ability of AI systems to fit seamlessly into existing workflows without disrupting them, avoiding additional workload, and providing recommendations that meet practical needs [16,22,23,29,31,34-37,40,44,45]. This includes (1) providing support at appropriate decision points, (2) moderate frequency of prompts and alerts, and (3) providing recommendations aligned with practical needs.

First, the time points of AI assistance must not disrupt or interrupt doctors' routines, increase workload, and extend time requirements for a workflow with well-integrated AI. It is essential to identify which complex decision points require AI assistance, rather than providing information or data that doctors already know [36]. This is especially problematic in high-patient-volume settings, where doctors are already under significant pressure. Workflow disruptions caused by AI systems

can negatively impact various aspects of the diagnostic and treatment process [16,22,29,34-37,40,44,45]. See illustrative description from Wenderott et al [40]:

When using AI-CAD, Seven radiologists were concerned about potential time constraints associated with the software. [Wenderott et al, 2024]

Second, attention must be paid to the frequency of AI alerts and notifications. Frequent alerts and information prompts from AI systems may lead to alert fatigue among doctors, making them desensitized to genuinely critical alerts. This can also result in information overload, further impairing doctors' decision-making abilities [23,31,36]. See illustrative description from Zheng et al [31]:

In-basket message was mentioned by many clinicians as a common type of active alarm. However, it is necessary to balance effective information delivery and alert fatigue as clinicians, especially physicians,

receive various alarms and notifications from multiple channels in their daily work. [Zheng et al, 2024]

Third, 3 studies highlighted that it is crucial for AI outputs to align with the context of HCPs [22,23,29]. AI recommendations must match the operational capacity of HCPs and health care facilities, especially in resource-limited community clinics. For instance, if an AI system suggests conducting laboratory tests that cannot be performed or prescribing medications that are unavailable, it may reduce clinicians' trust in the system and negatively impact its practical effectiveness [25]. Therefore, thorough local validation is essential when introducing AI systems to ensure they function appropriately within specific health care environments [29]. See illustrative description from Wang et al [22]:

In addition, since our research sites are first-tier community clinics, they are only capable of performing a limited number of laboratory examinations (e.g., none of the research sites have CT scan equipment). They also have very limited medication resources in stock. However, AI-CDSS would suggest a variety of laboratory tests, and treatment and medicine options, which clinicians often cannot prescribe. In this case, these recommendations are often ignored by the clinician users. [Wang et al, 2021]

System Compatibility

System compatibility is a critical dimension of AI integrability, particularly in the health care field, where it primarily refers to integration with electronic health records (EHRs). A total of 5 studies highlighted that many clinicians are willing to integrate AI systems with patient EHRs to provide more comprehensive and relevant information during clinical decision-making [23,31,35,37,40]. Integrating AI risk models into EHR systems can offer clinicians more valuable references, improving patient management while reducing repetitive tasks and minimizing information omissions in clinical workflows. See illustrative description from Kinney et al [35]:

Physicians cited the diverse factors that impact a treatment plan that is not able to be captured in an electronic system as a reason it may not be helpful. [Kinney et al, 2024]

Usability

Usability refers to the ease with which users can interact with and effectively use a system to achieve their goals [47]. In the

context of AI systems, usability is a key aspect, with 5 studies highlighting it as a critical factor that directly impacts the acceptance and effectiveness of the system in clinical settings [22,23,30,40,42].

Simplicity of User Interface

A user-friendly interface needs to be intuitive and easy to operate while providing timely and useful information without disrupting clinical workflows. The display and organization of interface functions are major dimensions of interface usability [22,40,42]. For example, one study found that overly frequent and space-consuming pop-up designs in clinical decision support systems hindered doctors' access to other important information, leading to a poor user experience [22].

Ease of Operation

In addition, whether the system is easy for doctors to master directly affects its use. Clinicians tend to reject AI systems if they require significant time to learn [23,30]. A mixed method study conducted in diabetes management surveyed HCPs' attitudes toward AI [23]. The results showed that 68% of participants considered usability (simple and easy operation) an important factor influencing their use. Another quantitative study, based on the technology acceptance model and diffusion of innovations theory, analyzed the key factors affecting the adoption of AI technologies among doctors and medical students, with 17.9% indicating the lack of user-friendly software and support systems as a barrier [30]. See illustrative description from Wang et al [22]:

A primary issue of AI-CDSS usability was that the system always pop up to occupy one-third of the screen, whenever the clinician opened a patient's medical record in EHR. If the monitor's screen size is small, the floating window of AI-CDSS may block the access to some EHR features (e.g., data fields). This frustrated many participants. To workaround this issue, clinicians had to minimize it while it was not in use. [Wang et al, 2021]

Quantitative Research

This study included 3 quantitative and 2 mixed method studies, focusing on HCPs' willingness to use AI, the key influencing factors, explainability needs, and integrability. Only 1 quantitative study addressed AI integrability. Owing to the limited number of quantitative studies, they primarily serve to complement and validate the qualitative analysis in this study. See Table 2 for details.

Table 2. Characteristics of the publications of quantitative data.

Study	Research methods	Sample size	Participants	Outcome variables	Concerned dimensions of explainability or integrability
Liaw et al [23], 2023	Semistructured interviews and surveys	22	Clinicians managing diabetes	Factors influencing the adoption of the tool, perception of the tool’s usefulness, and ease of use	Transparency, usability, and impact on clinic workflows need to be tailored to the demands and resources of clinics and communities.
Schoonderwoerd et al [38], 2021	Domain analysis, interviews, surveys, and scenario experiment	6	Pediatrician clinicians	Diagnosis, information they have used in their decision-making, and the importance ranking of different types of explanations in various contexts	The information that is used to make a diagnosis, the information that supports the diagnosis, how certain the clinician is of the diagnosis, and the relevance of the information for their diagnosis
Panagoulas et al [30], 2023	Survey	39	Medical personnel (including medical students and medical practitioners)	Suggested level of explainability, knowledge of AI ^a , ways to better integrate AI, and AI concerns	The overall system functions, user-friendly software, and impact on workflow
Ghanvatkar and Rajan [11], 2024	Theoretical construction and case analysis	— ^b	Clinicians	Usefulness of AI explanations for clinicians	Local explanations and global explanations
Fogliato et al [44], 2022	Scenario experiment	19	Radiologists	Anchoring effects; human-AI team diagnostic performance and agreement; time spent and confidence in decision-making; perceived usefulness of the AI	Do not waste time and no additional workload.

^aAI: artificial intelligence.

^bNot available.

Two studies explored factors affecting physicians’ willingness to adopt AI, with a focus on explainability and ease of use. One mixed method study found that 77% of HCPs managing diabetes were willing to use AI, citing ease of use (68%) as a key factor [23]. Another study revealed that 25.6% of participants identified a lack of understanding of underlying technology as a barrier [30], which confirms the focus of HCPs on usability and explainability described in the previous section of the research.

