Review Article

Artificial Intelligence as a Second Reader for High-Risk Prescriptions rather than a First Decision Maker

Sara Al-Fahd[1]*, Noura Al-Khalifa[1], Omar Al-Turki[2]

[1]Department of High-Risk Prescription Safety, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia. [2]Department of AI-Assisted Clinical Review, College of Pharmacy, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia.

Abstract

Artificial intelligence is increasingly positioned near medication decisions, yet the safest location for algorithmic assistance remains unresolved. Treating AI as the first reader may frame the case before a pharmacist has independently interpreted the prescription, while treating it as a decision maker may blur professional authority, encourage automation bias, and conceal uncertainty. This article proposes an Independent Human–AI Prescription Second-Reader Model for selected high-risk prescriptions. The model begins with a human first read, followed by a separately configured AI review activated through a locally validated risk-trigger gate. The AI may detect, prioritize, challenge, explain, or abstain, but it may not prescribe, authorize, independently release a prescription, or adjudicate unresolved conflict. Human and AI findings are reconciled through six proposed states: concordant clearance, concordant concern, AI-only concern, human-only concern, competing interpretations, and unresolved uncertainty or abstention. Disagreement initiates data verification, contextual recovery, independent professional reassessment, prescriber clarification, and proportionate escalation. The model also incorporates auditability, subgroup assessment, workload monitoring, drift surveillance, version control, and suspension or requalification after material change. Evaluation must address incremental interception of clinically important prescription problems together with false reassurance, misleading challenge, delay, alert burden, inequity, and responsibility diffusion. The contribution is conceptual rather than validated: it organizes independent review, authority boundaries, conflict resolution, and governance into a testable human–AI safety architecture. Its applicability remains conditional on prescription type, data quality, professional capacity, local workflow, external validation, and prospective evaluation of the combined work system.

Keywords: Digital pharmacy, Artificial intelligence, Pharmacy practice, Medication safety, Human–AI collaboration, Clinical decision support

Introduction

The First-Reader Assumption

Preventable patient harm occurs across healthcare settings, but its existence does not establish that another automated decision layer will necessarily improve care [1]. Medication errors arise during prescribing, dispensing, administration, and monitoring, with risk shaped by treatment complexity, patient vulnerability, communication failures, and incomplete clinical information [2]. High-risk prescriptions therefore pose a dual problem: important abnormalities may be missed, yet indiscriminate checking can add noise, delay, and false reassurance.

Clinical decision support is often discussed as though its position in the workflow were neutral. It is not. A system shown before professional review can define what the user notices, narrow the hypothesis space, and turn subsequent checking into confirmation. A system permitted to recommend or clear action can also displace attention from contextual factors that were absent from its inputs. Clinical decision-support systems may improve information access and consistency, but their effects depend on data quality, interface design, workflow fit, alert burden, and user response [3]. These dependencies make "who reads first" a safety question rather than a presentation preference.

The first-reader assumption is the implicit belief that AI should inspect a prescription before the responsible professional and present a prioritized interpretation for acceptance or override. This arrangement may be efficient for some low-consequence tasks, but it is epistemically fragile in

Address for correspondence: Sara Al-Fahd, Department of High-Risk Prescription Safety, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia.
s.alfahd@ksu.edu.sa
Received: 26 January 2025; **Accepted:** 25 April 2025

How to cite this article: Al-Fahd S, Al-Khalifa N, Al-Turki O. Artificial Intelligence as a Second Reader for High-Risk Prescriptions rather than a First Decision Maker. Arch Pharm Pract. 2025;16(2):35-43. https://doi.org/10.51847/ROT7TCFRhD

high-risk medication decisions. The professional may cease to form an independent judgment, while a non-alert may be interpreted as evidence of safety. Conversely, excluding AI entirely may forgo a complementary pattern-recognition process capable of detecting atypical combinations, dose patterns, or overlooked contextual contradictions.

Clinical decision support and pharmacy AI encompass heterogeneous assistive functions, and the existence of an algorithmic output does not establish that the system should become the first or final medication decision maker [3, 4]. This article therefore proposes a different role: AI as an independent second reader for selected high-risk prescriptions. "Second reader" denotes a bounded challenge function performed after a human first read but before final disposition. It does not denote a second prescriber, a substitute pharmacist, or an autonomous authorization mechanism. The article develops the model's components, reconciliation states, workflow conditions, evaluation requirements, and governance boundaries without claiming clinical validation.

