Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

AI-Driven EHR and Clinical Decision Support Integration for Reliable, Efficient, and Patient-Centered Care: A Scoping Review

Gideon Owusu
owusugideon2347@gmail.com
Michigan Technological University
David Tetteh Akuaku Blemano
Michigan Technological University
Frederick Damptey
Michigan Technological University
Benjamin Asomaning Odoom
Michigan Technological University

Research Article

Keywords: Artificial intelligence, electronic health records, clinical decision support systems, machine learning, clinical workflow, patient-centered care, healthcare interoperability, implementation

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11275950/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Artificial Intelligence (AI) and Electronic Health Records (EHRs) are increasingly working together to enhance predictive ability, decision-making, workflow efficiency, and deliver care. But good algorithmic performance is no guarantee of successful clinical integration or patient benefit.

Objective

The objective of the scoping review was to identify the current evidence regarding the use of AI to enhance EHRs and clinical decision support, exploring clinical applications, EHR integration, reliability and safety, workflow efficiency, patient-centered care and implementation maturity.

Methods

English language peer-reviewed articles between 2015 and 2026 were searched in 8 multidisciplinary information sources. The search was a combination of concepts involving AI, EHRs/EMRs and clinical decision support/workflow integration. There were 2,028 records identified with 184 being duplicates and 1,844 being screened. Sixty-two full-text reports were assessed for eligibility, of which 16 publications met the inclusion criteria and were included in the review. Evidence was compiled descriptively and thematically by care delivery domains and a continuum of evidence maturities.

Results

The types of AI-powered EHR/CDSS applications included risk prediction, early warning, diagnosis and prognosis, triage, treatment and medication support, documentation, and workflow optimization. Technical performance and selected reliability outcomes were best supported by evidence, while prospective evaluation, external validation, sustained deployment in the real-world, and directly measured outcomes from a patient perspective were less well developed. Reported benefits of workflow were noted, but interoperability concerns, alert and cognitive burden, usability, clinician trust, explainability and governance were still key implementation concerns.

Conclusion

AI has developed from prediction to clinically integrated decision support, but has not reached a level of translational maturity. Systems that are interoperable and prospectively evaluated and systemically human-supervised should be emphasized in future progress, and should be shown to have measurable advantages in safety, efficiency, clinician workflow, patient-centered outcomes, and real-world systems of care delivery.

1. Introduction

Over time, Electronic Health Records (EHRs) have transformed from a place to store and retrieve patient data into increasingly complex digital systems designed to facilitate communication, care coordination, clinical surveillance and decision-making. Extensive use of this has created extensive amounts of structured and unstructured clinical data that could be leveraged to enhance the timeliness, uniformity and personalization of care. These data have been augmented by the capabilities of AI technologies like machine learning (ML), deep learning, natural language processing (NLP), and more recently large language models (LLMs) that broaden the ability to extract clinically relevant information and provide predictions, classifications, alerts, and recommendations at various stages of the care journey. AI-powered EHR applications have grown to predict risks, warn of potential issues, diagnose patients, process incoming patient inputs, create treatment plans, safeguard patient medications, document patient care, and optimize workflows (Daskalo et al., 2026; Naderian et al., 2025; Talianu et al., 2026; Ye et al., 2024).

Clinical decision support systems (CDSSs) are very important interfaces between these analytical capabilities and clinical practice. Traditional CDSSs rely on patient data to trigger alerts, reminders, recommendations, diagnosis assistance or other information designed to support decision making. With the power of AI, these capabilities can be enhanced by recognizing intricate patterns in longitudinal clinical data, tailoring predictions to the individual patient, handling unstructured data, and generating more contextually tailored recommendations. Hence, AI combined with CDS enabled by EHRs can pave the way for a more proactive clinical information landscape. AI's potential is demonstrated in various clinical scenarios such as prostate cancer treatment (Naderian et al., 2025), sepsis diagnosis and intervention (Al-Juhani et al., 2025), perioperative risk prediction and triage (Chinn et al., 2026), rare-disease diagnosis (Özçetin et al., 2026), and bacterial infection management (Talianu et al., 2026).

The value of this integration lies not only in predictions, but in the entire process of creating a strategy. AI-powered CDS can enhance clinical reliability through earlier detection of deterioration, risk stratification, diagnostics, and safer treatment decisions. It could also increase efficiency by prioritizing information, minimizing repetitive cognitive tasks, coordinating work flow and funneling more time and attention to patients that need it. Importantly, opportunities for patient-centered care may exist when a variety of sources of clinical and patient generated information is included in decision making processes. For instance, Ye et al. (2024) found some potential benefits of integrating AI with EHRs and patient-generated health data for shared decision-making, patient-provider relationships, risk identification, and more accurate recommendations. From a practical standpoint, Owora et al. (2026) showed that a ML asthma risk marker integrated into an EHR could help improve the accuracy of pediatricians' prognosis and allow for assessment of usability,

acceptability and feasibility. The examples illustrate the importance of the clinical relevance of AI not only in predicting the disease, but also in how well the predictions are integrated in the decision making process as well as in workflows and patient interactions.

However, going from algorithmic capability to routine clinical value is still challenging. The literature on AI-EHRs has focused on the development and validation of models in a retrospective setting, with much of this work centered on discrimination, accuracy, or other technical performance metrics, but not on how a model works in clinical practice. Use of the data derived from EHR does not mean that it is an EHR, nor does it automatically lead to clinical decision support, although it can form an important basis for development of AI. Even a technically sound model can be of little clinical value if it produces results at the wrong time, is hard to interpret by clinicians, does not fit into clinical workflows, or cannot be supported by the necessary implementation processes. Emerging evidence increasingly highlights interoperability, workflow compatibility, usability, clinician acceptance, explainability, governance and prospective evaluation as key factors in successful clinical translation (Alruwaili et al., 2026; Joshi et al., 2022).

Such a translational challenge is reflected in more general reviews of clinically validated AI. AI applications employing imaging with physiological signals, EHRs, and wearable devices have the potential to aid in earlier disease detection and may offer clinically significant lead time (Uhoranishema et al., 2025). But, the validation, workflow integration, governance, user training and monitoring in real-world settings were the essential pillars of success in translation. The review also highlighted that there could be a degradation in the performance of the model at different institutions or at different time due to dataset shift, calibration, clinical environment differences, etc. The results confirm the difference between technically successful AI and clinically reliable AI – predictive use alone does not imply that an application is ready for clinical use and sustainable integration in everyday care.

Tensions are further revealed in real-world implementation which may not be as obvious when looking back at the models during assessment. When CDSSs are poorly targeted or lack the necessary connection to clinical practice, they can lead to alert fatigue, disruption of workflow, automation bias, and new types of technology-induced errors. While AI could potentially enhance the specificity and adaptability of decision support, it also raises potential concerns regarding algorithmic bias, transparency, generalizability, accountability, and overreliance on automated recommendations (Alruwaili et al., 2026). Experience with the clinician is, therefore, a critical aspect of implementation. In primary care, Goldberg et al. (2025) noted that workflow fit and usability, clinician burden, adoption, and implementation context are important considerations for EMR-integrated CDS. Likewise, Joshi et al. (2022) demonstrated that the model features, in addition to characteristics of the models themselves, also influence the implementation of sepsis prediction tools, such as the design of the alerts, acceptance of the tools by the clinicians, integration practices, and organizational barriers. These results help to frame the application of AI-powered CDS as a sociotechnical practice, rather than a standalone algorithm.

There is also an uneven evidence translational maturity. A few applications have advanced to either a prospective or real-world evaluation. For instance, Wang et al. (2026) assessed an EHR-integrated, LLM-based surgical triage approach within a clinical workflow, and Owora et al. (2026) explored the use of an EHR-integrated ML risk marker in a randomized clinical trial design. Retrospective or early stage studies are still the main source of information in other areas. In their systematic review of the use of LLMs in EHRs for paediatric and surgical care, Daskalo et al. (2026) found that most studies were conducted in a retrospective manner and that few studies assessed the models in real clinical workflows. Likewise, Chinn et al. (2026) identified that while AI-driven CDSSs are being studied for preadmission risk prediction, triage, and periadmission workflows are relatively under-researched, with prospective studies and embedding into workflows being less common. Technically sophisticated systems should not be taken as a sign that clinical integration, ongoing implementation, or health outcomes have proven.