Two studies focused on explainability needs. One found that post hoc local explanations, such as those provided by logistic regression and Shapley additive explanations (SHAP), received higher usability scores from clinicians than model-level explainability [11]. Another study found that clinicians rated diagnostic information, certainty, and related reasoning as very important, particularly when their diagnoses conflicted with AI recommendations [38]. *These quantitative results support and validate the qualitative findings that post hoc local explainability is crucial for HCPs.*

One study examined AI integration into workflows, comparing its placement in different stages of decision-making. It found that AI support at the start of a diagnostic session increased participants’ confidence and perceived usefulness but also

highlighted that poor integration could increase task complexity and workload [44].

Discussion

Principal Findings

To enhance HCPs’ trust and use of AI-CDSS in future real-world clinical settings, this study adopted a mixed systematic review approach to synthesize evidence regarding AI explainability and integrability from the HCPs’ perspective. To the best of our knowledge, this study is the first to systematically summarize the concept of “AI integrability” from the HCPs’ perspective. It refers to the ability of AI systems to be easily and seamlessly integrated into workflows, providing timely, appropriately scaled, and practically relevant prompts or recommendations at the right points, without requiring excessive effort from the HCPs. HCPs’ needs for AI integrability are primarily reflected in 3 aspects: system compatibility, usability, and workflow adaptation.

Second, this study decodes the components of AI explainability based on HCPs’ lived experiences. It identifies that the core HCPs’ requirements of AI explainability can be divided into 3 stages, namely, data preprocessing, the AI model itself, and postprocessing explainability. Unlike the results from AI

developers and researchers, the study found that general users (HCPs) are more focused on the explainability of the postprocessing stage, particularly local explainability, such as the importance of specific output features and the certainty of results. From the HCPs' perspective, an explainable AI must clearly present data sources, processing workflows, model structures, decision mechanisms, and their rationales, using tools such as visualization and user-comprehensible language and logic to help HCPs understand and trust AI.

Comparison With Existing Research

This study is the first to systematically review AI integrability from a comprehensive perspective. Current discussions about easily integrable AI only sporadically mention its compatibility with other systems and its ability to integrate into HCPs' workflows (eg, the location of the AI and when it provides assistance), but lack in-depth and systematic exploration [48-50]. For instance, the study by Maleki Varnosfaderani and Forouzanfar [51] discussed the possibility of integrating AI with medical practice but did not thoroughly examine the specific needs faced by HCPs during the integration process. A study from rural clinics in China reported various tensions between AI-CDSS design and the rural clinical environment, such as misalignment with local environments and workflows, technical limitations, and usability barriers [22]. Another study concerning AI-CDSS in emergency departments identified integrability factors (eg, time, treatment processes, and mobility) through interviews with 12 emergency department doctors, but it was limited to a specific environment with a small sample size [45], resulting in poor extrapolation capability.

This study proposes, from the perspective of HCPs' needs, that AI explainability should not only focus on technical transparency but also emphasize HCPs' understanding and trust, particularly in clinical settings, where AI explanations should support HCPs in making more accurate and effective decisions. Existing research in explainable AI primarily focuses on algorithms, with related reviews mainly discussing taxonomies of explainability and technological innovations [25,52]. For example, Markus et al [53] proposed an explainability framework that mainly focuses on providing better tools for developers but largely explores explainability from an algorithmic perspective. Similarly, Amann et al [14] highlighted ethical and technical issues in explainable medical AI, pointing out the multidimensional nature of explainability, but their research remains focused on algorithm optimization and technical compliance. This study emphasizes the user perspective, offering more practical guidance for the design and promotion of explainable medical AI. The study found that HCPs focus more on postprocessing local explainability, meaning how specific predictions made by the model can explain changes in the patient's condition or decision-making basis. This aligns with Shin [54], who emphasized that local explainability and causal relationships are key to user trust. A systematic review from medical and technical perspectives also supports this view [55]. Unlike data experts, who focus on model and data-layer explainability [25], users prioritize post hoc explainability. Some experimental research shows users (eg, doctors) have a higher understanding and acceptance of post hoc explanations, which are more actionable than traditional

technical explanations [56,57]. This preference stems from 3 key clinical needs. First, post hoc local explanations help clinicians understand AI predictions in the context of individual patient cases, enabling more personalized and relevant decision-making [54]. Second, given their professional and legal responsibilities, doctors need to justify their choices based on understandable and traceable reasoning. Post hoc explainability provides the transparency required to assess whether AI outputs align with clinical guidelines and ethical standards [53]. Third, in high-pressure clinical environments, HCPs prioritize usability over theoretical clarity. Local, case-specific explanations are more practical and immediately applicable, which enhances trust and facilitates integration into routine workflows [58]. Thus, this study's conclusion better matches real clinical scenarios, offering insights for developers to create AI-CDSS that meet the needs of HCPs.

Challenges for Explainability and Integrability

Despite these advances, significant challenges remain in achieving explainability and integrability of AI-CDSS in clinical practice. These challenges are as follows.

Lack of Tailored Explainability Methods for HCPs

A key challenge is the lack of explainability methods tailored to HCPs [59,60]. HCPs focus on the explainability of the postprocessing stage, especially local explainability. To address this, post hoc explainability techniques such as local interpretable model-agnostic explanations (LIME) and SHAP [61] explain decision-making in black-box models, helping HCPs understand how predictions are made based on input data. For instance, Alabi et al [62] demonstrated the use of SHAP and LIME in prognostic modeling for nasopharyngeal carcinoma, highlighting their potential in clinical decision support. Simplifying output by avoiding technical jargon and using graphical explanations [63] allows HCPs to adjust detail levels to avoid overload. Medical AI can also offer personalized explanations based on roles, preferences [64], or feedback after explanations [65].

Dynamic Nature of AI Models Affecting Explanation Consistency

A significant challenge in explainable AI for clinical use lies in the evolving nature of explanations as AI models are continuously updated. These updates—whether for improving performance, incorporating new data, or aligning with emerging medical knowledge—can change a model's internal logic, rendering previously valid explanations obsolete or misleading [66]. In clinical settings, where trust and transparency are paramount, outdated explanations may lead to incorrect interpretations or reduced confidence in AI recommendations. To address this, explainability methods must be adaptive—capable of automatically regenerating explanations following model updates, tracking changes over time, and surfacing the rationale behind those changes [67]. Therefore, maintaining the temporal validity of explanations is as crucial as ensuring their initial explainability, especially as AI systems become increasingly dynamic and responsive to new clinical evidence.

Limited Technical Compatibility With Existing Information Systems

Compatibility with existing information systems is a major challenge [68]. AI-CDSS often require large amounts of patient data to provide decision support, but if these data cannot be electronically retrieved, clinicians must manually input them, leading to frustration and abandonment [69]. Integrability difficulties are also linked to the lack of semantic interoperability standards [70]. To address this, standardized application programming interface and data format protocols should be developed to enable AI systems to automatically retrieve patient data from EHRs, reducing manual workload. In addition, implementing standards such as the Fast Healthcare Interoperability Resources and Unified Medical Language System can facilitate integrability with various patient information systems [71,72].

