

When Does AI-Assisted Medication Review Add Clinical Value? An Evidence-Mapping Review of Prescription Screening, Automation Error, Verification Workload, Pharmacist Oversight, and Patient-Level Consequences, 2017–2026

Ahmed Ben Ali^{1*}, Samira Gharbi¹, Nabil Jebali²

¹ Department of AI-Assisted Medication Review, Faculty of Pharmacy, University of Tunis, Tunis, Tunisia.

² Department of Automation Error and Pharmacist Oversight, Faculty of Pharmacy, University of Sfax, Sfax, Tunisia.

*E-mail ✉ ahmed.benali@fpharm.u-tunis.tn

Received: 05 April 2026; Revised: 24 June 2026; Accepted: 29 June 2026

ABSTRACT

Artificial intelligence is increasingly used to prioritize prescriptions, suppress low-value alerts, identify medication-error risk, and support pharmacist verification. However, technical detection performance is frequently discussed alongside workflow or patient outcomes as though these endpoints demonstrate equivalent clinical value. This evidence-mapping review examines where AI-assisted medication review has progressed beyond technical screening and where evidence remains limited across automation error, verification workload, pharmacist oversight, and patient-level consequences. Peer-reviewed evidence published during 2017–2026 was mapped by medication-review task, care setting, data input, AI approach, validation design, outcome class, automation-failure characteristics, verification burden, pharmacist role, and proximity of outcomes to patients. Methodological and reporting standards were used to preserve distinctions between model performance, early clinical evaluation, workflow effects, and patient consequences. Evidence is most developed for bounded detection and prioritization tasks, including prescription-risk identification, medication-alert targeting, atypical-order detection, and selection of patients for pharmacist review. Human–AI studies indicate that uncertainty presentation and erroneous algorithmic advice can alter pharmacist decisions and verification time. Patient-level evaluation is less mature and context dependent. The review therefore proposes that clinical value should be judged across separable layers: detection, verification burden, pharmacist actionability, and patient consequence. AI-assisted medication review should not be judged by discrimination performance alone. Evidence becomes more clinically informative when model outputs are evaluated within human verification, pharmacist action, organizational workflow, and patient-level outcome pathways, with explicit attention to error recovery and transferability.

Keywords: Artificial intelligence, Medication review, Clinical pharmacy, Medication safety, Machine learning, Pharmacist oversight

How to Cite This Article: Ben Ali A, Gharbi S, Jebali N. When Does AI-Assisted Medication Review Add Clinical Value? An Evidence-Mapping Review of Prescription Screening, Automation Error, Verification Workload, Pharmacist Oversight, and Patient-Level Consequences, 2017–2026. *Ann Pharm Pract Pharmacother*. 2026;6(1):189-99. <https://doi.org/10.51847/NrVBaxjhib>

Introduction

Artificial intelligence has entered medication-review workflows through alert optimization, prescription screening, risk prioritization, and broader pharmacy applications. Recent reviews nevertheless describe a heterogeneous evidence base in which technical model performance, implementation experience, and clinical outcomes are unevenly represented [1–3]. This heterogeneity makes the question “does AI work?” too coarse for medication safety. A system may accurately identify unusual orders yet produce little clinical value if pharmacists must verify excessive false positives, if unsafe recommendations are difficult to recognize, or if predicted risk never changes treatment.

This review therefore treats clinical value as a layered rather than unitary construct. The proposed distinction separates four questions: whether an AI system detects or prioritizes relevant medication problems; whether its output changes verification workload; whether pharmacists can appropriately accept, reject, or recover from its recommendations; and whether downstream patient consequences are demonstrated. Human–AI experiments show that erroneous advice and the presentation of uncertainty can alter pharmacist decision-making [4]. More recent prospective evidence also shows that medication-safety outcomes can be evaluated downstream of AI-supported decision support rather than inferred from model discrimination alone [5]. These layers are analytically connected but should not be collapsed.

Evidence-mapping questions and classification scheme

The evidence map was designed to determine where AI-assisted medication review has been evaluated and how close each evaluation comes to clinically consequential action. Existing clinical-pharmacy AI literature supports separating task type from development and validation maturity because tools performing apparently similar “medication review” functions may automate different decisions and undergo substantially different evaluation pathways [6].

The classification scheme therefore distinguished medication-review task, setting, data source, AI approach, comparator, validation design, automation-error characteristics, verification workload, pharmacist oversight, and outcome level. This review additionally proposes a four-layer clinical-value distinction: detection performance, verification burden, pharmacist actionability, and patient consequence. These layers are not treated as mandatory chronological stages. A technically strong model may remain weakly actionable, while an apparently useful workflow intervention may depend on local staffing, interface design, or other co-interventions. Experimental pharmacist evidence demonstrates that human performance changes when AI advice is incorrect or its uncertainty is communicated differently [4]. Patient-level evaluation represents a separate evidential level rather than an assumed extension of such task performance [5].