Existing reviews have provided important but relatively segmented perspectives on this developing field. There have been a few reviews, but they are rather disjointed views of this evolving area. AI-driven CDSSs have been reviewed across specific clinical domains, such as prostate cancer (Naderian et al., 2025), perioperative care (Chinn et al., 2026), and rare disease (Özçetin et al., 2026), wound management (Sa'ban et al., 2026), and infection management (Talianu et al., 2026). Other syntheses have focused on specific technological or implementation aspects, such as using EHR data to support LLM applications (Daskalo et al., 2026), FHIR-based ML clinical information systems (Alruwaili et al., 2026), emergency triage CDSSs (Michel et al., 2024), workflow improvements powered by AI (Dave et al., 2025), medication safety and human-AI interaction (Ye et al., 2024), and using EHR and patient-generated data to aid clinical decisions (Alruwaili et al., 2026). Together, these studies illustrate significant advances and highlight the fact that the capacity of artificial intelligence, interoperability of EHRs, CDS functionality, workflow implementation, and clinical outcomes are often studied individually.

Importantly, there is no dearth of research on AI, EHRs, or clinical decision support alone, but rather a lack of overall synthesis of this entire translation process from AI capability to meaningful impact in care delivery. Whether AI applications move from the use of EHR data to "true" EHR integration, EHR integration to clinically actionable decision support, and clinically actionable decision support to proven improvements in reliability, efficiency, patient-centeredness, and routine clinical use is not a consistent pattern. This is especially relevant as there is a growing body of literature describing predictive models, which can give the illusion of clinical maturity, but with little evidence for prospective modeling assessment, workflow integration, interoperability, sustained use, or patient-level benefit.

A scoping review is thus the right type of review to map this diverse and fast-changing evidence base, categorize the various levels of translational maturity of this evidence, and identify gaps in the evidence for technical development that has or has not been scaled with meaningful integration into clinical practice. Therefore, this review will focus on four interrelated aspects of AI to enhance EHRs and support clinical decision making: clinical reliability and safety, workflow and operational efficiency, patient-centeredness, and implementation and translational immaturity. Instead of considering AI performance as the same as clinical effectiveness, the review assesses the journey from AI functioning to EHRs and the integration of AI tools into workflow, to actionable clinical decision support, workflow interaction, and care delivery outcomes.

The review will focus on answers to the following questions: (1) What are the AI-based EHR and clinical decision-support applications being studied in clinical specialties and in health care settings? (2) What is the connection between AI technologies and EHR/EMR systems and clinical workflows? (3) What is the evidence of their impact on clinical reliability and safety, efficiency in tasks, and patient-centered care? (4) What is the maturity of implementation and

evidence (technical validation, prospective evaluation, deployment in practice) of these applications? and (5) What methodological and methodological-technological, clinical and implementation gaps remain? The review aims to offer a comprehensive overview of the progress of AI-powered EHR and CDS technologies, moving from algorithm development to their clinical deployment, while considering the answers to the above questions as interdependent.

2. Methods

2.1 Review design

A scoping review design was used to map and synthesize evidence on the use of artificial intelligence (AI) and electronic health records (EHRs) and clinical decision support systems (CDSSs) to enhance clinical reliability and safety, workflow efficiency, patient-centered care, and real-world implementation. A scoping approach was felt to be appropriate given the diversity of AI technologies, clinical specialties, healthcare setting, study design, implementation phase and outcome measure in the literature and not every study could be aggregated statistically to achieve the primary aim.

This review was undertaken following a set of methodological guidance for scoping reviews which included identification of the review questions, systematic identification and selection of relevant evidence, data charting and synthesis and presentation of findings. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) was followed when reporting. Therefore, the methodological framework guided the process of conducting the review, while PRISMA-ScR was applied to report the review process and study selection in a transparent manner.

2.2 Information sources and search strategy

A systematic literature search was conducted in eight multidisciplinary information sources (PubMed/ MEDLINE, Scopus, Embase, Web of Science, IEEE Xplore, CINAHL, ScienceDirect, and ACM Digital Library). The search spanned English-language peer-reviewed literature published from January 2015 to 2026, to reflect the latest advances in AI-driven EHRs, machine learning, NLP, generative AI, LLMs, interoperability, and clinical decision support.

The search strategy consisted of three major concept blocks: (1) artificial intelligence, (2) electronic health records, and (3) clinical decision support and integration. Artificial Intelligence and other AI related terms featured variations of AI, machine learning, deep learning, natural language processing, generative AI, large language model and neural network. The terms associated with EHR were: Electronic health record, Electronic medical record, Electronic patient record, EHR, and EMR. The third concept included elements of clinical decision support, CDSS, clinical decision-making, AI-assisted decision-making, clinical workflow, workflow integration, and integration with EHR/EMR. Boolean, truncation and field-specific syntax were modified to suit the needs of the individual databases.

Reliability, efficiency, patient centeredness and implementation were considered as eligibility domains and synthesis domains, rather than as required search blocks, so as not to miss out on relevant studies reporting specific outcomes without using the broader terms.

2.3 Eligibility criteria

Empirical primary studies were eligible for inclusion if they were peer-reviewed and empirical, as well as secondary evidence syntheses (such as systematic reviews, scoping reviews, meta-analysis, and narrative or structured narrative reviews). Because the review sought to map evidence across heterogeneous clinical, technological, and implementation contexts, narrative and structured narrative reviews were not excluded solely on the basis of review type. For secondary publications, though, the methodological transparency had to be adequate to make it possible to assess how the evidence was identified and selected. To be eligible, a secondary review must be able to outline the scope or review question, the sources or literature search strategy employed to identify evidence, explicit criteria (or a clear enough process) for selecting the evidence that was included in the review, and a synthesis of the evidence of the secondary review that is traceable to the review questions. Narrative or structured narrative reviews that failed to give enough information to determine how the evidence included was identified and selected were excluded, on methodological grounds.

Eligibility was operationalized around four substantive criteria. The publication had to include some meaningful part of AI, whether it be machine learning, deep learning, natural language processing, generative AI, large language models, neural networks, or like technologies. Second, the AI application needed to show some relevance to an EHR/EMR environment as opposed to being a set of "retrospective" data derived from an EHR/EMR. Third, the application had to offer or assess clinically relevant decision support features such as risk stratification, alerts, early warning, triage, diagnostic or prognostic support, treatment suggestions, prescribing support, or similar clinician-facing features. Lastly, the publication was required to discuss at least one outcome that was either clinical reliability/safety, workflow/operational efficiency, patient-centered care, or implementation/real-world use.

2.4 Study selection

The database searches identified 2,028 records, comprising PubMed/MEDLINE (n = 1,743), Scopus (n = 102), Embase (n = 62), Web of Science (n = 51), IEEE Xplore (n = 25), CINAHL (n = 21), ScienceDirect (n = 6), and ACM Digital Library (n = 18). After eliminating 184 duplicate records from the data set, a total of 1,844 unique records were left for screening the titles and abstracts.

After the removal of duplicates, 1,844 unique records were entered into the title and abstract screening phase. The search strategy was designed to be as broad as possible in order to find articles in the overlapping fields of artificial intelligence, EHR, and CDS, thus identifying articles that would be relevant to one or more parts of the review question. Four reviewers took part in study selection, and two reviewers independently checked the records against the predetermined eligibility criteria. Disagreements were addressed by discussion and consensus, with the involvement of other reviewers as needed.

Title and abstract screening identified evidence that included a substantive AI component, clinically actionable decision support and meaningful integration with EHR/EMRs. Studies that were not primarily focused on clinical decision support or integration using EHR-derived data and did not involve AI model

development were excluded. When it was not possible to reliably determine eligibility based on the title and abstract, records were kept for assessment of the full text. At this stage, 1,782 of these 1,844 records were discarded and 62 records were requested from the reports.

Sixty-four reports was retrieved and assessed for full text screening. Of those, 16 publications met the inclusion criteria and were included in the review and 46 publications were excluded. The most common reasons for full text exclusion were that the publications included only EHR data or AI/model development without adequate, clinically integrated decision support (n = 17), a lack of methodological rigor for eligible reviews (n = 10), inadequate EHR/EMR integration (n = 8), ineligible publication types (n = 5), insufficient relevance to AI/ML (n = 3), failure to report an outcome of relevance to the review domains (n = 1), or failure to provide actionable AI-enabled clinical decision support (n = 2). For each full-text report that was excluded, a single exclusion reason was specified. The complete study-selection process is presented in the PRISMA-ScR flow diagram (Fig. 1).

