Saudi Journal of Medicine and Public Health

https://saudijmph.com/index.php/pub
https://doi.org/10.64483/202631861

AI-Driven Clinical Decision-Making Across Primary and Emergency Healthcare: A Narrative Review Integrating Family Medicine, Nursing Assessment, Pharmacological Safety, Resident Training, EMS Triage, and Diagnostic Radiology.

Arwa Mubark F Aldawsari[(1)] , Abdulrahman Eid Alrashdi[(2)] , Khloud Saleh Hamad Alhafdi[(3)] , Salman Abdulrahman Alotaibi[(4)] , Muneera Abdulrahman Alotaibi[(5)] , Hassan Abdulraouf Abdulrhem Faris[(6)] , Bushra Hassan Al-Buraiki[(7)] , Omar Marzouq Aloufi[(8)] , Rawan Alnogaidan[(9)] , Darwish Mohammed Baalgayth Almaghrabi[(10)] , Hassan Jaber Hassan Ammash[(11)]

(1)Al Imam Abdulrahman Al Faisal Hospital (IAFH), Ministry of Health, Saudi Arabia
(2)How Medical Clinic, Riyadh, Private Sector, Ministry of Health, Saudi Arabia
(3)Al-Ais Hospital, Al-Madinah Al-Munawwarah (Al-Ais Governorate), Ministry of Health, Saudi Arabia
(4)Riyadh First Health Cluster, Ministry of Health, Saudi Arabia
(5)Ministry of National Guard, Health Affairs (MNGHA), Dammam, Saudi Arabia
(6)Erada and Mental Health Hospital, Jazan City, Ministry of Health, Saudi Arabia
(7)King Abdullah bin Abdulaziz University Hospital, Riyadh, Ministry of Education, Saudi Arabia
(8)Maternity, Obstetrics, and Children's Hospital, Medina, Ministry of Health – Medina Health Cluster, Saudi Arabia
(9)Medical Cities of the Ministry of Interior, Riyadh, Ministry of Interior, Saudi Arabia
(10)Dammam Prison Health Center, Dammam, General Directorate of Prison Health Services – Ministry of Interior, Saudi Arabia
(11)Saudi Red Crescent Authority, Saudi Arabia

Abstract

Artificial intelligence (AI) is moving from retrospective validation toward routine use across primary care, acute care, and emergency services. This narrative review synthesises literature published between 2017 and 2026 on AI-driven clinical decision-making in six interdependent domains: family medicine, nursing assessment, pharmacy, resident physician training, emergency medical services (EMS), and diagnostic radiology. Evidence suggests that AI can support earlier diagnosis, risk stratification, deterioration detection, medication safety, triage accuracy, and imaging prioritisation. However, performance often declines under external validation, alert burden threatens adoption, and algorithmic and automation bias remain unresolved. Prospective evidence of improved patient outcomes remains limited to a few use cases. We propose an integrated, cross-professional framework emphasising shared data, human oversight, competency-based education, and lifecycle governance, and we outline implementation and research priorities for health systems seeking safe and equitable adoption.

Keywords: artificial intelligence; clinical decision support; emergency medical services; patient safety; medical education.

Introduction

Clinical decision-making is the central cognitive task of healthcare, and it is increasingly performed under conditions of information overload, time pressure, and workforce strain. Clinicians must integrate longitudinal records, laboratory results, imaging, physiological signals, medication lists, and patient preferences, often within minutes. Artificial intelligence (AI), and particularly machine learning (ML) and deep learning, has been proposed as a way to extract clinically useful patterns from these data and to return them to clinicians as predictions, prioritisation signals, or draft recommendations (Rajkomar et al., 2019; Topol, 2019). Over the past decade the field has progressed from proof-of-concept classifiers to a broad ecosystem that includes risk scores embedded in electronic health records (EHRs), autonomous screening devices, triage engines, and, most recently, large language models (LLMs) capable of summarising records and answering clinical questions (Haug & Drazen, 2023; Singhal et al., 2023).

The promise of these tools is frequently framed around single specialties, most prominently radiology and ophthalmology, where image-based deep learning has produced striking results (Esteva et al., 2019; Rajpurkar et al., 2022). Yet most patients do not experience care within one specialty. A person with chest pain may first consult a family physician, be assessed by a nurse, have medications reviewed by a pharmacist, be transported by EMS, undergo imaging interpreted by a radiologist, and be managed by a resident physician who is simultaneously learning the craft of diagnosis. AI introduced

*Corresponding author arwamubark20@gmail.com (Arwa Mubark F Aldawsari).
Receive Date: 5 September 2026, Revise Date: 30 October 2026, Accept Date: 16 November 2026

independently into each of these steps may improve local performance while creating fragmentation, duplicated alerts, inconsistent trust calibration, and unclear accountability across the pathway. A cross-professional perspective is therefore needed.

The evidence base also warrants caution. Systematic reviews have shown that many studies claiming parity between AI and clinicians were retrospective, poorly reported, and at high risk of bias, with few prospective trials or comparisons performed in real clinical settings (Liu et al., 2019; Nagendran et al., 2020). Reviewers have emphasised that the principal obstacles to clinical impact are no longer algorithm development but data quality, workflow integration, generalisability, and governance (Kelly et al., 2019; Sutton et al., 2020). Regulatory data show a rapidly growing number of approved AI-enabled devices, concentrated in imaging and often supported by limited prospective evidence (Muehlematter et al., 2021).

This narrative review therefore aims to (a) summarise how AI is being applied to clinical decision-making in family medicine, nursing, pharmacy, resident training, EMS, and diagnostic radiology; (b) appraise the strength and limitations of the supporting evidence; (c) identify cross-cutting risks, including bias, automation dependence, and alert fatigue; and (d) propose an integrated framework and research agenda. The domains are organised along the patient journey from continuity-based primary care through prehospital and in-hospital acute care, with resident education treated as the process through which the next generation of clinicians learns to work with these systems.