Independent Review in High-Risk Medication Decisions

Independent review is not synonymous with duplicating the same task. It is a second examination designed to preserve the possibility of identifying a different error pathway. Evidence on medication double checking shows that the practice is widespread but variably defined, inconsistently implemented, and not supported by conclusive evidence of effectiveness across settings [5]. The relevant lesson is neither that checking is useless nor that every prescription needs two readers. It is that review quality depends on independence, task design, error detectability, timing, and the action available when concern is identified.

Direct observation in paediatric medication administration similarly indicates that recording or performing a double check does not by itself establish lower error rates [6]. A nominal second check can become social confirmation, especially when the second reviewer sees the first conclusion, assumes that key calculations were completed, or lacks time to reconstruct the decision. For high-risk prescribing, independence should therefore be understood across four proposed dimensions: cognitive independence from the first conclusion; analytical independence in the detection process; temporal independence before authorization or administration; and authority independence, under which the second reader can challenge but cannot silently determine the final action.

Human expertise remains central, but an additional professional review should not be idealized as automatically effective. A randomized pharmacist intervention involving patients prescribed high-risk medicines after hospitalization did not establish universal improvement across measured medication-safety outcomes [7]. The value of review is conditional on access to relevant information, intervention fidelity, communication, acceptance, and follow-through. An AI second reader should consequently be evaluated against strengthened current practice, not against an artificial comparator in which human review is assumed to be flawless.

High risk is also multidimensional. Incident analyses involving high-alert medicines show that errors vary across medicines and process stages [8]. The proposed trigger for a second read should therefore consider potential severity, reversibility, dose sensitivity, route, monitoring dependence, patient vulnerability, treatment transition, incomplete information, contextual complexity, and time criticality. These dimensions are proposed organizing criteria rather than a validated score. They should determine whether a second read is appropriate, what it examines, and how quickly disagreement must be resolved.

Table 1 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for definitions, components, relationships, and intended uses for the article Artificial Intelligence as a Second Reader for High-Risk Prescriptions rather than a First Decision Maker.

Table 1. Definitions, components, relationships, and intended uses for the article Artificial Intelligence as a Second Reader for High-Risk Prescriptions rather than a First Decision Maker.

Component or scientific dimension	Problem addressed	Inputs or determinants	Proposed mechanism or relationship	Expected contribution	Evidence required	Failure or uncertainty risk	Boundary statement
High-risk prescription trigger	Uniform review wastes attention or misses context-sensitive risk	Potential severity, reversibility, dose sensitivity, route, monitoring, patient vulnerability, transition, and time criticality [2, 8]	**Proposed synthesis:** activate a second read only when locally specified risk conditions are met	Concentrate independent review where plausible safety value is greatest	Local prevalence, severity, actionability, and workload evidence	Overinclusive triggers create burden; underinclusive triggers miss harm	Not a universal medicine list or validated risk score
Human first read	AI-first display may pre-frame professional interpretation	Prescription, indication, patient state, treatment history, and professional knowledge	**Proposed:** the accountable professional forms an initial assessment before seeing the AI conclusion	Preserve contextual and cognitive independence	Human-factors and workflow evaluation	Superficial first reading may become ceremonial	Human-first does not imply human infallibility

Independent second read	Repetition may reproduce the same omission	Structured clinical data, prescription features, and separately specified detection logic	**Proposed:** AI reviews without authority to alter the first assessment	Add a distinct error-search pathway	Comparative assessment of incremental detection	Shared data and design assumptions can create correlated failure	Independence is partial and must be measured
Review quality	A documented check may not be an effective check	Time, task clarity, data access, reconstruction, and escalation capacity [5, 6]	Review effectiveness depends on how checking is performed	Prevent "double check" from becoming a process label	Observation of fidelity and resolution	Confirmation, interruption, and responsibility diffusion	Presence of a check is not evidence of benefit
Human expertise and intervention	Additional review may fail without implementation support	Expertise, communication, intervention acceptance, and follow-through [7]	Professional interpretation converts concern into contextual action	Retain medication authority and accountability	Prospective workflow and outcome evidence	Expertise may be unavailable or inconsistently deployed	Human oversight is necessary but not sufficient
AI challenge function	Important patterns may be overlooked by one reader	Dose, route, duplication, interaction, monitoring, and atypicality signals	**Proposed:** detect, prioritize, explain uncertainty, or abstain	Create a complementary safety challenge	External validation and human–AI evaluation	False reassurance, misleading challenge, or alert burden	AI may not prescribe, authorize, or adjudicate conflict
Escalation pathway	Detection without resolution does not protect patients	Severity, uncertainty, missing data, and disagreement type	**Proposed:** route unresolved concern to proportionate human review	Connect detection to accountable disposition	Timeliness, feasibility, and safety-event assessment	Delay, unresolved alerts, or escalation overload	Escalation rules require local validation