2.5 Data charting

Data from eligible publications were charted using a standardized extraction framework developed around the review questions. Author and publication year, geographic setting, study design, publication design, clinical specialty, healthcare setting, function of CDSS, mode of integration with the EHR/EMR, approach to validation, and implementation status were extracted. Outcomes were tracked by clinical reliability and safety, workflow and operational efficiency, patient-centered care, and implementation/adoption.

Other data was retrieved regarding interoperability, usability, clinician acceptance, human oversight, implementation barriers and facilitators, prospective evaluation, real-world deployment, principal findings, and limitations (if applicable). Data charting and data synthesis were conducted separately by identifying primary empirical studies and secondary reviews. When the same underlying primary study evidence was represented in other primary studies in the included literature, findings reported through secondary reviews were coded as contextual or synthesis-level evidence, instead of as additional independent observations. The evidence from primary and secondary publications was, therefore, not brought together or added to as evidence of effect. This distinction remained in effect during interpretation, to minimize the danger of double counting of the same evidence.

2.6 Critical appraisal

A formal risk-of-bias assessment and/or study-level quality scoring was not performed. With the mapping purpose of this scoping review in mind, the purpose was to describe the breadth and nature of the evidence available, and its clinical integration and translation maturity, not to quantify a pooled treatment effect or to assess comparative effectiveness. The methodological aspects relevant to interpretation (study design, validation procedure, status of implementation, out-of-the-lab application, reported limitations) were instead plotted and taken into account during the descriptive and thematic synthesis. Analysis of findings was determined by the level of supporting evidence that was found, and evidence from retrospective, review-level or early-stage studies was not considered as the same as prospective or real-world clinical evaluation. As a result, the synthesis was designed to detect patterns and gaps in evidence and not to determine causal relationships or definitive comparative effectiveness.

2.7 Evidence synthesis

The evidence was synthesized descriptively and thematically rather than statistically pooled because of the diversity of AI technologies, clinical settings, study designs, EHR systems, and reported results. The synthesis was structured around four analytical domains as follows: (1) clinical reliability and safety, (2) workflow and operational efficiency, (3) patient-centered care and (4) implementation and translational maturity.

To separate technical development from the evidence of clinical value, evidence maturity was also taken into account along a continuum that ranges from model development, to internal validation, to external or temporal validation, to workflow and usability evaluation, to prospective clinical evaluation, to real-world deployment, to demonstrated care-delivery impact. This helped to distinguish between applications primarily focused on algorithmic performance and those that had advanced to the use of EHR in meaningful ways, to actionable clinical decision support, and to measurable impact in clinical care.

3. Results

3.1 Study selection

After excluding 184 duplicates, 1,844 unique records were then screened for their abstracts and titles and then further screened for their focus and eligibility. Sixty-two high priority reports were retrieved and fully assessed, 16 of which were eligible and 46 were excluded. The PRISMA-ScR flow diagram (Fig. 1) shows the study-selection process and the principle reasons for exclusion of full texts.

3.2 Characteristics of the included evidence

The 16 publications published in this period ranged from 2022 to 2026 (with the majority from 2025 onward), which reflects the recent focus of current research on integrating AI with EHR and CDS. Most included publications were secondary research, such as systematic and scoping reviews and a prospective evaluation of a LLM integrated with EHRs, with only a few primary studies. The evidence base was dominated by secondary evidence, that is, systematic, scoping reviews and a handful of qualitative implementation studies, a pilot randomized controlled trial and an evaluation of a LLM integrated with EHRs. The clinical contexts were diverse such as prostate cancer, sepsis and emergency care, surgical and peri-operative care, rare diseases, bacterial infection management, wound care, primary care, paediatric asthma, medication safety, and broader applications for healthcare workflow (Al-Juhani et al., 2025; Chinn et al., 2026; Goldberg et al., 2025; Naderian et al., 2025; Wang et al., 2026).

There was also diversity in the technologies involved. The most common approaches were machine learning-based prediction and risk stratification, followed by NLP, deep learning, generative AI and LLMs, knowledge-based CDS systems and hybrid systems. Integration of EHRs varied from incorporating structured and unstructured EHRs into prediction and decision support workflows to incorporating systems available to clinicians into their workflows. One architectural

path of evidence was an interoperability-focused approach, especially with FHIR (Fast Healthcare Interoperability Resources) (Balch et al., 2023). The properties of the publications included are listed in Table 1.

Table 1
Characteristics of the included evidence

Study	Evidence type	Country/context	Design	Clinical setting	Principal AI technology	EHR/CDSS application	Integr
Naderian et al. (2025)	Secondary	Multistudy/international	Systematic review	Prostate cancer	ML, neural networks, reinforcement learning	Diagnosis, prognosis, treatment planning	Mixed integr
Al-Juhani et al. (2025)	Secondary	Multistudy	Systematic review	Emergency/sepsis care	ML	Sepsis early warning and risk detection	EHR-li imple
Wang et al. (2026)	Primary	USA	Prospective quality-improvement study	Surgical care	LLM	Surgical comanagement triage	EHR-i huma
Özçetin et al. (2026)	Secondary	Multistudy	Scoping review	Rare-disease diagnosis	ML/AI, knowledge and ontology systems	Diagnostic reasoning/support	Heter limite integr
Talianu et al. (2026)	Secondary	Multistudy	Narrative review	Infection management	ML/AI	Diagnosis, prescribing and treatment support	Predo develo deplo
Sa'ban et al. (2026)	Secondary	Multistudy	Scoping review	Wound management	AI/ML	Assessment, documentation and CDS	Workf heter
Chinn et al. (2026)	Secondary	Multistudy	Scoping review	Preadmission/perioperative care	ML, DL, NLP	Risk prediction, triage and perioperative CDS	Mostl conce
Michel et al. (2024)	Secondary	Multistudy	Scoping review	Emergency telephone triage	ML/DL and knowledge-based CDS	Triage, orientation and prioritization	Limite conne
Goldberg et al. (2025)	Primary	USA	Qualitative study	Primary care	CDS including predictive/ML tools	EMR-integrated clinical support	Real-v imple
Owora et al. (2026)	Primary	USA	Pilot randomized clinical trial	Pediatric care	ML	Asthma prognostic risk marker	EHR-i
Dave et al. (2025)	Secondary	Multistudy	Scoping review	Healthcare workforce/workflow	NLP, ML, generative AI	Documentation, workflow and decision support	Mixed
Joshi et al. (2022)	Primary	USA	Qualitative implementation study	Hospital/sepsis care	ML and rule-based models	Sepsis surveillance and alerts	Real-v imple
Alruwaili et al. (2026)	Secondary	Multistudy	Structured narrative review	Medication safety	AI/ML-enabled CDS	Prescribing and adverse-event prevention	EHR-i
Daskalo et al. (2026)	Secondary	Multistudy	Systematic review	Pediatric and surgical care	BERT/transformers, ChatGPT and other LLMs	Documentation, classification and CDS	Predo retros
Balch et al. (2023)	Secondary	Multistudy	Scoping review	Clinical information systems	ML	FHIR-enabled analytics and CDS	Intero focus
Ye et al. (2024)	Secondary	Multistudy	Structured review	Patient-generated/EHR data	AI/ML	Integrated data for clinical decision support	EHR a gener integr

3.3 AI technologies and EHR/CDSS applications

Across the included evidence, AI-enabled EHR/CDSS applications clustered around four overlapping functions: risk prediction and early warning, diagnostic and prognostic support, triage and prioritization, and treatment or workflow support. ML-based prediction was particularly prominent. In prostate cancer, applications addressed diagnosis, recurrence and survival prediction, and radiotherapy or brachytherapy planning; gradient boosting, neural networks, and deep reinforcement learning were among the approaches reported (Naderian et al., 2025). Sepsis applications similarly used EHR-derived vital signs, laboratory results, and other clinical information for early-warning systems. Al-Juhani et al. (2025) reported AUROC values commonly above 0.80 and warnings occurring approximately two to four hours before clinical recognition in several studies.

More recent evidence showed progression towards generative and language-model based decision support. Daskalo et al. (2026) found BERT and its variants as well as ChatGPT in 52.3% and 29.5% of the 44 included studies respectively. However, only retrospective, unstructured data was used in 90.9% of the studies included. In contrast, Wang et al. (2026) examined an LLM system integrated with an EHR in a prospective manner. The SCM Navigator automated the electronic documentation and organization of pre-op information and added structured data to help classify surgical patients for the use of hospitalist comanagement without losing the physician review process.