2. Review Approach

This article is a narrative review rather than a systematic review or meta-analysis. Narrative reviews are suited to broad, heterogeneous topics in which the aim is interpretive synthesis rather than pooled estimation. The review draws on peer-reviewed literature published between 2017 and 2026 addressing AI or ML applied to clinical decision support, diagnosis, risk prediction, monitoring, medication safety, triage, medical education, or image interpretation in the six domains of interest. Priority was given to randomised or prospective studies, multicentre validations, systematic and scoping reviews, and widely cited conceptual or policy analyses in high-quality journals. Retrospective model-development studies were included where they illustrated a mechanism, a limitation, or a pivotal advance.

Because the evidence was selected purposively, the review does not claim exhaustive coverage, and publication and selection bias are possible. Quantitative findings are reported as stated by the original authors and should be interpreted in the context of each study's population, comparator, and design. Throughout, the term *AI-driven decision-making* is used to denote AI-assisted decision-making in which a clinician remains responsible for the final decision, with the exception of a small number of autonomous diagnostic devices that are discussed explicitly.

3. Family Medicine: Early Diagnosis, Risk Stratification, and Continuity of Care

3.1 Early diagnosis in undifferentiated presentations

Family medicine is characterised by low disease prevalence, undifferentiated symptoms, and diagnostic uncertainty that unfolds over time. These features make early diagnosis difficult for clinicians and for algorithms alike. A scoping review of AI in primary care research concluded that the literature was dominated by early-stage development, with relatively few studies evaluating implemented tools in routine practice and with limited attention to primary care workflows (Kueper et al., 2020). This gap matters because models trained in hospital or specialty populations often perform poorly when applied to the lower-prevalence, less selected populations seen in the community.

The most mature primary care applications are those in which the task is narrow, the input is standardised, and the output is actionable. The autonomous detection of diabetic retinopathy is the clearest example. In a pivotal prospective trial conducted in primary care offices, an autonomous AI system achieved a sensitivity of 87.2% and a specificity of 90.7% for more-than-mild diabetic retinopathy, and it produced an interpretable result without specialist review in the large majority of participants (Abràmoff et al., 2018). The significance of this study lies less in the accuracy figure than in its demonstration that AI can relocate a screening task from a scarce specialist resource to the point of primary care, improving access for patients who might otherwise be lost to follow-up. Comparable approaches in dermatology, cardiology, and respiratory medicine remain at earlier stages of validation, although deep learning in these areas has shown clinician-level performance on curated datasets (Esteva et al., 2019; Rajpurkar et al., 2022).

LLMs have extended the debate from narrow classification to open-ended diagnostic reasoning. Benchmark studies show that such models can encode substantial clinical knowledge and answer licensing-style questions at a high level (Singhal et al., 2023), and they have been proposed as tools for generating differential diagnoses from free-text histories (Thirunavukarasu et al., 2023). For family practice, the attraction is the ability to prompt consideration of uncommon diagnoses in patients whose symptoms do not fit a typical pattern. The risk is equally clear: LLMs can produce fluent but incorrect reasoning, and benchmark performance does not guarantee safety in real consultations in which histories are incomplete and patients are heterogeneous.

3.2 Risk stratification and population health

Risk stratification is the second major role for AI in primary care. ML models applied to EHR data can estimate the likelihood of future events such as hospital admission, cardiovascular disease, or deterioration of chronic illness, supporting proactive outreach and resource allocation (Rajkomar et al., 2019). Compared with conventional regression-based tools, ML approaches can draw on high-dimensional longitudinal data, including diagnoses, laboratory trends, prescriptions, and encounter patterns, and can in principle update continuously as new information arrives.

The central caution is the choice of prediction target. In a widely cited analysis of a commercial algorithm used to allocate care-management resources to millions of patients, Obermeyer et al. (2019) showed that using healthcare cost as a proxy for health need led to systematically lower risk scores for Black patients than for White patients with the same burden of illness, because less money was historically spent on their care. The finding illustrates that a statistically well-calibrated model may still perpetuate structural inequity when its label encodes it. For family medicine, whose mission includes equitable, population-wide care, risk models must therefore be audited by subgroup, and their targets must be clinically meaningful outcomes rather than convenient administrative proxies (Char et al., 2018).

3.3 Continuity of care, documentation, and patient communication

Continuity is a defining value of family medicine and is also a significant source of information burden. Longitudinal records are lengthy, fragmented across care settings, and difficult to review in a ten-minute consultation. Generative AI offers the potential to summarise records, draft correspondence, prepare pre-visit briefs, and reduce documentation time, thereby giving clinicians more attention for the patient (Haug & Drazen, 2023). Early evaluations have also examined the communication quality of AI-generated text. In a cross-sectional study of public forum questions, evaluators preferred chatbot responses to physician responses for both quality and empathy in a majority of evaluations (Ayers et al., 2023), although the study did not assess accuracy against the full clinical context or patient outcomes and was limited to written answers to anonymous questions.

These results suggest a plausible role for AI in drafting patient-facing messages that clinicians review and personalise, but they do not support unsupervised use. A continuity relationship rests on trust and on the clinician's knowledge of the person, which no summary can fully replace. The appropriate design principle is that AI should preserve, rather than substitute for, the relational core of family medicine, absorbing clerical work while leaving interpretation, shared decision-making, and responsibility with the physician.4. Nursing: AI-Assisted Patient Assessment, Monitoring, and Deterioration Detection

4.1 The nursing role in AI-enabled care

Nurses spend the most time at the bedside, perform the bulk of vital-sign monitoring, and are frequently the first to recognise that a patient is worsening. An international think-tank of nursing and AI leaders concluded that nursing is underrepresented in AI development, governance, and research, and called for nurses to be engaged as co-designers rather than as end users of systems built without their input (Ronquillo et al., 2021). This is more than a representation issue. Nursing knowledge, including subtle cues about behaviour, appearance, and family concern, is often poorly captured in structured data, and models trained largely on physician-authored or administrative data may therefore miss what nurses actually observe.