Materials and Methods

The review was conducted as an evidence map rather than as a pooled effectiveness synthesis because the included literature encompasses prediction-model development, human–AI experiments, system evaluations, implementation studies, reviews, and methodological guidance with non-equivalent endpoints. Reporting was structured using PRISMA 2020 principles for transparent review identification, screening, and inclusion [7]. Scoping-review reporting principles were additionally applied because the objective was to characterize heterogeneous evidence domains, evaluation maturity, and gaps rather than estimate a single treatment effect [8]. Methodological conduct followed contemporary guidance for scoping reviews, including explicit formulation of the review question, eligibility logic, evidence charting, and structured synthesis [9]. The unit of mapping was the eligible article, while study-level coding focused on the evaluated AI task and its evidential contribution. Technical performance metrics were not interpreted as interchangeable across different prediction targets, reference standards, or care contexts.

No meta-analysis was planned because the review question concerned the distribution and maturity of evidence rather than a common quantitative effect. Evidence density was likewise not treated as evidence of effectiveness. Where studies reported implementation, workload, or patient outcomes, these were mapped separately from model-development results so that progressively more patient-proximal evidence could be identified without implying a validated causal sequence.

Information sources and eligibility criteria

Search and verification procedures were documented to distinguish actual retrieval from later citation checking and targeted metadata verification, consistent with PRISMA-S expectations for reproducible literature-search reporting [10]. Searches covered literature published from 2017 through 2026 and focused on AI or machine-learning applications relevant to prescription screening, medication-error detection, alert optimization, pharmacist verification, medication-review prioritization, oversight, and downstream medication-safety consequences.

Eligible sources were peer-reviewed journal articles with a verifiable DOI and direct relevance to at least one prespecified evidence-map domain or a necessary methodological foundation. Prediction-model reports were required to provide sufficient information to classify data sources, outcome definition, model-development or

validation design, and intended use; these reporting elements are central to TRIPOD+AI [11]. Model performance was also interpreted separately from risk of bias and applicability, consistent with PROBAST+AI [12].

Studies were excluded when they were outside the date window, duplicated another article version, lacked adequate source verification, or were too indirectly related to support a specific medication-review claim. Differences in population, care setting, intervention target, and outcome definition were preserved rather than harmonized by assumption.

Study selection and screening

Eligibility was assessed after DOI- and title-level deduplication. Machine-assisted review methods can increase screening efficiency, but human verification remains necessary when eligibility decisions affect the evidence base [13]. Accordingly, no record was treated as eligible solely because of automated ranking or semantic similarity. Clinical AI reports were additionally examined for information needed to interpret human interaction, workflow integration, safety, and failure handling. DECIDE-AI identifies these dimensions as important during early clinical evaluation of AI-driven decision support [14]. Pharmacy-specific reporting considerations were also applied using the MedinAI framework, particularly for defining input data, model purpose, validation, pharmacist interaction, and implementation context [15]. Imperfect reporting did not automatically exclude an otherwise relevant study, but missing information constrained what could subsequently be coded or inferred.

Figure 1 shows the source-selection record. Thirty-nine unique records underwent screening; 14 were excluded and 25 were included. Earlier retrieval, deduplication, and full-text transition totals were unavailable from the preserved search records.

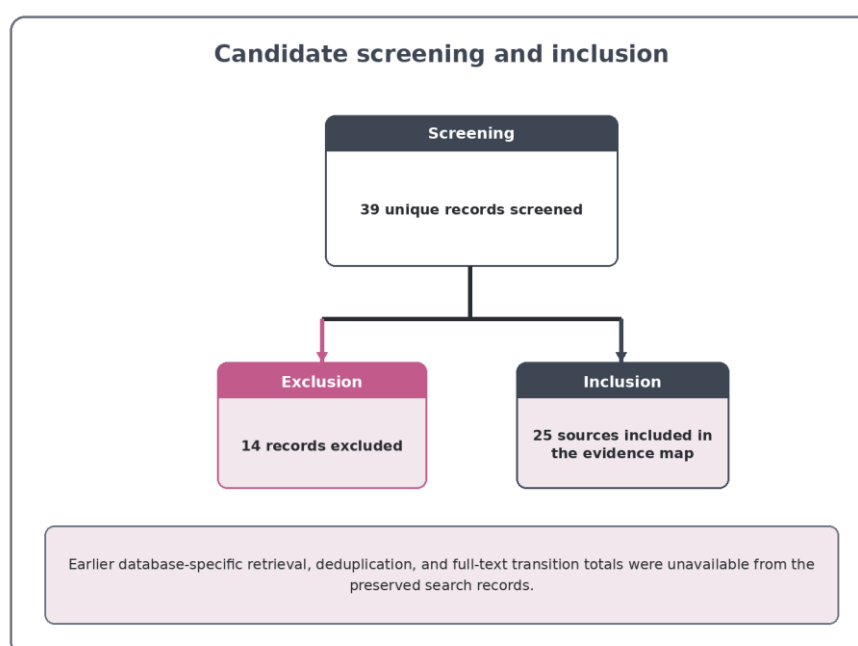


Figure 1. Selection of evidence for the AI-assisted medication-review evidence map

Data extraction and evidence-map coding

Data extraction preserved distinctions between model, clinician, patient, and health-system levels. Coding fields included medication-review task, setting, data input, AI approach, comparator, validation design, error type, workload outcome, pharmacist oversight, patient-level outcome, and evidence maturity. Governance and postdeployment considerations were retained as separate variables because responsible AI-enabled clinical decision support depends on validation, monitoring, transparency, and clinician preparation beyond predeployment model performance [16].