Complexity of AI Integrability in Clinical Settings

While this review identifies key enablers of AI integrability—namely, system compatibility, usability, and workflow adaptation—it is important to emphasize that integration is rarely seamless in real-world clinical settings. These dimensions, although essential, do not guarantee smooth adoption. For example, perceptions of usability often vary among different clinical roles, leading to inconsistent engagement [73]. Embedding AI tools into existing workflows can require significant adaptation, redefinition of tasks, and role negotiations. Even when systems are technically compatible, they may introduce new tensions, including resistance from clinicians or disruption to established routines [22]. Thus, integrability should not be treated purely as a technical process, but as a complex challenge shaped by cultural norms, institutional readiness, and professional autonomy [16,40,74].

To address this, frequent involvement of HCPs during system design, continuous feedback loops, and adaptation to local workflows are crucial. Methods such as human-computer interaction with expert input [75] and consumer journey mapping [76] have been used to enhance AI-CDSS integration. At the same time, it is necessary to develop a standardized diagnostic support framework that aligns AI with specific clinical needs [77]. Another promising direction to enhance integrability is dynamic adaptation, where AI-CDSS adjust their level of support based on contextual factors such as patient volume, emergency status, or resource availability [22,78]. In high-pressure situations (eg, during emergencies or when clinician workload is high), the AI system could provide more proactive or autonomous recommendations. Conversely, during low-acuity periods, it could take a more supportive or background role, allowing clinicians greater control. Such adaptability can reduce disruption, improve acceptance, and ensure that AI interventions align with real-time clinical needs and capacities.

Ethical Concerns

In addition, the ethical implications of explainability and integrability should not be overlooked. Explainability is ethically significant in supporting informed consent, accountability, and clinicians' ability to critically evaluate AI recommendations

[14]. When clinicians can understand how an AI system reaches its conclusions, they are better equipped to maintain professional autonomy and protect patient rights. Integrability also has ethical implications. If an AI system is not well integrated into clinical workflows, clinicians may not know when or how to use it properly. This can create confusion about who is responsible for decisions influenced by AI. For example, if an AI recommendation appears at the wrong time in the workflow or is difficult to interpret in context, a clinician might follow it without full understanding or ignore it when it should have been considered. In both cases, the boundaries of responsibility become blurred [22,79]. These issues highlight the necessity of designing AI systems that align not only with technical and operational requirements but also with core ethical principles, such as transparency, fairness, and trustworthiness in health care.

Implementation Strategies Based on the Exploration, Preparation, Implementation, and Sustainment Framework

As noted in the previous section, AI faces persistent challenges in explainability and system integration. These cannot be resolved through isolated interventions but require a structured, phased approach. The exploration, preparation, implementation, and sustainment framework—comprising exploration, preparation, implementation, and sustainment—offers a widely validated model for supporting health care technology adoption and enhancing clinician acceptance of AI systems [80]. To systematically address these challenges, we draw on the exploration, preparation, implementation, and sustainment framework, providing structured strategies for each phase of implementation.

In the exploration phase, institutions should identify clinical needs and collaborate with multidisciplinary teams (eg, physicians, nurses, and IT staff) to assess AI integration opportunities. Prioritizing explainability, especially alignment with clinical reasoning, is critical. Models supporting SHAP, LIME, or other visual local explanation tools are recommended, alongside qualitative feedback collection from end users [81].

The preparation phase focuses on resolving integration barriers and tailoring explanations for different roles. Multilevel explanation interfaces can accommodate varying expertise levels [32,57]. Close coordination with IT departments is essential to embed AI tools into existing EHR systems, minimizing manual input and workflow disruption [71,72]. Organizational readiness and role-specific training are also key to successful adoption [82].

In the implementation phase, a phased deployment strategy helps minimize disruption and support gradual clinician adaptation. AI-CDSS tools can first target low-risk, supportive tasks (eg, risk alerts and abnormal laboratory flagging), then gradually expand to core decision-making such as diagnosis or treatment support. To ensure alignment with clinical needs, implementation should combine performance metrics (eg, alert response times and override rates) with user feedback (eg, satisfaction ratings and suggestion boxes). Regular log reviews

and focus groups can surface usability issues and guide iterative improvement [83].

The sustainment phase focuses on the long-term integration of AI into clinical workflows. Continuous monitoring of system performance and user experience is essential to ensure sustained adoption [84]. As models evolve, transparent update mechanisms—such as automatically generated explanation revisions and change logs—should be maintained to support clinician trust and promote continued engagement with the system.

Strengths and Limitations

This study offers a systematic exploration of AI integrability from the HCPs' perspective, providing a conceptual framework to guide medical AI design and development. Unlike previous studies focusing on technical developers or researchers' perspectives [85,86], it emphasizes the needs of actual HCPs, such as physicians. It establishes an AI explainability framework based on their priorities in data preprocessing, model structure, and postprocessing. This approach facilitates the development of user-centered AI-CDSS, promoting its acceptance and use by HCPs.

The limitation of this systematic review is the limited number of quantitative studies, which restricts quantitative analysis and statistical inference. Future research should include high-quality quantitative studies to validate and complement the conclusions.

Although we used a comprehensive set of keywords, the decision not to use truncation (eg, an asterisk) may have led to the omission of some relevant studies. This choice was made

to maintain specificity, but it may have limited the search breadth. To address this, we supplemented the search with citation tracking. This limitation is acknowledged in the review to clarify the scope of our search strategy.

In addition, while emphasizing user perspectives, this review provides limited analysis of varying needs among different medical roles. This is partly because of the small number of eligible studies and the lack of detail regarding specific clinical tasks, settings, or user groups in many of the included papers. These limitations made it difficult to conduct deeper, context-sensitive analysis of HCPs' perceptions of explainability and integrability. In this regard, we recommend that future research draw on implementation frameworks such as the Consolidated Framework for Implementation Research or Promoting Action on Research Implementation in Health Service framework to better account for the contextual and role-specific factors that shape HCPs' experiences. These frameworks can support more nuanced analyses of clinical settings and tasks, guiding the development of AI tools that are better aligned with real-world practices.

Conclusions

In conclusion, the explainability and integrability of medical AI are key factors influencing its acceptance and use in clinical settings. On the basis of the user-centered conceptual framework proposed in this study, future AI design should focus on HCPs' needs to enhance explainability and integrability, thereby promoting HCPs' acceptance and use and improving its effectiveness in real-world clinical applications.

Acknowledgments

This study was supported by the Shenzhen Basic Research Program (Natural Science Foundation; JCYJ20240813115806009); the National Natural Science Program of China (project 72004066); the Humanities and Social Sciences Research Project; the Ministry of Education, China (24YJAZH086 and 24YJCZH284); the Knowledge Innovation Project of Wuhan (2023020201020471); the Teaching Research Project of Huazhong University of Science and Technology (2023146); the Teaching Ability Training Curriculum Development Project of Huazhong University of Science and Technology (202408) and the China Scholarship Council. The authors confirm that generative artificial intelligence tools were used only for English language polishing and not for drafting the original or revised versions of the manuscript, which were entirely generated by the authors.