AI-assisted assessment in nursing currently takes several forms: automated extraction of risk factors from documentation, predictive models for falls and pressure injury, continuous analysis of wearable or bedside monitor signals, and decision support at telephone triage and in community nursing. Many of these are still in early evaluation. The most consequential application, because it is the most directly linked to mortality, is the detection of clinical deterioration.

4.2 Deterioration and sepsis prediction

Early warning systems (EWS) based on vital signs have long been used to prompt escalation of care. ML-based variants aim to improve upon these by using many more variables and by modelling trajectories. A scoping review of ML-based EWS for clinical deterioration found a rapidly growing literature, but one dominated by retrospective development, with comparatively little evidence on prospective impact, human factors, or workflow effects (Muralitharan et al., 2021). Similarly, a systematic review and meta-analysis of ML models for sepsis prediction found promising discrimination in many settings but substantial heterogeneity and concerns about study quality, with performance depending strongly on the prediction window and the clinical setting (Fleuren et al., 2020).

The consequences of inadequate external validation became highly visible when a widely implemented proprietary sepsis model was evaluated in a large academic health system. Wong et al. (2021) reported poor discrimination (an area under the curve of 0.63), a low sensitivity, and alerts on roughly one in six hospitalised patients, a combination that would generate considerable alert burden while missing the majority of patients who developed sepsis. By contrast, a different system, the Targeted Real-time Early Warning System (TREWS), was evaluated prospectively across multiple hospitals. Adams et al. (2022) found that when clinicians confirmed the alert

within three hours, adjusted mortality was lower, with a relative reduction of 18.7%, although the observational design limits causal inference. A companion analysis of provider adoption found that clinician engagement with alerts was central to any benefit, and that adoption depended on how the system fit clinical work (Henry et al., 2022). Taken together, these findings suggest that the same technology class can be either unhelpful or beneficial depending on model quality, local calibration, and the extent to which front-line staff trust and act upon its output.

4.3 Alert fatigue and the nursing workflow

Alerts do not help if they are ignored. Research on clinical decision support alerts has repeatedly shown that repeated, low-value alerts reduce clinician acceptance; in one analysis of an EHR alerting system, the likelihood of accepting an alert fell with each additional reminder of the same type, and also with higher workload (Ancker et al., 2017). Because nurses receive many alarms and notifications from monitors, infusion pumps, and EHRs, an additional AI-generated alert imposes costs that must be weighed against its yield. Design choices that matter include thresholds calibrated to unit capacity, tiered escalation pathways, clear explanations of why the alert was triggered, and feedback loops that allow nurses to flag false alarms.

The ethical dimension is also relevant. Nursing judgment is grounded in holistic assessment, and algorithms that reduce the patient to a score may inadvertently reduce the value placed on that judgment. Governance frameworks should specify that a nurse's concern is sufficient grounds for escalation even when the algorithm is silent, and that low algorithmic risk never overrides clinical suspicion (Ronquillo et al., 2021).

5. Pharmacy: Medication Reconciliation, Prescribing Support, and Prevention of Drug-Related Errors

5.1 The medication safety problem

Medication errors and adverse drug events are among the most frequent and preventable causes of patient harm. Conventional clinical decision support, such as rule-based interaction and dose alerts, has improved prescribing safety but is notorious for low specificity, which produces overrides and alert fatigue (Ancker et al., 2017; Sutton et al., 2020). Pharmacists, who bear much of the burden of verifying orders and reconciling lists across transitions of care, are therefore among the clinicians most likely to benefit from smarter prioritisation.

A scoping review on AI and patient safety concluded that medication safety is one of the most active areas of AI safety research, with applications in the prediction of adverse drug events, the detection of prescribing errors, and the identification of high-risk patients, although many implementations had yet to be rigorously evaluated in clinical practice (Bates et al., 2021). A more specific scoping review of use cases to reduce the frequency of adverse drug events highlighted the opportunity to target prediction at the patient, drug, and system levels, while noting that the majority of work was still at the development or validation stage and that reporting on implementation outcomes was sparse (Syrowatka et al., 2022).

5.2 Prescription screening and alert optimisation

One of the most practical roles for ML in pharmacy is to prioritise which prescriptions most need expert review. Corny et al. (2020) developed an ML-based decision support system in a hospital setting to identify prescriptions at high risk of medication error, trained on prescriptions that pharmacists had previously flagged for intervention. The tool was designed to complement, not replace, rule-based alerts, directing pharmacist attention to orders that deviated from expected patterns. Such approaches exploit the fact that pharmacists' historical interventions represent a rich record of expert judgment about what constitutes a clinically significant problem. The limitations are equally instructive: models learn from the biases and blind spots of past review, and performance may not transfer to institutions with different formularies, patient populations, or prescribing habits.

Alert optimisation is a related opportunity. If ML can suppress low-value alerts, or promote high-risk ones, it may counter the alert fatigue that undermines the effectiveness of existing decision support (Ancker et al., 2017). Care must be taken that suppression does not remove alerts for rare but dangerous events, so any such system requires monitored deployment, explicit safety thresholds, and audit of overridden alerts.

5.3 Medication reconciliation and LLM-assisted workflows

Medication reconciliation, the process of establishing an accurate list of what a patient is actually taking, is labour-intensive and error-prone, particularly at admission, transfer, and discharge. Sources conflict, patients may not remember names or doses, and primary care, hospital, and community pharmacy records are rarely aligned. LLMs and natural-language processing can extract medication information from free-text notes, discharge summaries, and patient messages, and can flag discrepancies for pharmacist review (Haug & Drazen, 2023). Early enthusiasm should be tempered by the well-described tendency of LLMs to produce plausible but incorrect output and by the sensitivity of their performance to prompting and context (Thirunavukarasu et al., 2023). In medication contexts, a single hallucinated dose or omitted anticoagulant can be harmful.