Outcome coding also separated workflow proxies from clinical consequences. For example, machine-learning models developed from pharmacy inquiry databases can predict dose-related inquiries and potentially prioritize pharmacist attention, but the predicted endpoint remains an inquiry rather than medication harm prevented [17].

Similar distinctions were applied to alert acceptance, atypical orders, pharmaceutical interventions, medication discrepancies, adverse drug events, and utilization outcomes.

The review therefore used a proposed endpoint hierarchy without assigning numerical weights: technical discrimination; task-level detection or prioritization; human verification or action; workflow consequence; and patient-level consequence. Movement toward a more patient-proximal endpoint was treated as increased clinical interpretability, not automatically as stronger causal evidence. Study design, risk of bias, comparator, external validation, and implementation context remained independent dimensions.

Results and Discussion

The mapped literature shows several distinct forms of AI-assisted medication review rather than a single technological intervention. Hospital studies have developed machine-learning systems to identify prescriptions at elevated risk of medication error and thereby prioritize pharmacist review [18, 19]. The same broad literature includes alert optimization, medication-reconciliation targeting, prediction of dose-related inquiries, and human–AI decision support. Because the prediction targets differ, direct comparison of reported discrimination metrics would be misleading.

Evaluation maturity also varied. Some studies remained at retrospective development or internal validation; others examined pharmacist responses to system output, simulated operational prioritization, or prospective clinical use. This spread supports treating “AI-assisted medication review” as an evidence field defined by multiple task–outcome combinations rather than by one effectiveness estimate.

A recurring distinction was whether the AI output represented the endpoint or merely initiated another clinical process. Flagging an unusual prescription, selecting a patient for medication reconciliation, or predicting a likely pharmacist intervention may have operational value, but each still requires downstream verification and action before patient benefit can plausibly occur. This separation became central to the subsequent task-level mapping.

Distribution of evidence across medication-review tasks

Medication-alert targeting represents one visible evidence cluster. Machine-learning-based clinical decision support has been evaluated for reducing low-value alert burden while identifying wrong-drug and look-alike/sound-alike medication risks [20]. Such outcomes capture system targeting and clinician-facing workload but should not automatically be interpreted as patient-level medication-safety benefit.

A second cluster concerns risk-based allocation of pharmacist attention. A recent hospital model used electronic health-record information to prioritize patients for admission medication reconciliation, illustrating how AI may reallocate a constrained pharmacy service toward cases predicted to contain medication errors [21]. This form of assistance differs analytically from autonomous screening because the model selects where human review should be concentrated rather than replacing that review.

The task inventory in **Table 1** makes these distinctions visible across data source, validation design, and endpoint class. It deliberately preserves unlike targets rather than combining them under a single measure of “accuracy.”

Table 1. Where AI enters medication review: task targets, validation routes, and outcome classes

Evaluated study/system	Medication-review task	Care setting	Data input	AI approach	Comparator	Validation design	Outcome class / external validation
Uncertainty-aware verification system	Medication verification	Simulated pharmacist task	Medication cases + AI advice	Uncertainty-aware AI	Black-box AI	Randomized crossover evaluation	Decision accuracy/time; no external clinical validation
AI-enhanced CDS network	Medication-error prevention	Multicenter hospitals	Clinical/order data	AI-enabled CDS	Usual decision support	Prospective multicenter randomized evaluation	Medication errors/ADEs; multisite clinical evaluation
Dose-inquiry predictor	Dose-review prioritization	Tertiary hospital	Pharmacy inquiry data base	Machine learning	Model comparators	Retrospective development/validation	Workflow proxy; no patient outcome
High-risk prescription system	Prescription prioritization	Hospital pharmacy	Prescription data	Machine learning	Existing review process/model comparators	Retrospective model evaluation	Medication-error risk; no external validation shown

Atypical-order detector	Order anomaly detection	Hospital pharmacy	Medication-order data	Machine learning	Pharmacist judgment	Prospective pharmacist evaluation	Perceived usefulness/actionability
Alert-targeting CDS	Alert optimization	Electronic prescribing	Medication/order alerts	Machine learning CDS	Existing alert logic	Clinical system evaluation	Alert burden/wrong-drug targeting
Reconciliation prioritizer	Patient selection for review	Hospital admission	EHR + reconciliation data	Machine learning	Existing targeting tool	Internal validation + simulation	Medication-error risk; no external-hospital validation

Note: The table includes directly evaluative medication-review evidence available at this point in the manuscript. Methodological and reporting standards are not listed as system evaluations. Technical, workflow, and patient-level endpoints are intentionally kept separate.