Data Availability

The data used and analyzed in this study will be made available by the corresponding author upon reasonable request.

Authors' Contributions

YL was responsible for original draft preparation, formal analysis, and software support. CL contributed to conceptualization, writing—review and editing, and methodology. JZ assisted with formal analysis and software development. CX and DW both contributed to conceptualization and supervision. All authors provided substantive contributions to the writing and revision of the manuscript and approved the final version before submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy for artificial intelligence (AI) explainability and for AI integrability from the user perspective.

[DOCX File, 16 KB-Multimedia Appendix 1]

Multimedia Appendix 2

PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) checklist.

[\[DOCX File, 32 KB-Multimedia Appendix 2\]](#)

References

1. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng*. Oct 2018;2(10):719-731. [\[FREE Full text\]](#) [doi: [10.1038/s41551-018-0305-z](https://doi.org/10.1038/s41551-018-0305-z)] [Medline: [31015651](#)]
2. Chen M, Zhang B, Cai Z, Seery S, Gonzalez MJ, Ali NM, et al. Acceptance of clinical artificial intelligence among physicians and medical students: a systematic review with cross-sectional survey. *Front Med (Lausanne)*. Aug 31, 2022;9:990604. [\[FREE Full text\]](#) [doi: [10.3389/fmed.2022.990604](https://doi.org/10.3389/fmed.2022.990604)] [Medline: [36117979](#)]
3. Delory T, Jeanmougin P, Lariven S, Aubert J, Peiffer-Smadja N, Boëlle PY, et al. A computerized decision support system (CDSS) for antibiotic prescription in primary care-Antibiocllic: implementation, adoption and sustainable use in the era of extended antimicrobial resistance. *J Antimicrob Chemother*. Aug 01, 2020;75(8):2353-2362. [doi: [10.1093/jac/dkaa167](https://doi.org/10.1093/jac/dkaa167)] [Medline: [32357226](#)]
4. Sambasivan M, Esmailzadeh P, Kumar N, Nezakati H. Intention to adopt clinical decision support systems in a developing country: effect of physician's perceived professional autonomy, involvement and belief: a cross-sectional study. *BMC Med Inform Decis Mak*. Dec 05, 2012;12(1):142. [\[FREE Full text\]](#) [doi: [10.1186/1472-6947-12-142](https://doi.org/10.1186/1472-6947-12-142)] [Medline: [23216866](#)]
5. Jeng DJ, Tzeng G. Social influence on the use of clinical decision support systems: revisiting the unified theory of acceptance and use of technology by the fuzzy DEMATEL technique. *Comput Ind Eng*. Apr 2012;62(3):819-828. [doi: [10.1016/j.cie.2011.12.016](https://doi.org/10.1016/j.cie.2011.12.016)]
6. Shibl R, Lawley M, Debusse J. Factors influencing decision support system acceptance. *Decis Support Syst*. Jan 2013;54(2):953-961. [doi: [10.1016/j.dss.2012.09.018](https://doi.org/10.1016/j.dss.2012.09.018)]
7. Jones C, Thornton J, Wyatt JC. Artificial intelligence and clinical decision support: clinicians' perspectives on trust, trustworthiness, and liability. *Med Law Rev*. Nov 27, 2023;31(4):501-520. [\[FREE Full text\]](#) [doi: [10.1093/medlaw/fwad013](https://doi.org/10.1093/medlaw/fwad013)] [Medline: [37218368](#)]
8. Burkart N, Huber MF. A survey on the explainability of supervised machine learning. *J Artif Intell Res*. Jan 19, 2021;70:245-317. [doi: [10.1613/jair.1.12228](https://doi.org/10.1613/jair.1.12228)]
9. Blezek DJ, Olson-Williams L, Missert A, Korfiatis P. AI integration in the clinical workflow. *J Digit Imaging*. Dec 22, 2021;34(6):1435-1446. [\[FREE Full text\]](#) [doi: [10.1007/s10278-021-00525-3](https://doi.org/10.1007/s10278-021-00525-3)] [Medline: [34686923](#)]
10. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med*. Jan 2019;25(1):30-36. [\[FREE Full text\]](#) [doi: [10.1038/s41591-018-0307-0](https://doi.org/10.1038/s41591-018-0307-0)] [Medline: [30617336](#)]
11. Ghanvatkar S, Rajan V. Evaluating explanations from AI algorithms for clinical decision-making: a social science-based approach. *IEEE J Biomed Health Inform*. Jul 2024;28(7):4269-4280. [doi: [10.1109/JBHI.2024.3393719](https://doi.org/10.1109/JBHI.2024.3393719)] [Medline: [38662559](#)]
12. Hua D, Petrina N, Young N, Cho J, Poon SK. Understanding the factors influencing acceptability of AI in medical imaging domains among healthcare professionals: A scoping review. *Artif Intell Med*. Jan 2024;147:102698. [\[FREE Full text\]](#) [doi: [10.1016/j.artmed.2023.102698](https://doi.org/10.1016/j.artmed.2023.102698)] [Medline: [38184343](#)]
13. Graziani M, Dutkiewicz L, Calvaresi D, Amorim JP, Yordanova K, Vered M, et al. A global taxonomy of interpretable AI: unifying the terminology for the technical and social sciences. *Artif Intell Rev*. Sep 06, 2023;56(4):3473-3504. [\[FREE Full text\]](#) [doi: [10.1007/s10462-022-10256-8](https://doi.org/10.1007/s10462-022-10256-8)] [Medline: [36092822](#)]
14. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak*. Nov 30, 2020;20(1):310. [\[FREE Full text\]](#) [doi: [10.1186/s12911-020-01332-6](https://doi.org/10.1186/s12911-020-01332-6)] [Medline: [33256715](#)]
15. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. *JAMA*. Dec 04, 2018;320(21):2199-2200. [doi: [10.1001/jama.2018.17163](https://doi.org/10.1001/jama.2018.17163)] [Medline: [30398550](#)]
16. Gu H, Liang Y, Xu Y, Williams CK, Magaki S, Khanlou N, et al. Improving workflow integration with xPath: design and evaluation of a human-AI diagnosis system in pathology. *ACM Trans Comput Hum Interact*. Mar 17, 2023;30(2):1-37. [doi: [10.1145/3577011](https://doi.org/10.1145/3577011)]
17. Mandl KD, Gottlieb D, Mandel JC. Integration of AI in healthcare requires an interoperable digital data ecosystem. *Nat Med*. Mar 30, 2024;30(3):631-634. [doi: [10.1038/s41591-023-02783-w](https://doi.org/10.1038/s41591-023-02783-w)] [Medline: [38291298](#)]
18. Interoperability in healthcare. Healthcare Information and Management Systems Society. URL: <https://legacy.himss.org/resources/interoperability-healthcare> [accessed 2025-05-09]
19. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bannetot A, Tabik S, Barbado A, et al. Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion*. Jun 2020;58:82-115. [doi: [10.1016/j.inffus.2019.12.012](https://doi.org/10.1016/j.inffus.2019.12.012)]
20. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit Med*. 2020;3:17. [\[FREE Full text\]](#) [doi: [10.1038/s41746-020-0221-y](https://doi.org/10.1038/s41746-020-0221-y)] [Medline: [32047862](#)]