The prudent model is therefore one of AI-prepared, pharmacist-verified reconciliation. The system proposes a reconciled list with provenance links to the source documents, and the pharmacist confirms or

corrects it with the patient. Such an approach captures time savings while preserving human verification at the point of highest risk. It also fits the broader ethical requirement that accountability for medication decisions remains clearly assigned (Char et al., 2018).

Table 1 summarises the principal applications, evidence, and risks across the six domains reviewed in this article.

Table 1. Summary of AI applications, evidence status, and principal risks across six clinical domains

Domain	Representative AI applications	Evidence status (selected sources)	Principal risks
Family medicine	Autonomous screening (e.g., diabetic retinopathy); EHR risk stratification; record summarisation; draft patient communication	Prospective validation for narrow screening tasks (Abràmoff et al., 2018); most primary care AI still early-stage (Kueper et al., 2020)	Label bias in risk models (Obermeyer et al., 2019); hallucination; erosion of relational continuity
Nursing	Early warning and sepsis prediction; fall and pressure injury risk; monitoring signal analysis	Prospective multi-site data for one system (Adams et al., 2022); poor external validation for another (Wong et al., 2021)	Alert fatigue (Ancker et al., 2017); under-representation of nursing data and voice (Ronquillo et al., 2021)
Pharmacy	Prescription risk screening; alert prioritisation; medication reconciliation; adverse drug event prediction	Mostly development and validation stage (Bates et al., 2021; Syrowatka et al., 2022); single-centre ML screening (Corny et al., 2020)	Missed rare events; LLM errors in doses; learned bias from past review
Resident training	LLM-supported reasoning; tutoring and feedback; simulated cases; diagnostic second opinions	Randomised trial showed no significant gain from LLM access alone (Goh et al., 2024); AI literacy called for (Wartman & Combs, 2018)	Automation bias (Gaube et al., 2021); deskilling (Cabitza et al., 2017)
EMS	Dispatcher recognition of cardiac arrest; prehospital risk prediction; triage; destination selection	Randomised trial without improvement in dispatcher recognition (Blomberg et al., 2021); ML prediction of critical care need (Kang et al., 2020)	Poor generalisation; limited prehospital data; accountability under time pressure
Diagnostic radiology	Detection and triage of critical findings; mammography screening support; report prioritisation	Large screening trial and multireader studies (Lång et al., 2023; McKinney et al., 2020; Seah et al., 2021)	Dataset shift; hidden bias; over-reliance; limited outcome data

6. Resident Physician: AI-Assisted Clinical Reasoning, Diagnostic Accuracy, and Competency Development

6.1 Reasoning and diagnostic accuracy

Residency is the period in which clinical reasoning is consolidated, and it is therefore the period in which exposure to AI is likely to have the greatest formative effect. Wartman and Combs (2018) argued that medical education must shift from emphasising memorisation of information to cultivating skills in data interpretation, judgment, and collaboration with intelligent systems, because the volume of medical knowledge now exceeds what any individual can retain. LLMs make this argument concrete. Models have achieved passing-level performance on licensing-style examinations and can generate structured differential diagnoses (Singhal et al., 2023; Thirunavukarasu et al., 2023), which suggests a role as an always-available tutor or second opinion for trainees working at night or in under-resourced settings.

The evidence from controlled studies is more nuanced than benchmark results suggest. In a randomised clinical trial of 50 physicians, including residents, access to an LLM in addition to conventional resources did not significantly improve diagnostic reasoning scores relative to conventional resources alone, whereas the LLM on its own scored considerably higher than either physician group (Goh et al., 2024). The result implies that capability of the model and benefit to the human-AI team are distinct quantities. Clinicians may not know how to query a model effectively, may discount correct suggestions that conflict with their initial impression, or may use the tool as a search engine rather than as a reasoning partner. Educational design, rather than technology alone, determines whether capability is translated into improved decisions.

6.2 Automation bias and deskilling

The corresponding risk is that trainees come to rely on AI before they have developed independent reasoning. Cabitza et al. (2017) warned that machine learning in medicine could produce unintended consequences, including automation bias and the loss of clinical skills when decision support is accepted uncritically. Experimental evidence supports concern about susceptibility to flawed advice. Gaube et al.

(2021) found that clinicians' diagnostic performance suffered when the advice they received was inaccurate, whether it was labelled as coming from an AI system or from a human. Likewise, in a study of skin cancer recognition, human-computer collaboration improved overall accuracy, but the benefit depended on the quality of the AI output and on the way the information was presented, and erroneous AI support could mislead readers (Tschandl et al., 2020).

For residents, who are by definition still calibrating their own judgment, the implication is that AI should be introduced in ways that preserve productive struggle. Strategies include requiring a documented independent assessment before an AI suggestion is revealed, deliberately inserting cases where the AI is wrong, and reviewing discordance between trainee and model in teaching conferences.

6.3 Competency development

Competency frameworks for graduate medical education should now explicitly include AI-related capabilities: understanding what a model was trained to do, recognising its failure modes, interpreting uncertainty, documenting AI use in the record, communicating AI-informed decisions to patients, and escalating when output conflicts with clinical judgment (Wartman & Combs, 2018). Assessment should evaluate the trainee's ability to detect and override erroneous AI advice, not merely to follow correct advice. Programmes should also use AI as an educational instrument, for example to generate simulated cases and formative feedback, while monitoring its accuracy. Faculty development is a prerequisite, since supervisors cannot teach critical appraisal of tools they do not themselves understand.