Prescription screening and detection performance

Prescription-screening systems commonly predict where pharmacist attention or intervention is likely to be needed, rather than establishing that harm has occurred. A recent hospital study used machine learning to predict pharmaceutical interventions during prescription review [22]. This endpoint is operationally meaningful because it can support prioritization, yet it remains dependent on local intervention practices, documentation, and clinical judgement. A predicted intervention should therefore not be interpreted as an adverse event prevented.

More clinically informative evaluations connect prediction with an implemented medication-management process. In a pharmacist-led surgical medicines optimization service, machine-learning support was incorporated into prescription review alongside pharmacist recommendations, with external validation and subsequent implementation outcomes reported [23]. The study is important because it places algorithmic output inside an actionable pharmacy service rather than evaluating discrimination in isolation. However, the combined intervention also limits attribution: changes in prescribing or patient outcomes cannot be assigned to the algorithm alone when pharmacist activity, workflow redesign, recommendation uptake, and other service components change concurrently.

Thus, detection performance becomes more clinically interpretable when the study identifies who acts on the output, what verification occurs, and which downstream endpoint is measured. It does not become equivalent to clinical effectiveness merely by being embedded in practice.

Automation error and failure characterization

Automation failure in medication review cannot be reduced to a single false-positive/false-negative distinction. Generative systems introduce additional risks because apparently coherent recommendations may omit clinically relevant context, respond inconsistently to prompt construction, or encourage unwarranted confidence. Evaluation of a retrieval-augmented large language model across multiple clinical specialties suggests that such systems can augment medication-safety reasoning under controlled conditions, while also retaining important limitations related to generalizability, bias, and prompt design [24]. Controlled performance therefore does not establish autonomous reliability in routine medication review.

Human-system failure is equally important. Medication alerts can lose value when excessive signaling, poor interaction design, or role-inappropriate information produces alert fatigue and disengagement [25]. In this review, automation failure is therefore treated as any system property that can distort detection, prioritization, interpretation, verification, or appropriate clinical response.

Table 2 distinguishes these failure modes according to whether pharmacists can detect and recover from them. The clinically important issue is not simply whether an algorithm makes an error, but whether that error remains visible before it affects medication action.

Table 2. Failure modes that determine whether AI medication-review errors remain recoverable

Failure mode	Typical source	Detectability	Potential consequence pathway	Verification needed	Pharmacist recovery role
False-positive flag	Low specificity or poorly calibrated threshold	Usually visible	Unnecessary review → interruption → alert burden	Clinical relevance check	Reject low-value flag; preserve attention
False-negative omission	Incomplete detection or out-of-distribution case	Often difficult	Relevant problem not surfaced → review opportunity missed	Independent medication review	Identify problem despite absent signal

Incorrect AI recommendation	Erroneous model reasoning or prediction	Variable	False assurance → inappropriate acceptance	Source and patient-context verification	Challenge recommendation and restore clinical judgement
Uncertainty hidden	Black-box presentation	Difficult	Confidence exceeds evidential support → automation bias	Explicit uncertainty appraisal	Discount unsupported certainty
Interaction/alert fatigue	Repetitive or role-mismatched signaling	Observable over time	Desensitization → dismissal of useful alerts	Workload and override review	Differentiate justified dismissal from unsafe habituation
Context-sensitive generative output	Prompt, retrieval, or contextual limitation	Variable	Inconsistent interpretation → unstable medication advice	Re-query plus independent verification	Reconcile output with patient-specific evidence

Note: Failure categories describe evidence-supported and evidence-informed distinctions across the mapped literature. They do not assign probabilities or imply that every AI system exhibits every failure mode.

Verification workload and human factors

AI assistance can reduce some search or prioritization work while creating new verification demands. In a randomized pharmacist experiment, incorrect AI advice affected medication-verification decisions, whereas communicating model uncertainty altered reaction time and reduced vulnerability to some erroneous recommendations [4]. The implication is that workload cannot be represented by alert count or processing speed alone. A faster decision may still require additional checking if pharmacists must establish whether the recommendation, its confidence, and its underlying context are trustworthy.

Automation may also redistribute rather than remove work. Prediction of dose-related pharmacy inquiries directs attention toward selected prescriptions [17], while patient-risk models can concentrate medication-reconciliation activity on patients more likely to contain medication discrepancies [21]. Such prioritization may lower the volume of cases receiving equal review but increase the complexity of cases reaching pharmacists.

Table 3 therefore treats verification burden as multidimensional. Time, cognitive checking, alert exposure, secondary confirmation, and susceptibility to automation bias can move in different directions. Claims of workload improvement should specify which dimension changed and what clinical work remained necessary.