21. Dowding D, Mitchell N, Randell R, Foster R, Lattimer V, Thompson C. Nurses' use of computerised clinical decision support systems: a case site analysis. *J Clin Nurs*. Apr 05, 2009;18(8):1159-1167. [doi: [10.1111/j.1365-2702.2008.02607.x](https://doi.org/10.1111/j.1365-2702.2008.02607.x)] [Medline: [19320785](https://pubmed.ncbi.nlm.nih.gov/19320785/)]
22. Wang D, Wang L, Zhang Z, Wang D, Zhu H, Gao Y, et al. "Brilliant AI doctor" in rural clinics: challenges in AI-powered clinical decision support system deployment. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021. Presented at: CHI '21; May 8-13, 2021:1-18; Yokohama, Japan. URL: <https://dl.acm.org/doi/10.1145/3411764.3445432> [doi: [10.1145/3411764.3445432](https://doi.org/10.1145/3411764.3445432)]
23. Liaw WR, Ramos Silva Y, Soltero EG, Krist A, Stotts AL. An assessment of how clinicians and staff members use a diabetes artificial intelligence prediction tool: mixed methods study. *JMIR AI*. May 29, 2023;2:e45032. [FREE Full text] [doi: [10.2196/45032](https://doi.org/10.2196/45032)] [Medline: [38875578](https://pubmed.ncbi.nlm.nih.gov/38875578/)]
24. Dwivedi R, Dave D, Naik H, Singhal S, Omer R, Patel P, et al. Explainable AI (XAI): core ideas, techniques, and solutions. *ACM Comput Surv*. Jan 16, 2023;55(9):1-33. [doi: [10.1145/3561048](https://doi.org/10.1145/3561048)]
25. Ali S, Abuhmed T, El-Sappagh S, Muhammad K, Alonso-Moral JM, Confalonieri R, et al. Explainable Artificial Intelligence (XAI): what we know and what is left to attain Trustworthy Artificial Intelligence. *Inf Fusion*. Nov 2023;99:101805. [doi: [10.1016/j.inffus.2023.101805](https://doi.org/10.1016/j.inffus.2023.101805)]
26. Miller T. Explanation in artificial intelligence: insights from the social sciences. *Artif Intell*. Feb 2019;267:1-38. [doi: [10.1016/j.artint.2018.07.007](https://doi.org/10.1016/j.artint.2018.07.007)]
27. Srinivasan R, Chander A. Explanation perspectives from the cognitive sciences—a survey. In: *Proceedings of the 29th International Conference on International Joint Conferences on Artificial Intelligence*. 2020. Presented at: IJCAI '20; January 7-15, 2020:4812-4818; Yokohama, Japan. URL: <https://dl.acm.org/doi/abs/10.5555/3491440.3492110> [doi: [10.24963/ijcai.2020/670](https://doi.org/10.24963/ijcai.2020/670)]
28. Pearson A, White H, Bath-Hextall F, Salmond S, Apostolo J, Kirkpatrick P. A mixed-methods approach to systematic reviews. *Int J Evid Based Healthc*. Sep 2015;13(3):121-131. [doi: [10.1097/XEB.0000000000000052](https://doi.org/10.1097/XEB.0000000000000052)] [Medline: [26196082](https://pubmed.ncbi.nlm.nih.gov/26196082/)]
29. Marco-Ruiz L, Hernández MÁ, Ngo PD, Makhlysheva A, Svenning TO, Dyb K, et al. A multinational study on artificial intelligence adoption: clinical implementers' perspectives. *Int J Med Inform*. Apr 2024;184:105377. [FREE Full text] [doi: [10.1016/j.ijmedinf.2024.105377](https://doi.org/10.1016/j.ijmedinf.2024.105377)] [Medline: [38377725](https://pubmed.ncbi.nlm.nih.gov/38377725/)]
30. Panagoulas DP, Virvou M, Tsihrintzis GA. An empirical study concerning the impact of perceived usefulness and ease of use on the adoption of AI-empowered medical applications. In: *Proceedings of the 23rd International Conference on Bioinformatics and Bioengineering*. 2023. Presented at: BIBE '23; December 4-6, 2023:338-345; Dayton, OH. URL: <https://ieeexplore.ieee.org/document/10431843> [doi: [10.1109/bibe60311.2023.00062](https://doi.org/10.1109/bibe60311.2023.00062)]
31. Zheng L, Ohde JW, Overgaard SM, Brereton TA, Jose K, Wi C, et al. Clinical needs assessment of a machine learning-based asthma management tool: user-centered design approach. *JMIR Form Res*. Jan 15, 2024;8:e45391. [FREE Full text] [doi: [10.2196/45391](https://doi.org/10.2196/45391)] [Medline: [38224482](https://pubmed.ncbi.nlm.nih.gov/38224482/)]
32. Wolf CT, Ringland KE. Designing accessible, explainable AI (XAI) experiences. *SIGACCESS Access Comput*. Mar 02, 2020;(125):1-1. [doi: [10.1145/3386296.3386302](https://doi.org/10.1145/3386296.3386302)]
33. Morais FL, Garcia AC, Dos Santos PS, Ribeiro LA. Do Explainable AI techniques effectively explain their rationale? A case study from the domain expert's perspective Publisher: IEEE. In: *Proceedings of the 26th International Conference on Computer Supported Cooperative Work in Design*. 2023. Presented at: CSCWD '23; May 24-26, 2023:1569-1574; Rio de Janeiro, Brazil. URL: <https://ieeexplore.ieee.org/document/10152722> [doi: [10.1109/cscwd57460.2023.10152722](https://doi.org/10.1109/cscwd57460.2023.10152722)]
34. Helman S, Terry MA, Pellathy T, Hravnak M, George E, Al-Zaiti S, et al. Engaging multidisciplinary clinical users in the design of an artificial intelligence-powered graphical user interface for intensive care unit instability decision support. *Appl Clin Inform*. Aug 04, 2023;14(4):789-802. [FREE Full text] [doi: [10.1055/s-0043-1775565](https://doi.org/10.1055/s-0043-1775565)] [Medline: [37793618](https://pubmed.ncbi.nlm.nih.gov/37793618/)]
35. Kinney M, Anastasiadou M, Naranjo-Zolotov M, Santos V. Expectation management in AI: a framework for understanding stakeholder trust and acceptance of artificial intelligence systems. *Heliyon*. Apr 15, 2024;10(7):e28562. [FREE Full text] [doi: [10.1016/j.heliyon.2024.e28562](https://doi.org/10.1016/j.heliyon.2024.e28562)] [Medline: [38576546](https://pubmed.ncbi.nlm.nih.gov/38576546/)]
36. Burgess ER, Jankovic I, Austin M, Cai N, Kapuścińska A, Currie S, et al. Healthcare AI treatment decision support: design principles to enhance clinician adoption and trust. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023. Presented at: CHI '23; April 23-28, 2023:1-19; Hamburg, Germany. URL: <https://dl.acm.org/doi/10.1145/3544548.3581251> [doi: [10.1145/3544548.3581251](https://doi.org/10.1145/3544548.3581251)]
37. Yoo J, Hur S, Hwang W, Cha WC. Healthcare professionals' expectations of medical artificial intelligence and strategies for its clinical implementation: a qualitative study. *Healthc Inform Res*. Jan 2023;29(1):64-74. [FREE Full text] [doi: [10.4258/hir.2023.29.1.64](https://doi.org/10.4258/hir.2023.29.1.64)] [Medline: [36792102](https://pubmed.ncbi.nlm.nih.gov/36792102/)]
38. Schoonderwoerd TA, Jorritsma W, Neerincx MA, van den Bosch K. Human-centered XAI: developing design patterns for explanations of clinical decision support systems. *Int J Hum Comput Stud*. Oct 2021;154:102684. [doi: [10.1016/j.ijhcs.2021.102684](https://doi.org/10.1016/j.ijhcs.2021.102684)]
39. Hong SR, Hullman J, Bertini E. Human factors in model interpretability: industry practices, challenges, and needs. *Proc ACM Hum Comput Interact*. May 29, 2020;4(CSCW1):1-26. [doi: [10.1145/3392878](https://doi.org/10.1145/3392878)]