7. Emergency Medical Services: Prehospital Decision Support, Emergency Triage, and Destination Selection

7.1 Dispatch and recognition of time-critical conditions

EMS decisions are made under uncertainty, with limited information, little time, and high stakes. AI has therefore been explored at the earliest point in the chain, the emergency call. Machine learning can analyse speech in calls to help dispatchers recognise out-of-hospital cardiac arrest, a condition in which early recognition drives early resuscitation. In a randomised clinical trial of more than 5,000 emergency calls, however, supplying dispatchers with an ML-based alert did not significantly improve their recognition of cardiac arrest relative to usual practice (Blomberg et al., 2021). This negative trial is instructive. A model that performs well retrospectively may not change behaviour if dispatchers are already skilled, if alerts arrive too late in the call, or if trust is low. It also demonstrates the value of prospective randomised evaluation, which in this field remains uncommon (Nagendran et al., 2020).

7.2 Prehospital risk prediction and triage

Outside the call centre, AI is being used to predict patient acuity from limited prehospital variables. Kang et al. (2020) developed ML models using data available in the prehospital phase to predict the need for critical care, illustrating the feasibility of risk estimation using vital signs and demographic information, although their data were retrospective and the work awaits prospective implementation. A scoping review of AI in emergency medicine similarly describes a broad range of promising but largely pre-implementation applications (Fernandes et al., 2020).

Evidence from emergency department (ED) triage offers a useful proxy for the prehospital setting. Raita et al. (2019) showed that ML models using data available at triage predicted critical care and hospitalisation more accurately than a conventional reference model based on the Emergency Severity Index. Levin et al. (2018) reported that an ML-based electronic triage tool differentiated patients with respect to clinical outcomes better than the Emergency Severity Index, suggesting potential to reduce both under-triage and over-triage. These studies relied on retrospective data, and both warn that models must be validated locally and monitored over time. Translation to EMS requires additional attention to data that are sparse, noisy, and frequently entered under duress, and to the fact that paramedic impressions are rarely captured in structured form.

7.3 Destination selection

Destination selection, namely directing a patient to a trauma centre, stroke centre, percutaneous coronary intervention facility, or local emergency department, is a high-leverage decision at the interface between EMS and hospital care. In principle, AI could combine prehospital physiological data, mechanism of injury, geography, and real-time hospital capacity to recommend the most appropriate destination. Realising this vision depends on interoperable data flows between ambulance services and receiving hospitals, which remain limited in many systems. It also requires explicit management of the trade-off between under-triage, which risks harm to patients taken to inadequate facilities, and over-triage, which overloads specialist centres. As with all prediction tools, models should be assessed for equitable performance across age, sex, ethnicity, and geography, since errors in prehospital triage compound downstream (Obermeyer et al., 2019; Sutton et al., 2020). Until prospective evidence is available, AI-based destination advice should remain advisory, with paramedics retaining authority and clear protocols for override.

8. Diagnostic Radiology: AI-Assisted Image Interpretation, Prioritisation, and Emergency Reporting

8.1 Interpretation performance

Radiology is the specialty in which AI is most mature, owing to the digital, standardised nature of imaging data (Hosny et al., 2018). A systematic review and meta-analysis found that the diagnostic performance of deep learning in medical imaging was equivalent to that of healthcare professionals, but only a small minority of studies compared the two on the same sample, and reporting standards were poor, limiting confidence in the pooled estimates (Liu et al., 2019). The most persuasive evidence comes from breast screening. McKinney et al. (2020) reported that an AI system, evaluated on UK and US mammography datasets, reduced both false positives and false negatives relative to radiologists, with absolute reductions in false positives of 5.7% (US) and 1.2% (UK) and in false negatives of 9.4% (US) and 2.7% (UK). In an interim safety analysis of the randomised MASAI trial, AI-supported screen reading achieved a cancer detection rate of 6.1 per 1,000 compared with 5.1 per 1,000 with standard double reading, while reducing screen-reading workload by 44.3% (Lång et al., 2023). These findings suggest that AI can both maintain safety and reduce workforce demand, an important consideration given global radiologist shortages.

8.2 Prioritisation and emergency reporting

In emergency care, the principal value of AI may lie less in replacing the radiologist than in reordering the worklist so that critical studies are read first. Chilamkurthy et al. (2018) developed deep learning algorithms to detect critical findings on head computed tomography, including intracranial haemorrhage, and demonstrated high discrimination on an independent dataset. Such algorithms can flag suspected haemorrhage, large-vessel occlusion, pneumothorax, or pulmonary embolism for immediate review, shortening the time to report in the situations where minutes matter. The benefit is conditional on the operating point, however. A highly sensitive triage tool that generates many false flags may displace genuinely urgent studies and fatigue reporting staff, so performance metrics should include time-to-report and clinical outcomes rather than accuracy alone.

8.3 Augmentation and generalisability

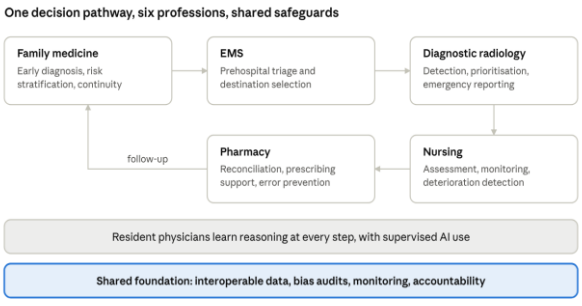
AI can also augment interpretation. In a multireader, multicase study, a comprehensive deep learning model for chest radiographs significantly improved radiologists' accuracy across a broad range of findings (Seah et al., 2021). The implication is that AI is most useful as a concurrent reader that draws attention to subtle findings, provided that clinicians remain critical of its output. Challenges nonetheless persist. Models are sensitive to dataset shift arising from different scanners, protocols, and populations (Kelly et al., 2019), many approved imaging devices have limited prospective evidence (Muehlematter et al., 2021), and reporting of evaluations remains inconsistent despite published guidelines for trials of AI interventions (Liu et al., 2020). Radiology departments should therefore treat each deployed algorithm as a clinical service requiring local validation, ongoing performance monitoring, and a defined pathway for discrepant results.