Table 3. How AI assistance can reduce, increase, or redistribute medication-verification work

Verification outcome	Verification task	Workload measure	Comparator	Direction suggested by mapped evidence	Automation-bias concern	Context boundary
Decision time	Evaluate AI-supported medication case	Reaction time	Alternative AI presentation	May change with uncertainty display	Faster response may reflect premature acceptance	Simulated verification task
Secondary checking	Confirm AI recommendation	Additional clinical review	Independent judgement	May increase when output is uncertain	Failure to check erroneous advice	Depends on access to clinical context
Alert exposure	Triage medication warnings	Number/relevance of signals	Existing alert logic	Potential reduction in low-value alerts	Excessive filtering may hide useful warnings	Local alert configuration
Case prioritization	Select prescriptions/patients for review	Review volume distribution	Routine or existing targeting	Work may shift toward higher-risk cases	Nonselected cases remain vulnerable to missed risk	Local case mix and model calibration
Cognitive vigilance	Decide whether to accept automation	Attention and contradiction checking	Unassisted review	Cannot be assumed to decrease	Overreliance on confident output	Human–AI interface and professional expertise

Note: “Direction” indicates the qualitative workload implication supported by the mapped evidence; it is not a pooled effect estimate.

Pharmacist oversight and override

Pharmacist oversight appeared in several forms across the mapped evidence. Pharmacists may provide the reference judgement against which an atypical medication order is evaluated, determine whether a generated warning warrants intervention, verify AI-supported recommendations, or operate an AI-enabled service in collaboration with prescribers. Prospective pharmacist evaluation of atypical-order detection illustrates that model

output and perceived usefulness remain distinct outcomes [19]. In the surgical medicines optimization study, pharmacists retained an active intervention role within the combined AI-supported service [23].

Override should therefore not automatically be interpreted as failure of adherence to decision support. Rejecting an irrelevant or erroneous alert can represent successful human recovery from automation error. Conversely, repeated dismissal can become unsafe when fatigue or overfamiliarity causes clinically meaningful signals to be ignored. The evidential question is consequently why the pharmacist accepted, modified, or rejected the output, not simply whether override occurred.

The review proposes that pharmacist oversight be classified by decision authority: verification, interpretation, intervention, escalation, or service-level monitoring. The required level may differ according to medication risk, task complexity, model transparency, organizational responsibility, and availability of patient-specific information.

Patient-level consequences and clinical value

Patient-level evidence is qualitatively different from evidence that an algorithm correctly classifies prescriptions. A recent multicenter randomized evaluation reported reductions in medication errors and adverse drug events with AI-enhanced clinical decision support [5]. Such evidence moves evaluation closer to medication-safety consequences because the endpoint occurs downstream of algorithmic prediction and clinician response. Nevertheless, transfer to other health systems depends on implementation conditions, baseline safety infrastructure, patient case mix, and replication.

The pharmacist-led surgical optimization study similarly assessed outcomes beyond predictive performance, including medication-management and downstream patient or service consequences [23]. Because the intervention combined machine learning with pharmacist activity and workflow redesign, however, observed changes should be attributed to the integrated service rather than the algorithm alone.

Table 4 formalizes the proposed distinction between evidence levels. Increasing proximity to patients improves the clinical interpretability of an endpoint, but does not erase design limitations. A nonrandomized patient-outcome study may remain more vulnerable to confounding than a rigorous technical evaluation, even though its endpoint is more clinically proximal.

Table 4. What increasingly patient-proximal AI medication-review outcomes permit—and do not permit—inference about clinical value

Evidence level	Typical endpoint	Patient proximity	Supporting evaluation on design	Inference permitted	Inference not permitted
Technical performance	Discrimination, sensitivity, ranking	Distant	Retrospective development/internal validation	Model identifies its defined target under evaluated conditions	Patient benefit or workflow effectiveness
Detection/prioritization	Flagged prescription, selected patient, predicted intervention	Indirect	Retrospective or prospective task evaluation	AI may direct attention toward predefined cases	Medication harm was prevented
Human verification	Accuracy, response time, acceptance/rejection	Intermediate	Human–AI experiment or pharmacist evaluation	AI changes clinician interaction with medication-review information	Improved patient outcomes
Workflow/action	Pharmacist intervention, alert burden, reconciliation targeting	Closer	Clinical-system or implementation evaluation	AI is associated with altered review processes	Algorithm alone caused outcome change
Patient consequence	Medication error, ADE/ADR, utilization outcome	Direct	Prospective comparative or implementation study	Clinical consequences can be evaluated downstream of AI-supported care	Universal effectiveness or transferability

Abbreviations: ADE, adverse drug event; ADR, adverse drug reaction.