The evidence also showed that the use of AI-EHR goes beyond prediction. In clinical documentation and workflow optimization, AI-supported systems have been employed in clinical documentation and workflow optimization, such as in an emergency department where AI is used to automate parts of the clinical documentation process, as described by Dave et al., 2025. In wound-management workflows, AI systems have been integrated to optimize documentation and facilitate compliance with workflow standards (Sa'ban et al., 2026). In medication safety, AI algorithms have been developed to detect medication errors in clinical records and to offer recommendations for correction (Alruwaili et al., 2026). For integrating patient-generated and EHR data for CDS, AI models have been created to transform patient-generated data into structured information that can support CDS (Ye et al., 2024). Based on the architecture, Balch et al. (2023) identified 39 FHIR-based ML clinical information systems, which comprised 18 CDSSs, 10 data-management and analytical platforms, and 11 auxiliary modules or application programming interfaces. Therefore, the evidence covered algorithm-level applications and infrastructures to enable interoperability of AI-based decision support with clinical information systems (CIS). Table 2 summarizes the AI technologies, integration features and clinical functions.

Table 2
AI technologies, EHR/CDSS integration characteristics and clinical decision-support functions

Study	AI technology/application	EHR/clinical data relationship	CDS function	Integration characteristics	Primary user/function
Naderian et al. (2025)	ML, neural networks, reinforcement learning	Clinical/EHR-supported cancer data	Diagnosis, prognosis, treatment planning	Integration heterogeneous; routine deployment limited	Clinician decision support
Al-Juhani et al. (2025)	ML early-warning models	EHR-derived vital signs, laboratory and clinical data	Sepsis prediction/early warning	EHR-linked alerts in some implementations	Emergency/acute-care clinicians
Wang et al. (2026)	LLM	Structured EHR data and preoperative documentation	Surgical triage/comanagement selection	Embedded EHR workflow with physician review	Surgical/hospitalist teams
Özçetin et al. (2026)	AI/ML and knowledge/ontology systems	Clinical and EHR-linked rare-disease information	Diagnostic support	Heterogeneous architectures; limited routine integration	Diagnostic clinicians
Talianu et al. (2026)	ML/AI	Clinical and EHR-related infection data	Diagnosis, prescribing and treatment support	Variable integration	Infection-management clinicians
Sa'ban et al. (2026)	AI/ML	Digital clinical/wound information	Assessment, documentation and CDS	Workflow integration constrained by interoperability	Multidisciplinary wound teams
Chinn et al. (2026)	ML, DL, NLP	Preoperative/EHR clinical information	Risk prediction and triage	Mostly developmental/proof-of-concept	Perioperative teams
Michel et al. (2024)	ML/DL and knowledge-based CDS	Telephone-triage and clinical data	Triage, prioritization and disposition	Real-world implementation in some systems; limited EHR connectivity	Emergency call personnel
Goldberg et al. (2025)	EMR-integrated CDS/predictive tools	Direct EMR integration	Clinical recommendations/support	Operational primary-care implementation	Primary-care clinicians
Owora et al. (2026)	ML risk prediction	EHR-integrated asthma risk information	Prognostic support	Embedded EHR risk marker	Pediatricians
Dave et al. (2025)	ML, NLP, generative AI	EHR/workflow information	Documentation and workflow support	Variable	Healthcare workers
Joshi et al. (2022)	ML and rule-based prediction	EMR clinical data	Sepsis risk alerts	Implemented across clinical sites	Clinicians/care teams
Alruwaili et al. (2026)	AI/ML-enabled CDS	Medication/EHR information	Prescribing support and adverse-event prevention	EHR/CDS interface	Nurses and physicians
Daskalo et al. (2026)	BERT/transformers, ChatGPT and other LLMs	Predominantly retrospective EHR text	Classification, documentation and CDS	Limited operational workflow evaluation	Clinical/analytical users
Balch et al. (2023)	ML	FHIR-standardized clinical information	Analytics and CDS	FHIR-enabled interoperability architecture	Clinical information systems
Ye et al. (2024)	AI/ML	EHR + patient-generated health data	Personalized CDS/shared decisions	Cross-source data integration	Clinicians and patients

3.4 Clinical reliability and safety

Evidence for reliability was strongest where studies measured predictive or decision-support performance, but considerably less consistent at the level of demonstrated patient outcomes. Wang et al. (2026) tested the LLM triage tool integrated with the EHR on 6,193 surgical cases, with a sensitivity of 0.94 and specificity of 0.74 when compared to the physician's determination. Importantly, post hoc review of the false negative cases found that only 2/19 cases were due to misclassification by the LLM, while discrepancies were due to clinical criteria, workflow characteristics, or variability among physicians.

Owora et al. (2026) gave randomized evidence at the clinician decision level. The EHR-integrated ML asthma risk marker improved pediatricians' prognostic accuracy compared with those who did not have access (0.83 versus 0.61; P = .008), with the tool showing positive usability and feasibility measures. Evidence with sepsis had similarly positive technical outcomes: Early-warning systems (EWS) for sepsis often performed better than conventional methods in detecting deterioration (Al-Juhani et al., 2025) and had an AUROC of > 0.80 in many cases. However, the outcomes of the patients were not universally favourable and prospective multicenter validation was limited.

The weaknesses were consistently found across the review evidence, external validation, calibration, generalizability, explainability, and human oversight. Despite favourable predictive performance, Naderian et al. (2025) found that the performance of prostate cancer AI-CDSSs was variable and external

validation was limited, as was real-world clinical implementation in rare-disease CDSSs, according to Özçetin et al. (2026). Daskalo et al. (2026) identified that there were fewer studies that examined clinical reliability, but where a comparator was included, LLMs performed better in 78.1% of the studies.

3.5 Workflow and operational efficiency

Evidence related to efficiency comprised direct measured impacts and/or potential or perceived impacts to workflow. Naderian et al. (2025) found a workflow CDSS which led to an operational benefit of 60% in decision-making time, which is measurable. Promptness of responding to the alert was also linked to AI sepsis alerts: Al-Juhani et al. (2025) found that in one multicenter AI implementation, time to antibiotics was shortened by around 1.8 hours when AI alerts were acted on quickly.

The overall evidence across the wider workflow indicated advantages with respect to documentation, administrative burden, cognitive load and information processing. In their review, Dave et al. (2025) identified clinical documentation as the most common workflow addressed and noted a correlation between the use of AI applications and reported benefits of a decreased administrative burden and increased workflow efficiency and job satisfaction. In multidisciplinary wound care, Sa'ban et al. (2026) also found that better documentation efficiency, increased diagnostic objectivity, and increased adherence to evidence-based guidance were aspects of the workflow that facilitate it.

Not all of these benefits were implemented equally. However, integration with workflow often led to new requirements. Sa'ban et al. (2026) found three challenges with the early adoption: cognitive workload, administrative workload, and lack of interoperability of EHRs, plus automation bias issues. Goldberg et al. (2025) found clinician resistance, poor infrastructure, organizational approval, and clinician burden to be barriers to EMR-integrated CDS adoption. Implementation was also found to be complex, time-consuming and interdisciplinary by Joshi et al. (2022) and there was considerable variation in how sepsis CDS tools were adopted into workflow. Measured efficiency gains were found in some applications and potential, perceived or context-specific efficiency gains were reported in much of the broader literature, but not in a consistent manner.

3.6 Patient-centered care

Patient-centered outcomes were less directly evaluated than technical performance, reliability, or workflow outcomes. The relevance for the patient was most clearly expressed in those areas where the use of AI-supported decisions had a direct impact on personalization, communication, shared decision-making, and the inclusion of patient-generated data. Ye et al. (2024) suggested that embedding patient-generated data into EHR systems could be a way to make CDS more personalized and to enhance shared decision making. Goldberg et al. (2025) conducted interviews with primary care clinicians and found that they were positive about improving the use of CDS when tools were used in conjunction with clinical judgment and patient preferences.