The table defines a conceptual structure rather than a clinical recommendation. The evidence supports examining independence and review fidelity, but it does not establish that AI second reading improves safety or that every high-risk prescription warrants the same pathway.

Proposed Second-Reader Model

Prescription-risk modelling demonstrates that machine learning can prioritize orders with a higher predicted likelihood of medication error [9]. Prediction, however, is not verification: a risk score can identify cases for attention without determining whether a prescription is wrong. The proposed model uses this capability only as part of a bounded second read. Entry occurs after a human first assessment and after a locally governed risk-trigger gate confirms that the task, data, and available escalation pathway are suitable.

AI-supported exclusion of prescriptions from full medication review illustrates both the attraction and danger of triage. Such systems may reduce review volume, but a false negative can remove the very scrutiny intended to protect a high-risk patient [10]. The proposed architecture therefore includes a data-sufficiency and domain-fit gate before interpretation. The AI must abstain when required inputs are absent, stale, contradictory, or outside the evaluated domain. A non-alert is not represented as clearance unless the relevant use and threshold have been prospectively justified.

Atypical-order models may provide pharmacists with an additional screening signal, although atypicality is neither equivalent to error nor sufficient for action [11]. The AI output should therefore be a structured finding rather than a verdict: issue category, supporting data trace, missing information, uncertainty, urgency, and applicable limitations. The human retains treatment-context interpretation and may maintain a concern even when the AI does not. This "human-only concern" pathway is essential because otherwise the second reader would quietly become the first authority.

Figure 1 presents the original conceptual synthesis for artificial intelligence as a second reader for high-risk prescriptions, showing how the article's principal components, evidence relationships, uncertainties, and decision boundaries are connected.