40. Wenderott K, Krups J, Luetkens JA, Weigl M. Radiologists' perspectives on the workflow integration of an artificial intelligence-based computer-aided detection system: a qualitative study. *Appl Ergon*. May 2024;117:104243. [FREE Full text] [doi: [10.1016/j.apergo.2024.104243](https://doi.org/10.1016/j.apergo.2024.104243)] [Medline: [38306741](https://pubmed.ncbi.nlm.nih.gov/38306741/)]
41. Verma H, Mlynar J, Schaer R, Reichenbach J, Jreige M, Prior J, et al. Rethinking the role of AI with physicians in oncology: revealing perspectives from clinical and research workflows. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023. Presented at: CHI '23; April 23-28, 2023;22; Hamburg, Germany. URL: <https://dl.acm.org/doi/fullHtml/10.1145/3544548.3581506> [doi: [10.1145/3544548.3581506](https://doi.org/10.1145/3544548.3581506)]
42. Tonekaboni S, Joshi S, McCradden MD, Goldenberg A. What clinicians want: contextualizing explainable machine learning for clinical end use. *arXiv*. preprint posted online on May 13, 2019. 2019. [FREE Full text] [doi: [10.5260/chara.21.2.8](https://doi.org/10.5260/chara.21.2.8)]
43. Brennen A. What do people really want when they say they want "explainable AI?" we asked 60 stakeholders. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020. Presented at: CHI EA '20; April 25-30, 2020;1-7; Honolulu, HI. URL: <https://dl.acm.org/doi/10.1145/3334480.3383047> [doi: [10.1145/3334480.3383047](https://doi.org/10.1145/3334480.3383047)]
44. Fogliato R, Chappidi S, Lungren MP, Fisher P, Wilson D, Fitzke M, et al. Who goes first? Influences of human-AI workflow on decision making in clinical imaging. In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 2022. Presented at: FAccT '22; June 21-24, 2022;1362-1374; Seoul, Republic of Korea. URL: <https://dl.acm.org/doi/10.1145/3531146.3533193> [doi: [10.1145/3531146.3533193](https://doi.org/10.1145/3531146.3533193)]
45. Salwei ME, Carayon P, Hoonakker PL, Hundt AS, Wiegmann D, Pulia M, et al. Workflow integration analysis of a human factors-based clinical decision support in the emergency department. *Appl Ergon*. Nov 2021;97(3 Suppl. 2):103498-103495. [FREE Full text] [doi: [10.1016/j.apergo.2021.103498](https://doi.org/10.1016/j.apergo.2021.103498)] [Medline: [34182430](https://pubmed.ncbi.nlm.nih.gov/34182430/)]
46. Fischer A, Rietveld A, Teunissen P, Hoogendoorn M, Bakker P. What is the future of artificial intelligence in obstetrics? A qualitative study among healthcare professionals. *BMJ Open*. Oct 24, 2023;13(10):e076017. [FREE Full text] [doi: [10.1136/bmjopen-2023-076017](https://doi.org/10.1136/bmjopen-2023-076017)] [Medline: [37879682](https://pubmed.ncbi.nlm.nih.gov/37879682/)]
47. ISO 9241-11: 2018 ergonomics of human-system interaction part 11: usability: definitions and concepts. International Organization for Standardization. URL: <https://inen.isolutions.iso.org/obp/ui#iso:std:iso:9241:-11:ed-2:v1:en> [accessed 2025-02-28]
48. Cutillo CM, Sharma KR, Foschini L, Kundu S, Mackintosh M, Mandl KD, et al. MI in Healthcare Workshop Working Group. Machine intelligence in healthcare-perspectives on trustworthiness, explainability, usability, and transparency. *NPJ Digit Med*. Mar 26, 2020;3(1):47. [FREE Full text] [doi: [10.1038/s41746-020-0254-2](https://doi.org/10.1038/s41746-020-0254-2)] [Medline: [32258429](https://pubmed.ncbi.nlm.nih.gov/32258429/)]
49. Wang L, Zhang Z, Wang D, Cao W, Zhou X, Zhang P, et al. Human-centered design and evaluation of AI-empowered clinical decision support systems: a systematic review. *Front Comput Sci*. Jun 2, 2023;5:15. [doi: [10.3389/fcomp.2023.1187299](https://doi.org/10.3389/fcomp.2023.1187299)]
50. Mebrahtu TF, Skyrme S, Randell R, Keenan A, Bloor K, Yang H, et al. Effects of computerised clinical decision support systems (CDSS) on nursing and allied health professional performance and patient outcomes: a systematic review of experimental and observational studies. *BMJ Open*. Dec 15, 2021;11(12):e053886. [FREE Full text] [doi: [10.1136/bmjopen-2021-053886](https://doi.org/10.1136/bmjopen-2021-053886)] [Medline: [34911719](https://pubmed.ncbi.nlm.nih.gov/34911719/)]
51. Maleki Varnosfaderani S, Forouzanfar M. The role of AI in hospitals and clinics: transforming healthcare in the 21st century. *Bioengineering (Basel)*. Mar 29, 2024;11(4):337. [FREE Full text] [doi: [10.3390/bioengineering11040337](https://doi.org/10.3390/bioengineering11040337)] [Medline: [38671759](https://pubmed.ncbi.nlm.nih.gov/38671759/)]
52. Combi C, Amico B, Bellazzi R, Holzinger A, Moore JH, Zitnik M, et al. A manifesto on explainability for artificial intelligence in medicine. *Artif Intell Med*. Nov 2022;133:102423. [FREE Full text] [doi: [10.1016/j.artmed.2022.102423](https://doi.org/10.1016/j.artmed.2022.102423)] [Medline: [36328669](https://pubmed.ncbi.nlm.nih.gov/36328669/)]
53. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. *J Biomed Inform*. Jan 2021;113:103655. [FREE Full text] [doi: [10.1016/j.jbi.2020.103655](https://doi.org/10.1016/j.jbi.2020.103655)] [Medline: [33309898](https://pubmed.ncbi.nlm.nih.gov/33309898/)]
54. Shin D. The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI. *Int J Hum Comput Stud*. Feb 2021;146:102551. [doi: [10.1016/j.ijhcs.2020.102551](https://doi.org/10.1016/j.ijhcs.2020.102551)]
55. Xu Q, Xie W, Liao B, Hu C, Qin L, Yang Z, et al. Interpretability of clinical decision support systems based on artificial intelligence from technological and medical perspective: a systematic review. *J Healthc Eng*. Feb 03, 2023;2023(1):9919269. [FREE Full text] [doi: [10.1155/2023/9919269](https://doi.org/10.1155/2023/9919269)] [Medline: [36776958](https://pubmed.ncbi.nlm.nih.gov/36776958/)]
56. Moradi M, Samwald M. Post-hoc explanation of black-box classifiers using confident itemsets. *Expert Syst Appl*. Mar 2021;165:113941. [doi: [10.1016/j.eswa.2020.113941](https://doi.org/10.1016/j.eswa.2020.113941)]
57. Wang D, Yang Q, Abdul A, Lim BY. Designing theory-driven user-centric explainable AI. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019. Presented at: CHI '19; May 4-9, 2019;1-15; Glasgow, UK. URL: <https://dl.acm.org/doi/10.1145/3290605.3300831> [doi: [10.1145/3290605.3300831](https://doi.org/10.1145/3290605.3300831)]
58. Jacobs M, Pradier MF, McCoy TH, Perlis RH, Doshi-Velez F, Gajos KZ. How machine-learning recommendations influence clinician treatment selections: the example of the antidepressant selection. *Transl Psychiatry*. Feb 04, 2021;11(1):108. [FREE Full text] [doi: [10.1038/s41398-021-01224-x](https://doi.org/10.1038/s41398-021-01224-x)] [Medline: [33542191](https://pubmed.ncbi.nlm.nih.gov/33542191/)]
59. Liao QV, Varshney KR. Human-centered explainable AI (XAI): from algorithms to user experiences. *arXiv*. preprint posted online on October 20, 2021. 2021. [FREE Full text] [doi: [10.48550/arXiv.2110.10790](https://doi.org/10.48550/arXiv.2110.10790)]