But, across the evidence included in this review, measures of direct patient experience, satisfaction, engagement, accessibility or shared decision making were rarely reported. A number of studies, rather, concluded that patient-centered benefit was achieved through benefits to diagnostic or prognostic decisions, early intervention, safer prescribing, or streamlined workflow. Further, concerns about the use of AI in wound care highlight the need for a human-in-the-loop approach, as evidence suggests that if not designed carefully, AI can diminish both relational and sensory components of care (Sa'ban et al., 2026). Overall, the representation of patient-centeredness as an intended consequence of improved clinical decision making was more common than as an outcome measured directly.

3.7 Implementation and evidence maturity

The included literature demonstrated marked variation in translational maturity. Much of the AI-EHR/CDSS research continued to be at the development and retrospective validation phase, with only a small number of studies moving on to evaluate workflows and then to prospective clinical evaluation and routine deployment. The most clear-cut example of this was the work by Daskalo et al. (2026), who found that while LLMs are increasingly being applied to EHR research, 90.9% of studies were retrospective, and pragmatic assessment of clinical workflows had been limited. In a separate study, Chinn et al. (2026) reported that most perioperative AI-CDSSs were still in the proof-of-concept phase with relatively few prospective or randomized evaluations.

A smaller group of evidence was more progressed in the translational pathway. Owora et al. (2026) prospectively assessed an EHR-integrated, human-in-the-loop, ML intervention, while Wang et al. (2026) prospectively assessed an operational human-in-the-loop LLM system integrated into EHRs. The Joshi et al. study (2022) also took place in 15 medical centers in the United States, which means they give real-world deployment data and not model-development data.

The challenge of interoperability was a significant tipping point from technical capability to deployment. Although the ML systems were designed to be interoperable with EHRs, they still lacked evidence of clinical efficacy, and lacked EHR interoperability, according to Balch et al. (2023). Likewise, 11 of the 19 emergency telephone-triage CDSSs identified by Michel et al. (2024) had been deployed in the real-world, but only 3 were linked to an EHR; and only 1 pragmatic and 1 randomized trial could be identified.

Immaturity of implementation was not a straightforward function of algorithmic sophistication across the evidence. Interoperability, external validation, workflow fit, clinician acceptance, usability, trust, explainability, human oversight, organizational resources and governance were found to be common determinants (Balch et al., 2023; Goldberg et al., 2025; Joshi et al., 2022; Özçetin et al., 2026). In the continuum of translation, evidence was highly localized at the development and validation phases, focused at a few points at the workflow and prospective evaluation phases, and relatively thin on the ground at the sustained real-world deployment and demonstrated care-delivery impact phases. Table 3 offers a cross domain synopsis of the primary care delivery and implementation findings. In addition, the resultant translation pattern is generated in Fig. 2.

The final stage encompasses three interrelated outcome dimensions: reliability and safety, efficiency, and patient-centeredness. While implementation maturity progresses from technical development and validation through prospective evaluation to real-world deployment and demonstrated clinical impact.

Table 3
Evidence of clinical reliability, workflow efficiency, patient-centered care and implementation among included studies.

Study	Reliability/safety	Efficiency/workflow	Patient-centered evidence	Implementation/maturity
Naderian et al. (2025)	Promising predictive performance; external validation limited	Some decision-time improvement	Limited direct evidence	Variable EHR integration; implementation gaps
Al-Juhani et al. (2025)	Strong discrimination in several sepsis models; clinical outcome evidence mixed	Earlier alerts/intervention reported	Mostly indirect	Some real-world systems; prospective evidence limited
Wang et al. (2026)	High sensitivity with physician comparator; discrepancies examined	Supports surgical triage	Indirect through appropriate care allocation	Prospective, EHR-integrated, human-in-the-loop
Özçetin et al. (2026)	Validation/generalizability limitations	Potential diagnostic efficiency	Potential earlier/more accurate diagnosis	Limited external validation and routine implementation
Talianu et al. (2026)	Potential improvement in infection decisions	Potential faster decision support	Primarily indirect	Translation/deployment remains variable
Sa'ban et al. (2026)	Diagnostic objectivity; automation-bias concerns	Documentation benefits but possible added burden	Human interaction may be affected	Interoperability, trust and workflow barriers
Chinn et al. (2026)	Promising risk prediction	Triage/workflow potential	Limited direct measurement	Predominantly proof-of-concept
Michel et al. (2024)	Evaluation heterogeneous	Triage/prioritization potential	Limited direct measurement	Some real-world deployment; limited EHR connectivity
Goldberg et al. (2025)	Clinicians emphasize judgment alongside CDS	Workflow fit and burden central	Patient preferences relevant	Adoption shaped by infrastructure, usability and organization
Owora et al. (2026)	Improved clinician prognostic accuracy	Favorable usability/feasibility	Indirect clinical benefit	Pilot randomized EHR-integrated evaluation
Dave et al. (2025)	Not primary outcome	Documentation/workload benefits reported	Mainly through workforce/care effects	Implementation heterogeneous
Joshi et al. (2022)	Safety dependent on alert design and trust	Implementation may create substantial work	Indirect	Real-world deployment; clinician acceptance critical
Alruwaili et al. (2026)	Medication-safety potential; technology-induced risk recognized	Alert burden/human vigilance important	Safer prescribing as indirect benefit	Human-AI interface and governance central
Daskalo et al. (2026)	Performance often favorable; clinical validity constrained by retrospective evidence	Potential documentation/CDS benefit	Sparse direct evidence	Predominantly retrospective; limited workflow evaluation
Balch et al. (2023)	Clinical efficacy evidence limited	Potential reusable/scalable integration	Sparse direct evidence	FHIR improves architecture but interoperability gaps remain
Ye et al. (2024)	Potential richer decision information	Potential data-integration benefits	Strongest emphasis on personalization/shared decisions	Integration of EHR and patient-generated data remains developing

4. Discussion

4.1 Principal findings

This scoping review revealed a fast-growing, but unevenly developed evidence base documenting the integration of AI tools into the EHR and clinical decision support. The five main findings were as follows. Second, AI applications are becoming more widespread and are being integrated into EHR-based decision making in a variety of clinical areas such as oncology, sepsis, peri-operative, rare diseases, infection management, wound management, primary care, medication safety, and pediatric risk prediction. But there were significant differences in integration maturity across applications. Second, technical development and retrospective validation occurred much more often than prospective clinical evaluations or long-term real-world implementation. Third, there was more evidence for the predictive reliability and for clinical decisions performance, than evidence for patient-centered outcomes. Fourth, the reported workflow benefits often co-occurred with new implementation challenges related to interoperability, clinician workload, alert design and trust, and organizational readiness. Lastly, only a small percentage of systems had reached a level of development where they could be evaluated for their impact on actual clinical decisions or care processes through the use of AI.

Combined, the results indicate that the question of whether AI-driven EHR and CDS are able to accurately predict has now become the secondary issue. But the more important question is whether such capabilities can be made into secure, usable, interoperable, and sustainable decision support tools in clinical settings. This separation was apparent throughout several reviews that were part of the package. For instance, Naderian et al. (2025) reported "generally favourable performance" of prostate cancer AI-CDSSs, but also noted some external validation of the tools and limited implementation of key elements of CDS, concern of tool interpretability, and poor integration with EHR-supported workflows. Despite the studies's performance gains over traditional methods,

Daskalo et al. (2026) identified that 90.9% of the studies that tested the performance of LLMs on pediatric or surgical EHR data remained retrospective. These patterns suggest a relationship between technical and translational progress but they are separable aspects of maturity.

4.2 Comparison with existing reviews

The present review's unique contribution is its focus on the level at which the evidence was linked, which is different from previous revisions. Existing reviews have tended to focus on one aspect of the AI-EHR-CDS pathway. Naderian et al. (2025) narrowed their focus to AI-CDSS development and performance in prostate cancer; Al-Juhani et al. (2025) discussed ML-based sepsis early-warning systems; Chinn et al. (2026) mapped AI-powered peri-operative risk prediction and triage; Daskalo et al. (2026) outlined the use of LLM using pediatric and surgical EHR data; and Balch et al. (2023) reviewed FHIR-enabled ML clinical information systems and interoperability. While these studies offer domain-specific or technology-specific insights, they are conducted in a variety of settings, making it challenging to assess whether AI is being developed in a coherent direction from model development to clinical adoption to care delivery impact.