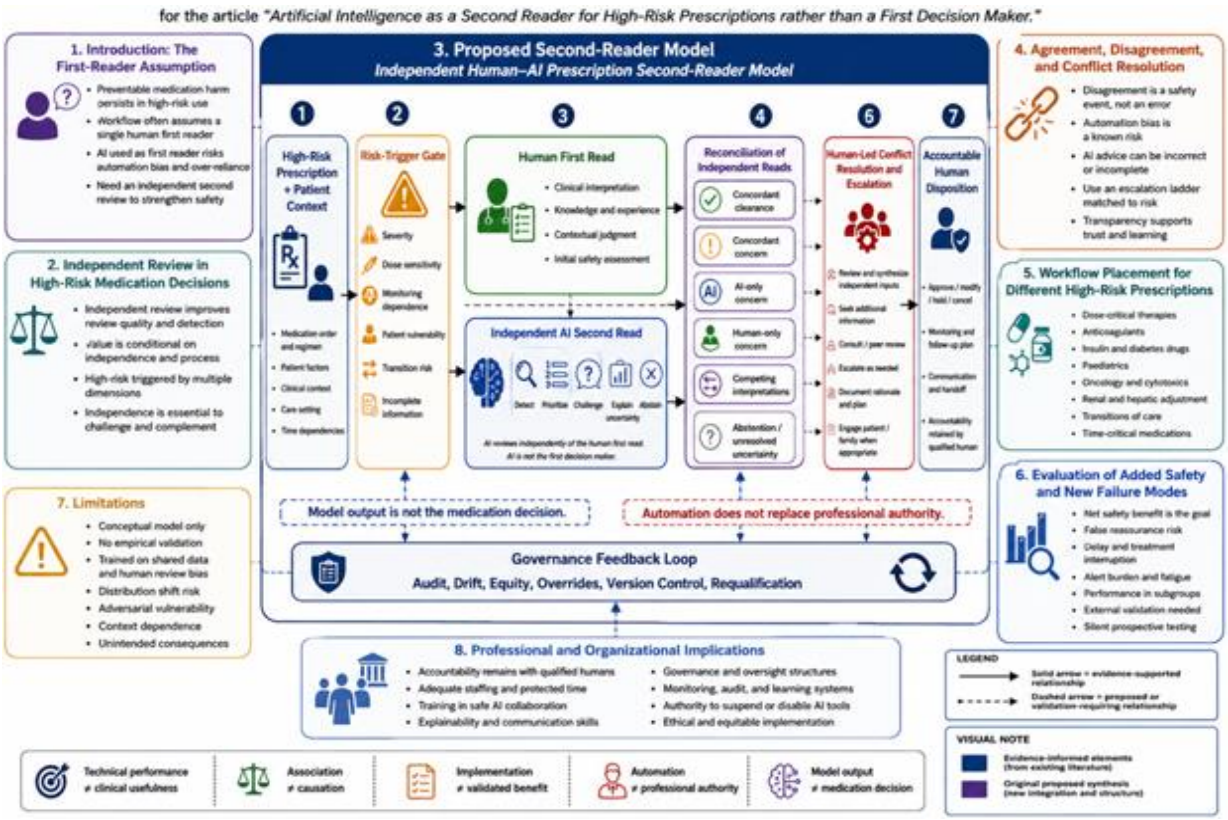

Figure 1. Artificial Intelligence as a Second Reader for High-Risk Prescriptions

The figure is an original conceptual synthesis developed for "Artificial Intelligence as a Second Reader for High-Risk Prescriptions rather than a First Decision Maker." Arrows and grouping indicate proposed or evidence-supported relationships rather than measured effect sizes. The figure does not represent a validated predictive model, clinical recommendation, regulatory determination, or deployment-ready system.

The central reconciliation layer receives two independently formed readings and classifies their relationship. Existing inpatient medication decision-support implementations demonstrate that probabilistic systems can be incorporated into prescription workflows, but they do not establish that the present architecture is safe across institutions or prescription types [12]. Existing prescription-risk, medication-review triage, and atypical-order models support the feasibility of AI-assisted screening, but they do not validate a universal autonomous prescription-review pathway [9-11]. The proposed model consequently prohibits a direct path from AI output to authorization.

The complete architecture comprises six linked stages: high-risk trigger; human first read; independent AI second read; reconciliation; human-led resolution; and governance surveillance. The AI can detect, prioritize, challenge, communicate uncertainty, or abstain. It cannot prescribe, release, suppress mandatory human review, or close an unresolved conflict. The final disposition is recorded by an authorized professional together with the rationale for accepting, modifying, rejecting, or escalating the prescription.

Table 2 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for evaluation criteria, evidence requirements, and failure modes for the article Artificial Intelligence as a Second Reader for High-Risk Prescriptions rather than a First Decision Maker.

Table 2. Evaluation criteria, evidence requirements, and failure modes for the article Artificial Intelligence as a Second Reader for High-Risk Prescriptions rather than a First Decision Maker.

Architecture layer or process stage	Core function	Information flow	Interaction with other components	Evaluation criterion	Potential failure mode	Human or organizational responsibility	Readiness boundary
Risk-trigger gate	Select cases for second reading	Prescription and patient-risk features enter before AI review	Activates or bypasses the second-reader pathway	Clinically important risk coverage and	Over-triggering, missed eligible cases, or	Approve local criteria and audit exclusions	**Proposed:** no deployment until locally and