Where evidence is concentrated

Across recent reviews, AI medication-review research is concentrated around detection, prioritization, alert optimization, and related pharmacy tasks, with considerable heterogeneity in design and outcome maturity [1–3]. The empirical studies mapped here reinforce that pattern: algorithms identify atypical orders, estimate medication-

error risk, prioritize medication reconciliation, predict pharmacist interventions, or refine alerts. These applications produce outputs relatively close to the information-processing stage of medication review.

Human-in-the-loop evaluation is present but less developed as an evidential layer. Pharmacist studies examine perceived usefulness, verification decisions, uncertainty, and intervention processes, yet these outcomes remain context sensitive. Patient consequences are evaluated in fewer, more heterogeneous designs and often coexist with workflow or service changes that complicate attribution.

The concentration of evidence near technical and task-level endpoints should not be read as evidence that these endpoints are the most clinically important. It more plausibly indicates that they are easier to measure during model development. The evidence map consequently distinguishes research density from evidential sufficiency: many technically evaluated systems can coexist with substantial uncertainty about their clinical value.

Where the evidence map remains empty

The most consequential gaps occur between successful prediction and reliable transfer into routine care. External validation remains critical because model bias and applicability cannot be inferred from internal discrimination alone [12]. Postdeployment monitoring is another weakly represented dimension; recommendations for responsible AI-enabled clinical decision support emphasize continuing validation, governance, transparency, and clinician preparation after implementation [16].

Evidence is also thin where several requirements intersect: diverse patient populations, longitudinal verification burden, changing model performance, pharmacist accountability, and direct patient consequences. A technically successful system may behave differently after changes in formulary, prescribing practice, staffing, documentation, or patient case mix. Generative systems introduce additional temporal instability because model versions, retrieval sources, and prompting practices may change.

Figure 2 locates these gaps relative to the better-developed detection and prioritization evidence. It is intentionally qualitative: marker position reflects the mapped evaluation target, while no bubble size or numerical density is invented.

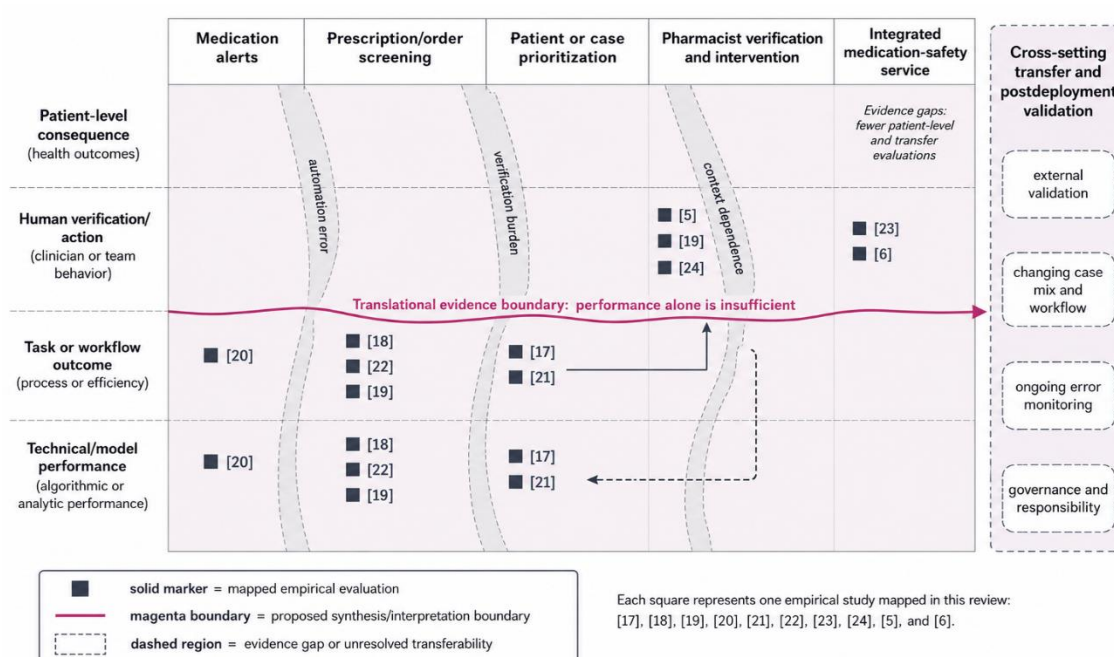


Figure 2. AI medication-review evidence is densest near detection and thins across verification, patient consequence, and transfer

The principal finding of this evidence map is not that AI-assisted medication review is either effective or ineffective. Rather, clinical value depends on where evidence is located along the path between model output and medication consequence. The mapped literature is comparatively mature in detection and prioritization, while

verification burden, pharmacist recovery from erroneous automation, sustained implementation, and patient consequences remain less uniformly evaluated.