The present synthesis explores the capability of AI, integration with EHR, actionable clinical decision support, interaction with workflow, and care-delivery outcomes as a continuous translational process, and identifies a recurring gap between the capability of AI and the actionable clinical decision support in the IDP.While examining the capability of AI, integration with EHR, actionable clinical decision support, interaction with workflow, and care-delivery outcomes as a connected translational pathway, this synthesis reveals a recurring discontinuity between the capability of AI and the actionable clinical decision support in the IDP. For instance, Michel et al. (2024) found 19 emergency telephone-triage CDSSs, with 11 being in the real-world, and only three with linkage to an EHR and only one pragmatic and one randomized trial. Likewise, Balch et al. (2023) found 39 clinical information systems (CIs) based on the FHIR standard, among which 18 were classified as CDS systems, but many of these CI systems did not have interoperability with EHRs, or had evidence of their clinical effectiveness that had yet to be validated external to the EHR community. This suggests that interoperable or clinical decision support systems can be designed but still not reach clinical use.

This integrated view of the world, therefore, provides an additional evidence mapping. Instead of questions about the existence of AI-CDSSs or the performance of algorithms, it focuses on the stage of the evidence's development: from data and algorithm to clinical value. It wasn't a smooth downgrade and a sure thing. Some systems were still in the model-development or retrospective-validation phases while others had developed to workflow evaluation, clinician facing implementation, randomized evaluation or real-world use. This maturity gradient is useful to compare otherwise diverse technologies and clinical areas.

Similarly, Uhoranishema et al. (2025) created an AI system that is clinically validated for early disease detection in the U.S. healthcare ecosystem. They found that clinically relevant AI is being applied to several image-based, physiological signal-based, and electronic health record (EHR) based, and even wearable, health data, and that generalizability, workflow integration, regulatory preparedness, equity, and post-deployment monitoring are important factors in adoption. The present review, on the other hand, is more narrowly targeted on the translation between AI capability, meaningful integration into EHRs, actionable clinical decision support, clinician and workflow engagement, and outcomes of care delivery. This wider integration process enables the separation of systems which are technically successful from those which have developed towards reliable and clinically usable systems.

4.3 From algorithmic performance to clinical integration

The need to separate model-centric evidence from care-delivery-centric evidence was one of the most obvious findings. The general AI-EHR literature focuses on assessing discrimination, model sensitivity, model specificity, AUROC, model classification accuracy, or other model performance metrics. Such outcomes matter, but do not necessarily imply clinical usefulness. Despite being accurate in terms of retrospective data, the model may be challenging to integrate, not clinically or workflow-timed, difficult to interpret, or not actually used in the clinic.

This difference is demonstrated with the sepsis evidence included. Al-Juhani et al. (2025) found that the AUROC values of the ML early-warning systems are often greater than 0.80, and that they were able to diagnose sepsis earlier than traditional systems, by two to four hours. Implementation studies also indicated quicker treatment, such as an average of 1.8 hours less time to antibiotics in one multicenter system. However, there was limited prospective validation and inconsistent patient-outcome benefits. Therefore, effective discrimination is not necessarily effective for the clinic.

Studies further down the translational continuum measured the impact of AI on clinician behavior or on clinician decision-making quality. Owora et al. (2026) showed that there was a generalized improvement in pediatrician prognostic accuracy, from 0.61 to 0.83, after introduction of the access to an EHR-integrated ML asthma risk marker, as well as favorable usability and feasibility. Wang et al. (2026) tested an EHR-integrated, human-in-the-loop LLM triage system in 6,193 surgical cases, finding it to be 94% sensitive and 74% specific compared to physician decisions. Unlike the studies that simply ask if AI can make a prediction, these studies ask if AI integrated into an operational workflow can result in a change of a clinically relevant decision.

As a result of this review, the evidence-maturity framework was created, which identifies different levels: AI capability, EHR integration, actionable CDS, workflow and clinician interaction, and care-delivery outcomes. Advancements in one stage of progress do not necessarily mean success at the next. It is a useful distinction between technological readiness and clinical readiness, and it is likely to make some sense of the co-existence of a huge AI literature, and a fairly small amount of clinical evidence of routine value.

4.4 Reliability, trust, and human oversight

The results also show that the clinical reliability is not synonymous with algorithmic correctness. But in practice, for any given task, reliability requires validation across populations and institutions, performance stability, interpretability, workflow context and appropriate human oversight. There were some recurring limitations, such as limited external validation, that were identified by several of the reviews included. Naderian et al. (2025) identified limited to no

external validation and calibration in prostate cancer AI-CDSS studies and Özçetin et al. (2026) noted that rare-disease CDSS studies were still hampered by cross-institutional validation, interoperability, explainability, and governance.

The article by Wang et al. (2026) is a welcome example of the ongoing need for human oversight even in cases of high model accuracy. Most of the discrepancies due to false negatives were not due to model failure alone, but modifiable clinical criteria, institutional workflows, and variation in physician practices. This discovery makes the simplistic categorisations between “correct AI” and “incorrect clinicians” more complicated. Clinical decision support is embedded in a sociotechnical context and relies on the interplay of model behavior, local policy, clinician judgment, data quality, and workflow design.

Rather, human-in-the-loop design is a means to safeguard the human system after the technical system has been developed. It's a component of the process by which AI's output becomes clinically interpretable and controllable. For systems that don't fit into a more intuitive narrative, clinician trust is particularly significant. Joshi et al. (2022) identified more confusion and distrust of the ML-based sepsis models compared to the RBS, and reported that adequate implementation required optimizing alerts, clinician buy-in, education, and expectation management. The findings indicated that the use of AI-CDSS needs to be grounded in transparency and trust both in terms of the content of the AI model itself but also in the transparency of the communication of how these tools operate, where uncertainty may arise and how clinicians should react to the recommendations.

4.5 Efficiency versus workflow burden

Efficiency emerged as a potentially important benefit of AI-enabled EHR/CDSS integration. Evidence indicates that AI can decrease and redistribute workloads. In certain places, direct efficiency improvements have been quantifiable. Naderian et al. (2025) found a workflow CDS application, which was correlated with a 60% time saving in decision making. In some implementation research, AI alerts led to quicker treatment, and a consistent finding across the various studies was that AI could help streamline clinical documentation, ease administrative burdens, and enhance workflow efficiency and job satisfaction.

But the benefit of workflow wasn't necessarily the same as the reduction of workload. In applications such as wound management, Sa'ban et al. (2026) identified greater efficiency in documentation and diagnostic objectivity, but noted that the initial implementation required more cognitive and administrative effort, limited interoperability, and automation-bias issues. Resistance by clinicians, lack of infrastructure, approval issues, workflow integration issues, and clinician burden were also identified as barriers to EMR-integrated CDS by Goldberg et al. (2025).

These results indicate that AI can automate one type of task and create a new one. Increased time saved in information retrieval, prediction, or documentation could be gained through alert review, review of AI output, a management of false positives, switching workflows, training needs, or added oversight. Therefore, future evaluations should separate task efficiency from net flow efficiency. A wider view of the implementation of healthcare-IT validates this distinction. Beyond predictive power, the potential impact of an ML-driven healthcare technology relies on workflow integration, a user-centred design, infrastructure, transparency, and end-user acceptance, according to Abaneme et al. (2025). Despite good technical performance, systems that do not fit current clinical practice could generate added operational costs. Their analysis is not in the context of clinical decision support but AI-EHR/CDSS is still applicable because of the emphasis on system fit, adoption needs, and organizational workload alongside performance. The length of automated task is a minimum measure technique that can miss out on the downstream work that can be generated by implementation.

4.6 Patient-centeredness as an underdeveloped evidence dimension

The least directly assessed of the domains explored in this review was patient centered care. While there were numerous studies that reported patients as being better “predicted”, earlier “diagnosed”, treated “safer”, or provided “efficient care”, few measured patient experience, preferences, engagement, communication, shared decision-making, equity, or satisfaction directly.

One of the most patient-centric viewpoints is provided by Ye et al. (2024), who argue for the integration of EHR data with patient-generated data, such as in the context of clinical decision support, when the integration is supported by AI. This integration may contribute to more personalized recommendations and shared decision-making by complementing the information in an EHR entered by a clinician. But, overall (across the larger included evidence), outcomes measured were more likely to be assumed than measured.

In a study by Goldberg et al. (2025), clinicians were positive about increased CDS adoption when CDS was integrated with clinical judgment and patient preferences. This is crucial as patient-centeredness does not need to be about AI making autonomous decisions for patients, but rather about AI helping clinicians to capture evidence and individual preferences more reliably. Meanwhile, Sa'ban et al. (2026) found that while digital addition can enhance patient-caregiver relationships, there is also a concern that AI systems could deteriorate these relationships when not integrated well with the care delivery system.