				manageable workload	inequitable selection		externally evaluated
Human first read	Form contextual professional assessment	Human interpretation precedes AI conclusion	Supplies an independent comparator, not real-time training feedback	Fidelity, completeness, and resistance to pre-framing	Ceremonial review or retrospective rationalization	Protect time, access, and accountability	Human-first sequencing alone does not establish safety
Data/domain gate	Determine whether AI review is supportable	Checks completeness, recency, provenance, and domain fit	Can route directly to abstention and escalation	Reliable detection of unsupported cases	Forced output from missing or shifted data	Define mandatory data and abstention rules	**Proposed:** abstention behaviour requires prospective testing
AI second read	Search for risk patterns or atypicality	Separately processed inputs produce structured findings [9, 11]	Sends concern, uncertainty, or abstention to reconciliation	External performance, calibration, and incremental detection	False reassurance, spurious concern, or correlated failure	Validate versions and disclose limitations	Prediction does not constitute prescription verification
Review-triage function	Prioritize attention without silently removing required review	Ranked or filtered cases inform workflow [10]	Interacts with workload and mandatory-review policy	Missed clinically important problems and workload effect	Unsafe rule-out or hidden false negatives	Set conservative limits and review overrides	No autonomous exclusion of mandatory human review
Reconciliation layer	Compare independently formed readings	Human and AI outputs meet only after both are recorded	Produces agreement, disagreement, or uncertainty state	Correct classification and timely routing	Agreement mistaken for correctness	Define states and ensure visible uncertainty	**Proposed taxonomy:** requires human-factors validation
Conflict-resolution pathway	Resolve concern through accountable review	Data and rationale move to pharmacist, prescriber, or specialist	Converts disagreement into documented disposition	Resolution quality, timeliness, and escalation burden	Delay, unresolved conflict, or authority diffusion	Maintain staffed escalation routes	AI cannot adjudicate unresolved conflict
Final disposition	Authorize, modify, defer, or reject through human authority	Professional decision and rationale enter the audit record	Closes the case and informs governance review	Traceability and appropriate action	Rubber-stamping or undocumented override	Named authorized professional owns the decision	No transfer of professional accountability to AI
Governance surveillance	Detect drift, inequity, incidents, and harmful interaction	Audits, overrides, events, and version changes return to oversight	May restrict, suspend, or require requalification	Sustained performance and net safety	Unmonitored degradation or normalization of deviance	Organization controls change, audit, and suspension	Implementation is not evidence of validated benefit

The proposed model is therefore a role-and-process architecture, not a clinical prediction rule. Its readiness depends on evidence that the combined system detects otherwise missed clinically important problems without producing unacceptable false reassurance, workload, delay, inequity, or responsibility diffusion.

Agreement, Disagreement, and Conflict Resolution

Agreement is not the primary safety endpoint of a second-reader system. Two readers may agree because both are correct, because both share the same missing information, or because one has influenced the other. Automation bias can produce omission errors when users fail to act because a system is silent and commission errors when they follow an incorrect suggestion; verification becomes harder as system reasoning and clinical context become more complex [13]. The proposed model therefore protects the timing of independent judgment and records both assessments before reconciliation.

Incorrect AI advice can shift human decisions toward error, including among trained clinicians [14]. This evidence does not directly validate behaviour in pharmacy review, but it establishes a plausible human-factors hazard. The AI result should not be displayed as a recommended final action, and confidence formatting should not imply authority. When the AI raises a concern, the reviewer should be able to inspect the relevant data and limitations; when it raises none, the professional must remain free to sustain and escalate an independently formed concern.

Human–AI collaboration can improve average performance in some tasks, but benefits vary across users, cases, and forms of assistance [15]. Accordingly, the model distinguishes six proposed reconciliation states. Concordant clearance means neither reader identifies a material concern; it is not proof of correctness. Concordant concern means both identify a potentially important problem, although they may differ on rationale or remedy. AI-only concern requires human examination of the flagged issue. Human-only concern proceeds to resolution despite AI silence. Competing

interpretations occur when both recognize the issue but disagree about its meaning or action. Unresolved uncertainty or AI abstention signals that available information cannot support a reliable disposition.

Correct assistance may help, whereas incorrect assistance may degrade human classification [16]. Evaluation must therefore examine not only whether the AI is right, but whether its presence changes professional behaviour appropriately. A useful second reader should increase attention to valid concerns without suppressing justified human doubt. Harmful interaction includes persuaded error, false reassurance, indiscriminate escalation, or prolonged delay even when the underlying model metric appears acceptable.

Conflict resolution follows a proposed graded ladder. First, verify patient identity, medicine, dose, route, timing, indication, laboratory data, allergies, treatment history, and information provenance. Second, recover missing context and determine whether the apparent anomaly is intentional. Third, inspect the AI evidence trace and uncertainty without treating explanation as validation. Fourth, obtain an independent pharmacist or clinician review. Fifth, clarify treatment intent with the prescriber. Sixth, escalate to specialist or institutional medication-safety pathways when potential severity, uncertainty, or irreversibility warrants it. The accountable professional then records the final disposition and rationale.

This ladder does not presume that disagreement identifies the correct party. It treats disagreement as a signal that the prescription cannot be safely closed through passive acceptance. Organizational policy must specify which conflicts halt progression, which permit time-bounded clarification, and which require urgent escalation. Those rules remain proposed until tested for timeliness, feasibility, workload, and safety in the intended environment.