9. Toward an Integrated Cross-Professional Framework

The preceding sections show that each profession encounters AI at a different point in the patient journey, with different data, time pressures, and risks. Yet the patient's trajectory links them. An early signal in primary care, such as a high risk score or an abnormal screening result, should inform nursing surveillance if the patient is admitted, pharmacist review of medications, and the triage assumptions of EMS crews who may later attend the same patient. Conversely, a deterioration alert, a medication discrepancy, or a radiology flag generated in hospital should flow back to the family physician responsible for continuity. Isolated deployments cannot deliver this integration, and they risk generating redundant or contradictory signals that clinicians must reconcile unaided (Sutton et al., 2020).

Figure 1 presents an integrated framework in which the six professions are treated as nodes of a single decision pathway resting on shared foundations. Three principles follow. First, data must be interoperable and attributable, so that predictions can be traced to their inputs across settings. Second, every AI output should be positioned as input to a human decision, with explicit accountability for the final action. Third, education must be embedded at each node, since the safe use of AI depends on the judgment of the people who use it, which is developed in training and sustained through feedback (Wartman & Combs, 2018).

Figure 1. Integrated framework linking family medicine, EMS, diagnostic radiology, nursing, pharmacy, and resident training through shared safeguards.



integrated pathway · 6 professions, 1 foundation

Table 2 translates the framework into practical safeguards, evaluation metrics, and accountable roles for each domain. The metrics deliberately extend

beyond model accuracy. Reviews of AI evaluations have repeatedly shown that discrimination statistics alone fail to predict clinical benefit, and that outcomes such as time to treatment, alert burden, and equity of performance should be measured prospectively (Kelly et al., 2019; Nagendran et al., 2020).

Table 2. Recommended safeguards, evaluation metrics, and accountable roles for AI-supported decision-making

Domain	Priority safeguard	Suggested evaluation metrics	Accountable role
Family medicine	Subgroup audit of risk models; clinician review of all generated patient text	Calibration by subgroup; diagnostic yield; time to diagnosis; patient-reported continuity	Practice lead with clinical informatics
Nursing	Tiered alert thresholds; explicit nurse-concern override	Alert positive predictive value; alert burden per shift; time to escalation; ICU transfer and mortality	Nurse manager with rapid-response team
Pharmacy	Pharmacist verification of AI-prepared reconciliation; monitored alert suppression	Prevented adverse drug events; intervention acceptance; override rates; audit of missed errors	Medication safety officer
Resident training	Independent-first reasoning; deliberate inclusion of AI-error cases	Diagnostic accuracy with and without AI; detection of erroneous advice; competency milestones	Programme director
EMS	Advisory-only destination support with a defined override protocol	Under- and over-triage rates; time to definitive care; dispatcher recognition; equity across groups	EMS medical director
Diagnostic radiology	Local validation; worklist monitoring; discrepancy review	Local sensitivity and specificity; time to report for critical findings; recall rate; reader agreement	Radiology lead with AI governance committee

10. Cross-Cutting Risks

10.1 Bias and equity

Bias is the most consequential risk across all six domains, because it can be invisible to the clinician who relies on the tool. The cost-based risk algorithm described earlier demonstrates how a proxy label can systematically under-identify the need of disadvantaged patients (Obermeyer et al., 2019). The same dynamic can arise in triage models trained on historical decisions, in deterioration scores trained on data in which some groups are monitored less intensively, and in imaging models trained on unrepresentative populations. Mitigation requires diverse and well-documented training data, subgroup performance reporting, and post-deployment surveillance for differential errors. Ethical analyses emphasise that fairness cannot be treated as an optional extra, since deploying a biased system can entrench inequity at scale (Char et al., 2018).

10.2 Automation bias, deskilling, and trust calibration

The second risk is behavioural. Clinicians may over-trust a confident output, or conversely ignore a useful one after a series of false alarms. Evidence that inaccurate advice degrades clinician performance (Gaube et al., 2021; Tschandl et al., 2020) shows that the safety of decision support depends on the human-AI interaction as much as on the model. Figure 2 depicts the decision loop and the principal risk and safeguard at each step. The central design intention is a clinician check that precedes, or is independent of, the AI suggestion where feasible, and that is supported by explanations, uncertainty indicators, and an easy mechanism to report errors.

Figure 2. Human-AI decision loop showing the principal risk and corresponding safeguard at each step.



human-AI decision loop · 4 steps, 4 risks

10.3 Generalisability, transparency, and generative AI

The third risk is that tools fail silently when they leave the environment in which they were built. External validation of the sepsis model discussed earlier (Wong et al., 2021) and the limited prospective evidence for approved devices (Muehlematter et al., 2021) illustrate this problem. Generative models add further hazards: they can fabricate references, doses, or findings with great fluency, and they pose privacy risks when clinical text is entered into external systems (Haug & Drazen, 2023; Thirunavukarasu et al., 2023). Institutions should specify which generative tools are approved, what data may be entered, and which outputs require verification against source records.

11. Governance, Regulation, and Implementation

Successful decision support depends on socio-technical factors, including workflow fit, clinician engagement, and ongoing maintenance, as much as on algorithmic accuracy (Sutton et al., 2020). A practical governance model treats each AI tool as a clinical intervention with a lifecycle (Figure 3). Before deployment, the tool is validated on local data, including subgroup analyses. A pilot phase, ideally running silently in the background before outputs are shown to clinicians, tests workflow effects. After deployment, the tool is monitored for drift, alert burden, and harms, and decisions to retrain or withdraw are made against pre-specified criteria. Across the lifecycle, a named owner is accountable for performance and for transparent reporting of incidents.