The lack of direct patient assessment of outcomes thus has a significant substantive evidence gap. The field seems a lot more advanced in its ability to predict outcomes of AI than in its ability to show what patients feel and perceive when receiving care with AI assistance. Patient-reported experience and outcome measures, shared-decision measures, communication outcome measures, accessibility and equity measures and patient trust measures (where appropriate to the intervention) should be incorporated into future studies.

4.7 Implications for research, practice, and system design

The results suggest a few implications for the next step in the development of AI-EHR/CDSS. For research, prospective, multicenter, externally validated studies should become a priority. While development is always needed in the future, there is a need to move the evidence beyond performance measures into comparative assessment of clinical decisions, safety, workflow impacts, patient outcomes, and sustained implementation. The low number of early stage evidence identified by Daskalo et al. (2026) and Chinn et al. (2026) suggests that real-world evaluation is not a routine part of AI development but rather a bottleneck. In order to integrate into clinical practice, systems need to be built around clinician workflow, not added on to it after being developed. Implementation needs are not just related to any single clinical algorithm, but also to the health care IT environment in which they are implemented. In

healthcare IT, Abaneme et al. (2025) highlight the importance of multidisciplinary engagement, user-centred design, seamless integration with existing workflows, interpretability, governance, and organisational preparedness for the effective and efficient adoption of ML. These factors emphasize the importance of implementing AI-EHR/CDSS as a sociotechnical problem-solving process and the necessity to build new technical performance, usability, infrastructure, professional acceptance, and governance, rather than one after the other.

According to Goldberg et al. (2025) and Joshi et al. (2022), key factors for successful adoption are interface usability, clinician education, alert optimization, technical assistance, organizational champions, trust, and the evidence of effectiveness. So instead, it is viewed as a co-design process with clinicians and other end users, and not as a final deployment step.

At the health-system level, interoperability is still at the core. However, Balch et al. (2023) demonstrated that the FHIR standards do not necessarily lead to interoperability or clinical effectiveness of EHRs. Health systems must have architectures that work across systems, secure data exchange, scalable infrastructure, monitoring, governance, and adaptable models to changing clinical data and workflows. Governance should not just cover approval, but also post-deployment monitoring for model drift, safety events, disparity in performance, and unexpected impacts on workflows.

The findings provide developers with a rationale and perspective to move beyond performance-based design to integration-based clinical value. The clinical decision and the workflow that the problem needs to be addressed should be the starting point for developing an AI-CDSS; then the minimum information needed for the problem to be solved should be determined, followed by the interpretation and action of the output and the definition of measurable care-delivery outcomes, and, finally, the choice of the modeling approach. A high predictive accuracy indicator is sufficient but not necessary. The more relevant question is: can AI be embedded within a clinical information system in a manner which contributes to decisions, lessens preventable burden, maintains human judgement, and generates measurable patient value?

5. Strengths and Limitations

5.1 Strengths

There are a number of strengths to this review. First, it had a multidisciplinary search approach using eight information sources from medicine, nursing, health informatics, engineering and computer science, fitting a topic that lies in both clinical and technical areas. Second, the criteria had to go beyond just having AI and EHR data. Publications were evaluated to determine if there was a substantive relationship between AI, EHR/EMR integration, clinically actionable decision support, and/or care-delivery/implementation outcomes, thereby separating out model development from clinically meaningful integration. This separation enhanced the review's capacity to gauge translation “maturity,” and not just to list predictive algorithms.

Another advantage was the independent screening and standardized eligibility criteria which included predetermined reasons for excluding the full text. In addition to the technical performance measures, the review looked at evidence on various dimensions of clinical value – reliability and safety, workflow and operational efficiency, patient centred care and implementation maturity. Primary empirical studies and secondary reviews were included in the analysis if they were appropriate, but were separated analytically to avoid the potential for double counting aspects of the same evidence. Last, the evidence-maturity approach offered a structured framework for comparing the heterogeneous studies with respect to the research phases from model development and validation to workflow evaluation, prospective clinical testing, real-world deployment and demonstrated care impact.

5.2 Limitations

There are a few constraints to note, as well. The review looked only at English-language publications and only at studies published between 2015 and 2026; this might have missed some relevant earlier articles written prior to 2015 or non-English evidence. Eight information sources were searched, but coverage and indexing differences in the databases might have led to some eligible studies being missed. The evidence base was very diverse in many aspects, such as clinical setting, AI methodology, study design, level of integration, outcome measurement, and so on, making it difficult to directly compare results from the studies and statistically combine the results throughout the evidence base.

The field is also dynamic, with new evidence arising following the publication of the review, and with generative AI and LLM technologies emerging, which may further change the evidence landscape in the near future. Also, there is the potential for overlap as there can be one or more review-level syntheses containing the same primary evidence. Data charting and data interpretation were therefore differentiated between ‘primary’ and ‘secondary’ publications and when ‘secondary-review’ findings were identified as secondary they were not regarded as additional independent observations when ‘primary’ evidence for the same observation could be found. However, since a formal "citation-overlap matrix" was not built, there is a possibility of some overlap in the secondary reviews and it is recommended that this be taken into account when interpreting the apparent breadth of the evidence base. A formal risk-of-bias assessment and study-level quality scoring were not performed. This was in line with the mapping aim of the scoping review, but precludes a formal, formal comparison of the methodological quality of individual studies or the certainty of the evidence. The conclusions should thus be read as a descriptive mapping of the existing evidence, and not as estimates of causal effect or as firm conclusions about comparative effectiveness. Finally, few studies included in many analyses assessed technical or workflow outcomes more fully than patient-centered outcomes, and conclusions regarding patient experience, shared decision making, equity, and other aspects of patient-centered outcomes that are directly measured were less robust.

6. Conclusion

Based on this scoping review, AI-enabled EHR and clinical decision support has come a long way from relying on electronic health data exclusively for retrospective algorithm development, but has not yet reached the stage of delivering reliable, efficient, patient-centered, and routinely integrated clinical decision support. The evidence shows more and more complex applications in many areas of risk prediction, early warning, diagnosis, triage, treatment

support, medication safety, and workflow optimization. But maturity is not uniform: technical development and retrospective validation are more prevalent than prospective testing, long-term deployment in the field, and impact on care delivery.

The results indicate that performance of the algorithm is not enough to assess clinical preparedness. The value of AI in the clinical context requires a more extensive translation cycle that links AI capability to EHR interoperability, actionable decision support, effective workflow integration, clinician uptake and oversight, and outcomes measurement. Direct evidence of patient-centered outcomes, such as patient experience, engagement, shared decision-making, equity, and satisfaction, is less well developed, while evidence of reliability and selected workflow benefits is emerging. Other factors that continue to limit the translation of technically promising systems to sustainable clinical use are interoperability, external validation, explainability, trust, governance, and workflow burden.

Future studies should therefore be less model based and more prospective, clinically integrated, and implementation based. Multicenter research is especially needed to assess real-world performance of AI-powered EHR and clinical decision support systems, as well as the performance of the healthcare and its outcomes. Evaluation should not be limited to predictive accuracy, but should also cover safety, reliability, workflow efficiency, clinician experience, patient-centred outcomes, interoperability, human oversight, and post-deployment monitoring. These should be taken into account during the development and implementation of the AI-CDSS process and not be considered as an after-thought when the model is already developed.

While AI has made strides from retrospective prediction based on healthcare data extracted from EHRs to more presence in clinical decision support, evidence of maturity along the entire pathway to routine clinical use is not consistent. Future steps should be measured not on the complexity of the models themselves but if the AI systems have the potential to be actionable, reliable, interoperable, trusted, efficient, patient-centric, and proven to be helpful in real-world care settings.

Declarations

7.1 Ethical Approval

Not applicable.

7.2 Consent to Participate

Not applicable.

7.3 Consent to Publish

Not applicable

7.4 Data Availability Statement

Not Applicable.

7.5 Funding

This research received no external funding.

7.6 Conflicts of Interest

The authors declare no conflict of interest.