60. Yang Q, Steinfeld A, Zimmerman J. Unremarkable AI: fitting intelligent decision support into critical, clinical decision-making processes. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019. Presented at: CHI '19; May 4-9, 2019:1-11; Glasgow, UK. URL: <https://dl.acm.org/doi/10.1145/3290605.3300468> [doi: [10.1145/3290605.3300468](https://doi.org/10.1145/3290605.3300468)]
61. Salih AM, Raisi - Estabragh Z, Galazzo IB, Radeva P, Petersen SE, Lekadir K, et al. A perspective on explainable artificial intelligence methods: SHAP and LIME. *Adv Intell Syst*. Jun 27, 2024;7(1):62. [doi: [10.1002/aisy.202400304](https://doi.org/10.1002/aisy.202400304)]
62. Alabi RO, Elmusrati M, Leivo I, Almangush A, Mäkitie AA. Machine learning explainability in nasopharyngeal cancer survival using LIME and SHAP. *Sci Rep*. Jun 02, 2023;13(1):8984. [FREE Full text] [doi: [10.1038/s41598-023-35795-0](https://doi.org/10.1038/s41598-023-35795-0)] [Medline: [37268685](https://pubmed.ncbi.nlm.nih.gov/37268685/)]
63. Frasca M, La Torre D, Pravettoni G, Cutica I. Explainable and interpretable artificial intelligence in medicine: a systematic bibliometric review. *Discov Artif Intell*. Feb 27, 2024;4(1):1-15. [doi: [10.1007/S44163-024-00114-7](https://doi.org/10.1007/S44163-024-00114-7)]
64. Delitzas A, Chatzidimitriou KC, Symeonidis AL. Calista: a deep learning-based system for understanding and evaluating website aesthetics. *Int J Hum Comput Stud*. Jul 2023;175:103019. [doi: [10.1016/j.ijhcs.2023.103019](https://doi.org/10.1016/j.ijhcs.2023.103019)]
65. Papadopoulos P, Soflano M, Chaudy Y, Adejo W, Connolly TM. A systematic review of technologies and standards used in the development of rule-based clinical decision support systems. *Health Technol*. May 27, 2022;12(4):713-727. [doi: [10.1007/S12553-022-00672-9](https://doi.org/10.1007/S12553-022-00672-9)]
66. Feng J, Phillips RV, Malenica I, Bishara A, Hubbard AE, Celi LA, et al. Clinical artificial intelligence quality improvement: towards continual monitoring and updating of AI algorithms in healthcare. *NPJ Digit Med*. May 31, 2022;5(1):66. [FREE Full text] [doi: [10.1038/s41746-022-00611-y](https://doi.org/10.1038/s41746-022-00611-y)] [Medline: [35641814](https://pubmed.ncbi.nlm.nih.gov/35641814/)]
67. Muschalik M, Fumagalli F, Hammer B, Hüllermeier E. Agnostic explanation of model change based on feature importance. *Künstl Intell*. Jul 12, 2022;36(3-4):211-224. [doi: [10.1007/S13218-022-00766-6](https://doi.org/10.1007/S13218-022-00766-6)]
68. Prado MD, Su J, Saeed R, Keller L, Vallez N, Anderson A, et al. Bonseyes AI pipeline—bringing AI to you: end-to-end integration of data, algorithms, and deployment tools? *ACM Trans Internet Things*. Aug 04, 2020;1(4):1-25. [doi: [10.1145/3403572](https://doi.org/10.1145/3403572)]
69. Jaspers MW, Smeulders M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. *J Am Med Inform Assoc*. May 01, 2011;18(3):327-334. [FREE Full text] [doi: [10.1136/amiain-2011-000094](https://doi.org/10.1136/amiain-2011-000094)] [Medline: [21422100](https://pubmed.ncbi.nlm.nih.gov/21422100/)]
70. Ahmadian L, van Engen-Verheul M, Bakhshi-Raiez F, Peek N, Cornet R, de Keizer NF. The role of standardized data and terminological systems in computerized clinical decision support systems: literature review and survey. *Int J Med Inform*. Feb 2011;80(2):81-93. [doi: [10.1016/j.ijmedinf.2010.11.006](https://doi.org/10.1016/j.ijmedinf.2010.11.006)] [Medline: [21168360](https://pubmed.ncbi.nlm.nih.gov/21168360/)]
71. Achour SL, Dojat M, Rieux C, Bierling P, Lepage E. A UMLS-based knowledge acquisition tool for rule-based clinical decision support system development. *J Am Med Inform Assoc*. Jul 01, 2001;8(4):351-360. [FREE Full text] [doi: [10.1136/jamia.2001.0080351](https://doi.org/10.1136/jamia.2001.0080351)] [Medline: [11418542](https://pubmed.ncbi.nlm.nih.gov/11418542/)]
72. Vorisek CN, Lehne M, Klopstein SA, Mayer PJ, Bartschke A, Haese T, et al. Fast Healthcare Interoperability Resources (FHIR) for interoperability in health research: systematic review. *JMIR Med Inform*. Jul 19, 2022;10(7):e35724. [FREE Full text] [doi: [10.2196/35724](https://doi.org/10.2196/35724)] [Medline: [35852842](https://pubmed.ncbi.nlm.nih.gov/35852842/)]
73. Longoni C, Bonezzi A, Morewedge CK. Resistance to medical artificial intelligence. *J Consum Res*. 2019;46(4):629-650. [FREE Full text] [doi: [10.1093/jcr/ucz013](https://doi.org/10.1093/jcr/ucz013)]
74. Reddy S, Shaikh S. The long road ahead: navigating obstacles and building bridges for clinical integration of artificial intelligence technologies. *J Med Artif Intell*. Mar 2025;8:7. [doi: [10.21037/jmai-24-148](https://doi.org/10.21037/jmai-24-148)]
75. Lee MH, Siewiorek DP, Smailagic A, Bernardino A, Bermúdez S. A human-AI collaborative approach for clinical decision making on rehabilitation assessment. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021. Presented at: CHI '21; May 8-13, 2021:1-14; Yokohama, Japan. URL: <https://dl.acm.org/doi/10.1145/3411764.3445472> [doi: [10.1145/3411764.3445472](https://doi.org/10.1145/3411764.3445472)]
76. LaMonica HM, Davenport TA, Ottavio A, Rowe SC, Cross SP, Iorfino F, et al. Optimising the integration of technology-enabled solutions to enhance primary mental health care: a service mapping study. *BMC Health Serv Res*. Jan 15, 2021;21(1):68. [FREE Full text] [doi: [10.1186/s12913-021-06069-0](https://doi.org/10.1186/s12913-021-06069-0)] [Medline: [33451328](https://pubmed.ncbi.nlm.nih.gov/33451328/)]
77. Goh E, Gallo R, Hom J, Strong E, Weng Y, Kerman H, et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. *JAMA Netw Open*. Oct 01, 2024;7(10):e2440969. [FREE Full text] [doi: [10.1001/jamanetworkopen.2024.40969](https://doi.org/10.1001/jamanetworkopen.2024.40969)] [Medline: [39466245](https://pubmed.ncbi.nlm.nih.gov/39466245/)]
78. Dapkins I, Prescott R, Ladino N, Anderman J, McCaleb C, Colella D, et al. A dynamic clinical decision support tool to improve primary care outcomes in a high-volume, low-resource setting. *NEJM Catal*. Mar 20, 2024;5(4):63. [doi: [10.1056/cat.23.0366](https://doi.org/10.1056/cat.23.0366)]
79. Morley J, Machado CC, Burr C, Cows J, Joshi I, Taddeo M, et al. The ethics of AI in health care: a mapping review. *Soc Sci Med*. Sep 2020;260:113172. [doi: [10.1016/j.socscimed.2020.113172](https://doi.org/10.1016/j.socscimed.2020.113172)] [Medline: [32702587](https://pubmed.ncbi.nlm.nih.gov/32702587/)]
80. Moullin JC, Dickson KS, Stadnick NA, Rabin B, Aarons GA. Systematic review of the Exploration, Preparation, Implementation, Sustainment (EPIS) framework. *Implement Sci*. Jan 05, 2019;14(1):1. [FREE Full text] [doi: [10.1186/s13012-018-0842-6](https://doi.org/10.1186/s13012-018-0842-6)] [Medline: [30611302](https://pubmed.ncbi.nlm.nih.gov/30611302/)]