The review therefore proposes a conditional clinical-value architecture. First, the model must identify or prioritize an actionable medication problem. Second, the output must not impose verification costs that overwhelm its informational value. Third, pharmacists or other clinicians must retain sufficient authority and contextual information to correct inappropriate automation. Finally, downstream consequences must be measured if patient benefit is claimed. Human–AI experimentation shows that erroneous advice and uncertainty presentation can materially alter pharmacist decisions [5]. Patient-outcome studies demonstrate that downstream evaluation is possible [6], while pharmacist-led implementation evidence illustrates how algorithmic effects can become inseparable from service redesign [23].

These conditions are dependencies, not a validated causal sequence. A system may perform well at one level and fail at another. Accordingly, model discrimination should be considered one component of an evidence hierarchy rather than the endpoint of clinical evaluation. Reporting quality, bias, applicability, human factors, and early clinical evaluation remain independently important when interpreting AI-supported decision systems [11, 12, 14].

Implications for ai-assisted medication review research

Future evaluations should specify what clinical action an AI output is expected to change, who retains responsibility for verification, how erroneous or uncertain recommendations are handled, and which downstream endpoint establishes value. Early clinical AI guidance supports explicit reporting of workflow integration, human interaction, safety, and failure handling [14]. Pharmacy-specific reporting guidance adds the need for transparent description of medication data, model purpose, validation, and pharmaceutical-service context [15]. Governance recommendations further emphasize monitoring and clinician preparation beyond initial deployment [16].

External validation should precede broad transfer claims. A model developed for one hospital's prescribing patterns, inquiry database, formulary, or pharmacist practices may not retain the same calibration or usefulness elsewhere. Prospective evaluations should therefore compare not only prediction performance but also verification effort, intervention appropriateness, automation-error recovery, and patient-proximal outcomes.

Research should additionally examine change over time. Model performance, clinical documentation, staffing, medication availability, and generative-AI behavior may drift after implementation. Longitudinal evaluation is therefore necessary to determine whether apparent efficiency persists or merely shifts work into less visible verification, monitoring, and recovery tasks.

Review Limitations

This evidence map is constrained by rapidly changing AI terminology, heterogeneous medication-review definitions, and uneven reporting across technical and clinical studies. Pharmacy applications range from conventional machine-learning prioritization to generative decision support, making a single performance metric inappropriate. Reporting frameworks can improve interpretability but cannot recover missing methodological detail [15].

Search-flow reporting is limited because early database-specific retrieval, deduplication, and full-text transition totals were unavailable from the preserved search records [7, 10]. The reported flow therefore begins with 39 unique records screened, of which 14 were excluded and 25 were included.

Finally, qualitative concentration in the evidence map should not be interpreted as a complete census of every AI medication-review implementation. Publication bias, indexing delays, local systems that remain unpublished, and rapid model turnover may alter apparent evidence density.

Conclusion

AI-assisted medication review has its clearest evidential footing in bounded detection, alerting, and prioritization tasks, but technical success alone does not establish clinical value. Medication-review AI becomes more clinically interpretable when evaluation also addresses verification burden, erroneous automation, pharmacist authority, actionability, and downstream patient consequences. These dimensions operate across model, clinician, patient, and health-system levels and may change with setting, workflow, case mix, and time.

The evidence map therefore supports a cautious translational principle: performance should be treated as the beginning of clinical evaluation rather than its conclusion. Patient-proximal outcomes strengthen interpretation, but study design, co-interventions, external validation, and transferability remain decisive. Future evidence should connect algorithmic output to human recovery, clinical action, sustained monitoring, and measurable medication consequences without assuming that success at one level guarantees success at another.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Graafsma J, Murphy RM, van de Garde EMW, Karapinar-Carkit F, Derijks HJ, Hoge RHL, et al. The use of artificial intelligence to optimize medication alerts generated by clinical decision support systems: A scoping review. *J Am Med Inform Assoc.* 2024;31(6):1411-22. doi:10.1093/jamia/ocae076
2. Johns E, Alkanj A, Beck M, Dal Mas L, Gourieux B, Sauleau EA, et al. Using machine learning or deep learning models in a hospital setting to detect inappropriate prescriptions: A systematic review. *Eur J Hosp Pharm.* 2024;31(4):289-94. doi:10.1136/ejhpharm-2023-003857
3. Hatzimanolis J, Riley B, El-Den S, Aslani P, Zhou J, Chaar BB. Applications of artificial intelligence in current pharmacy practice: A scoping review. *Res Social Adm Pharm.* 2025;21(3):134-41. doi:10.1016/j.sapharm.2024.12.007
4. Lester C, Rowell B, Zheng Y, Co Z, Marshall V, Kim JY, et al. Effect of uncertainty-aware AI models on pharmacists' reaction time and decision-making in a web-based mock medication verification task: Randomized controlled trial. *JMIR Med Inform.* 2025;13:e64902. doi:10.2196/64902
5. Shakarbaev N. AI-enhanced clinical decision support reduces medication errors and adverse drug events in a multicenter teaching hospital network: A prospective randomized controlled trial. *Int J Med Inform.* 2026;215:106457. doi:10.1016/j.ijmedinf.2026.106457
6. Ranchon F, Chanoine S, Lambert-Lacroix S, Bosson JL, Moreau-Gaudry A, Bedouch P. Development of artificial intelligence powered apps and tools for clinical pharmacy services: A systematic review. *Int J Med Inform.* 2023;172:104983. doi:10.1016/j.ijmedinf.2022.104983
7. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71. doi:10.1136/bmj.n71
8. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Ann Intern Med.* 2018;169(7):467-73. doi:10.7326/M18-0850
9. Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. *JBIM Evid Synth.* 2020;18(10):2119-26. doi:10.11124/JBIES-20-00167
10. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *Syst Rev.* 2021;10(1):39. doi:10.1186/s13643-020-01542-z
11. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ.* 2024;385:e078378. doi:10.1136/bmj-2023-078378
12. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, et al. PROBAST+AI: An updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ.* 2025;388:e082505. doi:10.1136/bmj-2024-082505
13. van de Schoot R, de Bruin J, Schram R, Zahedi P, de Boer J, Weijdemans F, et al. An open source machine learning framework for efficient and transparent systematic reviews. *Nat Mach Intell.* 2021;3(2):125-33. doi:10.1038/s42256-020-00287-7