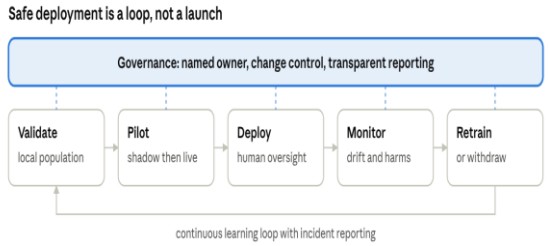


Figure 3. Lifecycle governance model for AI-supported clinical decision tools.

Regulatory frameworks continue to evolve. Analyses of approvals indicate that most AI-enabled devices are authorised on the basis of retrospective or limited evidence, so regulatory clearance should not be mistaken for proof of clinical benefit (Muehlematter et al., 2021). Health systems can complement regulation by requiring that trials and evaluations follow reporting guidance for AI interventions (Liu et al., 2020), by contributing to registries of adverse events, and by sharing local validation results with peers. Finally, the nursing, pharmacy, EMS, and trainee workforces should be represented on the committees that select and oversee these tools, since they bear the consequences of design choices (Ronquillo et al., 2021).

12. Research Priorities

Several gaps limit the field. First, outcome-focused prospective trials are scarce in primary care, nursing, pharmacy, and EMS relative to radiology, and pragmatic or cluster-randomised designs would help to establish whether AI changes patient outcomes rather than process measures (Kueper et al., 2020; Muralitharan et al., 2021). Second, research on human-AI interaction is needed to determine how explanations, uncertainty displays, and sequencing affect reasoning, particularly in trainees (Goh et al., 2024). Third, interoperable pathway-level studies, linking primary care, EMS, hospital, and radiology data, are needed to test whether the integrated approach proposed here outperforms isolated deployments. Fourth, equity-focused evaluations should be standard, with reporting by age, sex, ethnicity, and socioeconomic status. Fifth, longitudinal studies of competency, including the effect of AI exposure on skill acquisition and retention, are required before curricula can be considered evidence-based.

13. Limitations

This review has limitations. It is a narrative synthesis with purposive rather than exhaustive selection of sources, so relevant studies may have been omitted and the balance of evidence may be imperfectly represented. The field changes quickly, particularly for generative AI, and findings from 2023 to 2026 may be superseded. Evidence is uneven across domains, with the strongest data in imaging and sepsis prediction and the weakest in family medicine, pharmacy, and prehospital care. Most cited studies were conducted in high-income settings, which limits generalisability to low-resource health systems. The integrated framework is conceptual and has not been empirically tested.

14. Conclusion

AI-driven decision support has demonstrated credible value in several well-defined tasks, including autonomous screening in primary care, early warning for sepsis, prioritisation of imaging, and mammography reading, but evidence that it improves outcomes across the full patient pathway remains incomplete. Negative and mixed results, such as the failed external validation of a widely used sepsis model and the absence of benefit in a dispatcher trial, are as informative as successes because they identify the conditions under which AI fails. The central lesson is that AI should be implemented as a sociotechnical intervention, with local validation, human oversight, equitable performance, and education across professions, rather than as a stand-alone technology. If health systems adopt an integrated, governed approach and invest in prospective evaluation, AI can support family physicians, nurses, pharmacists, residents, paramedics, and radiologists in making safer and more timely decisions for the patients they share.

References

1. Abràmoff, M. D., Lavin, P. T., Birch, M., Shah, N., & Folk, J. C. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ digital medicine*, *1*(1), 39.
2. Adams, R., Henry, K. E., Sridharan, A., Soleimani, H., Zhan, A., Rawat, N., ... & Saria, S. (2022). Prospective, multi-site study of patient outcomes after implementation of the TREWS machine