81. Khedkar S, Gandhi P, Shinde G, Subramanian V. Deep learning and explainable AI in healthcare using EHR. In: Dash S, Acharya BR, Mittal M, Abraham A, Kelemen A, editors. Deep Learning Techniques for Biomedical and Health Informatics. Cham, Switzerland. Springer; 2020:129-148.
82. Abell B, Naicker S, Rodwell D, Donovan T, Tariq A, Baysari M, et al. Identifying barriers and facilitators to successful implementation of computerized clinical decision support systems in hospitals: a NASSS framework-informed scoping review. Implement Sci. Jul 26, 2023;18(1):32. [FREE Full text] [doi: [10.1186/s13012-023-01287-y](https://doi.org/10.1186/s13012-023-01287-y)] [Medline: [37495997](https://pubmed.ncbi.nlm.nih.gov/37495997/)]
83. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ. Apr 02, 2005;330(7494):765. [FREE Full text] [doi: [10.1136/bmj.38398.500764.8F](https://doi.org/10.1136/bmj.38398.500764.8F)] [Medline: [15767266](https://pubmed.ncbi.nlm.nih.gov/15767266/)]
84. Davis SE, Embí PJ, Matheny ME. Sustainable deployment of clinical prediction tools-a 360° approach to model maintenance. J Am Med Inform Assoc. Apr 19, 2024;31(5):1195-1198. [FREE Full text] [doi: [10.1093/jamia/ocae036](https://doi.org/10.1093/jamia/ocae036)] [Medline: [38422379](https://pubmed.ncbi.nlm.nih.gov/38422379/)]
85. Abdul A, Vermeulen J, Wang D, Lim BY, Kankanhalli M. Trends and trajectories for explainable, accountable and intelligible systems: an HCI research agenda. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018. Presented at: CHI '18; April 21-26, 2018:1-18; Montreal, QC. URL: <https://dl.acm.org/doi/10.1145/3173574.3174156> [doi: [10.1145/3173574.3174156](https://doi.org/10.1145/3173574.3174156)]
86. Holzinger A, Biemann C, Pattichis CS, Kell DB. What do we need to build explainable AI systems for the medical domain? arXiv. Preprint posted online on December 28, 2017. 2017. [FREE Full text]

Abbreviations

AI: artificial intelligence

AI-CDSS: artificial intelligence clinical decision support systems

EHR: electronic health record

HCP: health care professional

LIME: local interpretable model-agnostic explanations

SHAP: Shapley additive explanations

Edited by J Sarvestan; submitted 05.03.25; peer-reviewed by I Scharlau, L Zheng, HS Yun, M Khosravi, O Oloruntoba, C Udensi, S Kath; comments to author 08.04.25; revised version received 12.05.25; accepted 02.06.25; published 07.08.25

Please cite as:

Liu Y, Liu C, Zheng J, Xu C, Wang D

Improving Explainability and Integrability of Medical AI to Promote Health Care Professional Acceptance and Use: Mixed Systematic Review

J Med Internet Res 2025;27:e73374

URL: <https://www.jmir.org/2025/1/e73374>

doi: [10.2196/73374](https://doi.org/10.2196/73374)

PMID:

©Yushu Liu, Chenxi Liu, Jianing Zheng, Chang Xu, Dan Wang. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 07.08.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org>, as well as this copyright and license information must be included.