7.7 Clinical Trial Number

Not Applicable

Author Contribution

G.O wrote the main manuscript text which include the Abstract, introduction, Methods, Results, and Conclusions and recommendations, however D.T.A.B contributed in writing the Methodology and F.D., B.O contrubuted in writing the Results. All authors reviewed the manuscript

References

1. Abaneme, O. G., Aboh, I., Moshood, M. R., & Assumang, D. K. (2025). Machine learning models for predictive risk assessment in healthcare IT projects. *Advances in Research, 26*(4), 372–384. https://doi.org/10.9734/air/2025/v26i41417
2. Al-Juhani, A., Desoky, R., Iskander, Z., Alshehri, K. T., Alshehri, A. A., Almuhaimid, A., et al. (2025). Advances in data-driven early warning systems for sepsis recognition and intervention in emergency care: A systematic review of diagnostic performance and clinical outcomes. *Cureus, 17*(8).
3. Alruwaili, M. F., Paul-Chima, U. O., & Nneoma, U. C. (2026). Integrating clinical decision support systems, nursing vigilance, and physician prescribing patterns to reduce preventable adverse drug events: A structured evidence-based narrative review on human-AI interface in medication safety. *Frontiers in Digital Health, 8*, 1831150.

4. Balch, J. A., Ruppert, M. M., Loftus, T. J., Guan, Z., Ren, Y., Upchurch, G. R., et al. (2023). Machine learning–enabled clinical information systems using fast healthcare interoperability resources data standards: Scoping review. *JMIR Medical Informatics, 11*, e48297.
5. Chinn, L. W., Nemeh, I., & Chinn, N. R. (2026). Artificial intelligence-enabled clinical decision support systems in preadmission testing: A scoping review of risk prediction, triage, and perioperative workflows (2020–2025). *Journal of Clinical Monitoring and Computing, 40*(2), 525–545.
6. Daskalo, C., Abu-Ashour, W., Tshimula, J. M., Amoei, M., Guadagno, E., & Poenaru, D. (2026). Large language models for electronic health records in pediatric and surgical care: A systematic review. *Journal of Pediatric Surgery*, 162956.
7. Dave, B., Martin, P., David, S. S., Kumar, S., & Chakraborty, T. (2025). Enhancing healthcare worker mental health via artificial intelligence-driven work process improvements: A scoping review. *International Journal of Medical Informatics*, 106122.
8. Goldberg, D. G., Soylu, T. G., Hoffman, C. F., Kishton, R. E., & Cronholm, P. F. (2025). Clinicians' perspectives on the adoption and implementation of EMR-integrated clinical decision support tools in primary care. *Digital Health, 11*, 20552076251334043.
9. Joshi, M., Mecklai, K., Rozenblum, R., & Samal, L. (2022). Implementation approaches and barriers for rule-based and machine learning-based sepsis risk prediction tools: A qualitative study. *JAMIA Open, 5*(2), ooac022.
10. Michel, J., Manns, A., Boudersa, S., Jaubert, C., Dupic, L., Vivien, B., et al. (2024). Clinical decision support system in emergency telephone triage: A scoping review of technical design, implementation and evaluation. *International Journal of Medical Informatics, 184*, 105347.
11. Naderian, S., Soleimanzadeh, F., Nikniaz, L., Sanaie, S., Sadeghi-Ghyassi, F., & Samad-Soltani, T. (2025). A systematic review of artificial intelligence-based clinical decision support systems in prostate cancer management. *Healthcare Technology Letters, 12*(1), e70026. https://doi.org/10.1049/htl2.70026
12. Owora, A. H., Jiang, B., & Shah, Y. (2026). Effect of an electronic health record-integrated machine learning asthma risk marker on pediatrician prognostic accuracy during preschool age: A pilot randomized clinical trial. *Scientific Reports, 16*(1), 20753.
13. Özçetin, E., Baş, S. E., & Özpay, F. (2026). Architectural and translational perspectives on clinical decision support systems for rare disease diagnosis: A scoping review. *International Journal of Medical Informatics*, 106442.
14. Sa'ban, F. Z., Purba, C. I. H., Rahayu, U., Aziz, M. A., Afriana, R., & Yudha, F. (2026). Artificial intelligence integration in multidisciplinary wound management: A scoping review of barriers and facilitators in clinical workflows. *Journal of Multidisciplinary Healthcare*, 619318.
15. Talianu, A., Fraser-Krauss, O., Bolton, W., Ming, D., Zhu, N., Hernandez, B., et al. (2026). Artificial intelligence for improving decision-making in bacterial infection management: A narrative review. *Journal of Antimicrobial Chemotherapy, 81*(1), dkaf470.
16. Uhoranishema, B., Odufuwa, P., Rabbles, K. E., Nsanzumukunzi, F., Rhoda, K. T., Shasere, B. E., & Ibrahim, M. S. (2025). Artificial intelligence applications for early disease detection in U.S. healthcare systems. *Journal of Advances in Medicine and Medical Research, 37*(12), 355–375. https://doi.org/10.9734/jammr/2025/v37i126024
17. Wang, J., Keyes, T., Liang, A. S., Ma, S. P., Shen, J., Liu, J., et al. (2026). An electronic health record–integrated, large language model–powered tool to triage surgical patients. *JAMA Network Open, 9*(8), e2629579.
18. Ye, J., Woods, D., Jordan, N., & Starren, J. (2024). The role of artificial intelligence for the application of integrating electronic health records and patient-generated data in clinical decision support. *AMIA Summits on Translational Science Proceedings, 2024*, 459.

Figures

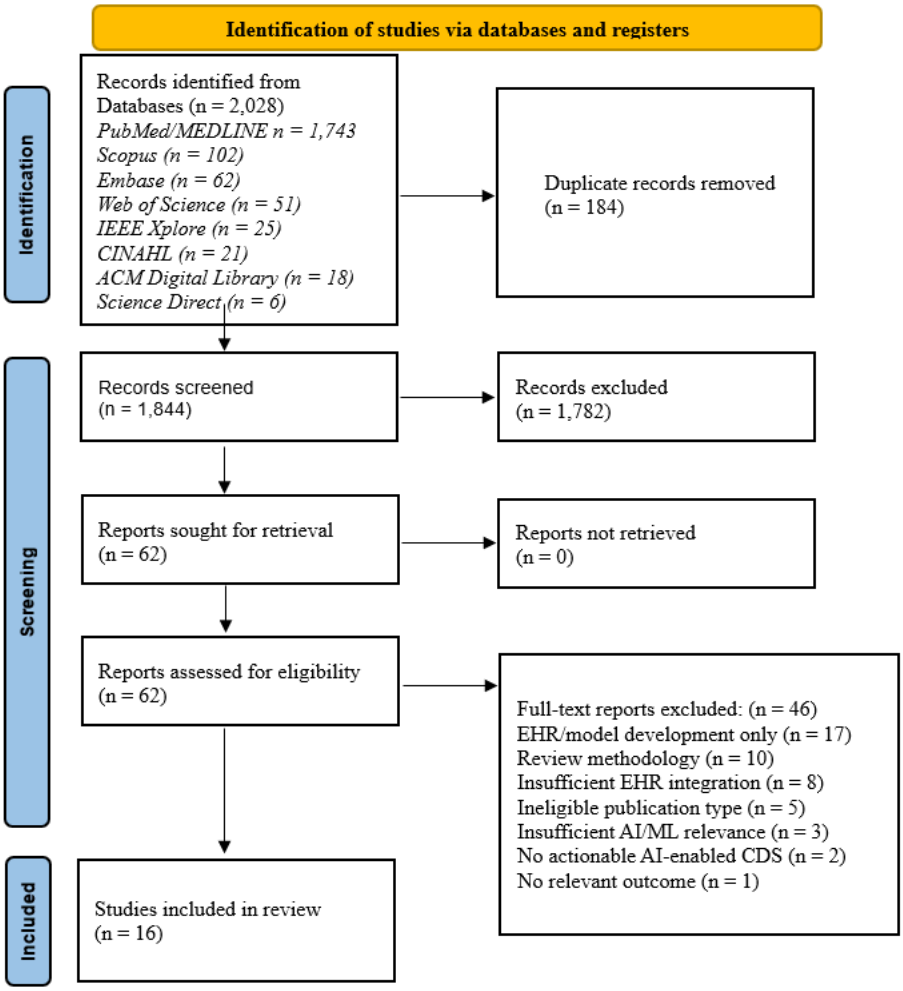


Figure 1

PRISMA-ScR flow



Figure 2

AI-EHR/CDSS translation and evidence-maturity framework from algorithmic capability to care-delivery impact (Author-developed conceptual framework derived from the scoping-review synthesis)