learning-based early warning system for sepsis. *Nature medicine*, *28*(7), 1455-1460.
3. Ancker, J. S., Edwards, A., Nosal, S., Hauser, D., Mauer, E., Kaushal, R., & With the HITEC Investigators. (2017). Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. *BMC medical informatics and decision making*, *17*(1), 36.
4. Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelley, J. B., ... & Smith, D. M. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA internal medicine*, *183*(6), 589-596.
5. Bates, D. W., Levine, D., Syrowatka, A., Kuznetsova, M., Craig, K. J. T., Rui, A., ... & Rhee, K. (2021). The potential of artificial intelligence to improve patient safety: a scoping review. *NPJ digital medicine*, *4*(1), 54.
6. Blomberg, S. N., Christensen, H. C., Lippert, F., Ersbøll, A. K., Torp-Petersen, C., Sayre, M. R., ... & Folke, F. (2021). Effect of machine learning on dispatcher recognition of out-of-hospital cardiac arrest during calls to emergency medical services: a randomized clinical trial. *JAMA network open*, *4*(1), e2032320.
7. Cabitza, F., Rasoini, R., & Gensini, G. F. (2017). Unintended consequences of machine learning in medicine. *Jama*, *318*(6), 517-518.
8. Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care—addressing ethical challenges. *The New England journal of medicine*, *378*(11), 981.
9. Chilamkurthy, S., Ghosh, R., Tanamala, S., Biviji, M., Campeau, N. G., Venugopal, V. K., ... & Warier, P. (2018). Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *The Lancet*, *392*(10162), 2388-2396.
10. Corny, J., Rajkumar, A., Martin, O., Dode, X., Lajonchère, J. P., Billuart, O., ... & Buronfosse, A. (2020). A machine learning–based clinical decision support system to identify prescriptions with a high risk of medication error. *Journal of the American Medical Informatics Association*, *27*(11), 1688-1694.
11. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... & Dean, J. (2019). A guide to deep learning in healthcare. *Nature medicine*, *25*(1), 24-29.
12. Fernandes, M., Vieira, S. M., Leite, F., Palos, C., Finkelstein, S., & Sousa, J. M. (2020). Clinical decision support systems for triage in the emergency department using intelligent systems: a review. *Artificial intelligence in medicine*, *102*, 101762.
13. Fleuren, L. M., Klausch, T. L., Zwager, C. L., Schoonmade, L. J., Guo, T., Roggeveen, L. F., ... & Elbers, P. W. (2020). Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive care medicine*, *46*(3), 383-400.
14. Gaube, S., Suresh, H., Raue, M., Merritt, A., Berkowitz, S. J., Lermer, E., ... & Ghassemi, M. (2021). Do as AI say: susceptibility in deployment of clinical decision-aids. *NPJ digital medicine*, *4*(1), 31.
15. Goh, E., Gallo, R., Hom, J., Strong, E., Weng, Y., Kerman, H., ... & Chen, J. H. (2024). Large language model influence on diagnostic reasoning: a randomized clinical trial. *JAMA network open*, *7*(10), e2440969.
16. Haug, C. J., & Drazen, J. M. (2023). Artificial intelligence and machine learning in clinical medicine, 2023. *New England Journal of Medicine*, *388*(13), 1201-1208.
17. Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. W. L. (2018). Artificial intelligence in radiology. *Nature Reviews Cancer, 18*(8), 500–510.
18. Kang, D. Y., Cho, K. J., Kwon, O., Kwon, J. M., Jeon, K. H., Park, H., ... & Oh, B. H. (2020). Artificial intelligence algorithm to predict the need for critical care in prehospital emergency medical services. *Scandinavian journal of trauma, resuscitation and emergency medicine*, *28*(1), 17.
19. Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC medicine*, *17*(1), 195.
20. Kueper, J. K., Terry, A. L., Zwarenstein, M., & Lizotte, D. J. (2020). Artificial intelligence and primary care research: a scoping review. *The annals of family medicine*, *18*(3), 250-258.
21. Lång, K., Josefsson, V., Larsson, A. M., Larsson, S., Högberg, C., Sartor, H., ... & Rosso, A. (2023). Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. *The Lancet Oncology*, *24*(8), 936-944.
22. Levin, S., Toerper, M., Hamrock, E., Hinson, J. S., Barnes, S., Gardner, H., ... & Kelen, G. (2018). Machine-learning-based electronic triage more accurately

differentiates patients with respect to clinical outcomes compared with the emergency severity index. *Annals of emergency medicine*, *71*(5), 565-574.

23. Liu, X., Faes, L., Kale, A. U., Wagner, S. K., Fu, D. J., Bruynseels, A., ... & Denniston, A. K. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The lancet digital health*, *1*(6), e271-e297.
24. Liu, X., Rivera, S. C., Moher, D., Calvert, M. J., Denniston, A. K., Ashrafian, H., ... & Yau, C. (2020). Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *The Lancet Digital Health*, *2*(10), e537-e548.
25. McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., ... & Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, *577*(7788), 89-94.
26. Muehlematter, U. J., Daniore, P., & Vokinger, K. N. (2021). Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. *The Lancet Digital Health*, *3*(3), e195-e203.
27. Muralitharan, S., Nelson, W., Di, S., McGillion, M., Devereaux, P. J., Barr, N. G., & Petch, J. (2021). Machine learning–based early warning systems for clinical deterioration: systematic scoping review. *Journal of medical Internet research*, *23*(2), e25187.
28. Nagendran, M., Chen, Y., Lovejoy, C. A., Gordon, A. C., Komorowski, M., Harvey, H., ... & Maruthappu, M. (2020). Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *bmj*, *368*.
29. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447-453.
30. Raita, Y., Goto, T., Faridi, M. K., Brown, D. F., Camargo Jr, C. A., & Hasegawa, K. (2019). Emergency department triage prediction of clinical outcomes using machine learning models. *Critical care*, *23*(1), 64.
31. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, *380*(14), 1347-1358.
32. Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature medicine*, *28*(1), 31-38.
33. Ronquillo, C. E., Peltonen, L. M., Pruinelli, L., Chu, C. H., Bakken, S., Beduschi, A., ... & Topaz, M. (2021). Artificial intelligence in nursing: Priorities and opportunities from an international invitational think-tank of the Nursing and Artificial Intelligence Leadership Collaborative. *Journal of advanced nursing*, *77*(9), 3707-3717.
34. Seah, J. C., Tang, C. H., Buchlak, Q. D., Holt, X. G., Wardman, J. B., Aimoldin, A., ... & Jones, C. M. (2021). Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. *The Lancet Digital Health*, *3*(8), e496-e506.
35. Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., ... & Natarajan, V. (2023). Large language models encode clinical knowledge. *Nature*, *620*(7972), 172-180.
36. Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ digital medicine*, *3*(1), 17.
37. Syrowatka, A., Song, W., Amato, M. G., Foer, D., Edrees, H., Kuznetsova, M., ... & Bates, D. W. (2022). Key use cases for artificial intelligence to reduce the frequency of adverse drug events: a scoping review. *The Lancet Digital Health*, *4*(2), e137-e148.
38. Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. (2023). Large language models in medicine. *Nature medicine*, *29*(8), 1930-1940.
39. Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine*, *25*(1), 44-56.
40. Tschandl, P., Rinner, C., Apalla, Z., Argenziano, G., Codella, N., Halpern, A., ... & Kittler, H. (2020). Human–computer collaboration for skin cancer recognition. *Nature medicine*, *26*(8), 1229-1234.
41. Wartman, S. A., & Combs, C. D. (2018). Medical education must move from the information age to the age of artificial intelligence. *Academic Medicine*, *93*(8), 1107-1109.

Wong, A., Otles, E., Donnelly, J. P., Krumm, A., McCullough, J., DeTroyer-Cooley, O., ... & Singh, K. (2021). External validation of a widely implemented proprietary sepsis prediction

model in hospitalized patients. *JAMA internal medicine*, *181*(8), 1065-1070.