- Ben Ali *et al.*, When Does AI-Assisted Medication Review Add Clinical Value? An Evidence-Mapping Review of Prescription Screening, Automation Error, Verification Workload, Pharmacist Oversight, and Patient-Level Consequences, 2017–2026
14. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat Med*. 2022;28(5):924-33. doi:10.1038/s41591-022-01772-9
 15. Bottacin WE, de Souza TT, Reis WCT, Melchioris AC. Guidelines for reporting artificial intelligence studies in medicines, pharmacotherapy, and pharmaceutical services: MedinAI development, validation and statement. *Int J Clin Pharm*. 2025;47(4):945-56. doi:10.1007/s11096-025-01905-3
 16. Labkoff S, Oladimeji B, Kannry J, Solomonides A, Leftwich R, Koski E, et al. Toward a responsible future: Recommendations for AI-enabled clinical decision support. *J Am Med Inform Assoc*. 2024;31(11):2730-9. doi:10.1093/jamia/ocae209
 17. Cho J, Lee AR, Koo D, Kim K, Jeong YM, Lee HY, et al. Development of machine-learning models using pharmacy inquiry database for predicting dose-related inquiries in a tertiary teaching hospital. *Int J Med Inform*. 2024;185:105398. doi:10.1016/j.ijmedinf.2024.105398
 18. Corny J, Rajkumar A, Martin O, Dode X, Lajonchère JP, Billuart O, et al. A machine learning-based clinical decision support system to identify prescriptions with a high risk of medication error. *J Am Med Inform Assoc*. 2020;27(11):1688-94. doi:10.1093/jamia/ocaa154
 19. Hogue SC, Chen F, Brassard G, Lebel D, Bussi eres JF, Durand A, et al. Pharmacists' perceptions of a machine learning model for the identification of atypical medication orders. *J Am Med Inform Assoc*. 2021;28(8):1712-8. doi:10.1093/jamia/ocab071
 20. Chen CY, Chen YL, Scholl J, Yang HC, Li YCJ. Ability of machine-learning based clinical decision support system to reduce alert fatigue, wrong-drug errors, and alert users about look alike, sound alike medication. *Comput Methods Programs Biomed*. 2024;243:107869. doi:10.1016/j.cmpb.2023.107869
 21. Abdo A, Gallay L, Vallecillo T, Clarenne J, Quillet P, Vuiblet V, et al. A machine learning-based clinical predictive tool to identify patients at high risk of medication errors. *Sci Rep*. 2024;14(1):32022. doi:10.1038/s41598-024-83631-w
 22. Johns E, Guendouz A, Dal Mas L, Beck M, Alkanj A, Gourieux B, et al. Using machine learning to predict pharmaceutical interventions during medication prescription review in a hospital setting. *Am J Health Syst Pharm*. 2025;82(22):1238-48. doi:10.1093/ajhp/zxaf089
 23. Li X, Yue X, Zhang L, Zheng X, Shang N. Pharmacist-led surgical medicines prescription optimization and prediction service improves patient outcomes - a machine learning based study. *Front Pharmacol*. 2025;16:1534552. doi:10.3389/fphar.2025.1534552
 24. Ong JCL, Jin L, Elangovan K, Lim GYS, Lim DYZ, Sng GGR, et al. Large language model as clinical decision support system augments medication safety in 16 clinical specialties. *Cell Rep Med*. 2025;6(10):102323. doi:10.1016/j.xcrm.2025.102323
 25. Hussain MI, Reynolds TL, Zheng K. Medication safety alert fatigue may be reduced via interaction design and clinical role tailoring: A systematic review. *J Am Med Inform Assoc*. 2019;26(10):1141-9. doi:10.1093/jamia/ocz095