Workflow Placement for Different High-Risk Prescriptions

A second-reader pathway should not be inserted indiscriminately into every prescription workflow. Repeated alerts, work complexity, and workload can reduce attention and contribute to alert fatigue [17]. A second read is therefore most defensible when the possible harm is substantial, the relevant error is detectable from available information, a timely corrective action exists, and the organization can resolve disagreement without displacing other essential work.

Machine learning may help prioritize medication alerts, but filtering introduces a safety trade-off between reducing low-value interruptions and suppressing clinically important concerns [18]. The proposed model distinguishes prioritization from clearance. AI may order cases by review priority or suppress optional low-value notifications only after task-specific validation; it should not silently remove mandatory professional review. A human concern must also remain actionable when the AI produces no alert.

Workflow placement should reflect the error pathway. Narrow-therapeutic-index or dose-critical prescriptions may require review of dose, organ function, concentration, monitoring, and recent physiological change. Paediatric and weight-based prescriptions may require verification of weight, units, formulation, concentration, and maximum-dose logic. Anticoagulant, insulin, opioid, sedative, oncology, renal-adjustment, and transition-of-care prescriptions present different combinations of severity, contextual dependence, reversibility, and time sensitivity. These examples are proposed categories rather than validated prescribing rules.

Hybrid systems may combine explicit rules for known contraindications or dose limits with learned detection of atypical orders and less obvious combinations [19]. The two approaches should remain distinguishable in the output because a violated rule, a statistical outlier, and a predicted risk are not equivalent forms of evidence. Each requires a different explanation, response, and validation strategy.

Current evidence on AI-supported medication alerts remains limited by internal validation, restricted settings, heterogeneous outcomes, and relatively little routine implementation evidence [20]. Workflow placement should therefore proceed by bounded use case, not institution-wide technological availability. Silent evaluation, simulation, and prospective observation should precede consequential use. Time-critical prescriptions may warrant only rapidly actionable, high-severity checks, whereas complex but non-urgent prescriptions may permit fuller contextual review. The appropriate placement is conditional on the prescription, patient, information environment, available professionals, and escalation capacity.

Evaluation of Added Safety and New Failure Modes

The model must be evaluated as a combined human–AI work system. Responsible clinical machine learning requires attention to stakeholders, data provenance, intended use, monitoring, and consequences across the system lifecycle [21]. Technical discrimination is necessary but cannot establish that the second-reader arrangement improves medication decisions.

Comparative evaluation should describe the AI intervention, comparator, user interaction, errors, exclusions, and analysis sufficiently to determine what produced any observed effect [22]. Relevant stages include retrospective technical validation, temporal and external validation, silent prospective testing, simulated human–AI review, early live evaluation, and comparison with strengthened current practice. Progression between stages should depend on

predefined evidence and stopping conditions rather than technological availability.

Early clinical evaluation must examine users, workflow integration, human–AI interaction, safety events, and implementation context [23]. Responsible evaluation therefore requires lifecycle oversight, transparent reporting of the AI intervention, and direct examination of users, workflow, human–AI interaction, and safety during early live evaluation [21-23].

The principal proposed criterion is net clinically important error interception: clinically important prescription problems intercepted because of the second-reader arrangement, considered together with new harms caused by false reassurance, misleading challenge, delay, excessive escalation, alert burden, responsibility diffusion, or displaced professional attention. Agreement, alert acceptance, sensitivity, or model accuracy alone cannot demonstrate net safety.

Evaluation should separately assess data completeness, domain fit, calibration, abstention, external performance, incremental detection, independence fidelity, disagreement resolution, response time, workload, override patterns, and sustained performance after workflow or model change. Subgroup analysis is essential because aggregate performance can conceal systematically greater underdetection in underserved populations [24]. Equity assessment should examine both model output and downstream response: an equivalent alert may not produce equivalent benefit where access to specialists, monitoring, or timely clarification differs.

A transition should stop when required inputs are unreliable, clinically important failures remain insufficiently detected, incorrect AI advice worsens human decisions, unresolved conflicts accumulate, delays become unsafe, or subgroup harm cannot be acceptably controlled. These are proposed evaluation boundaries, not validated thresholds.

Professional and Organizational Implications

Clinical impact depends on generalizability, data quality, workflow compatibility, monitoring, professional adoption, and the capacity to act on model output [25]. The organization, rather than an individual pharmacist alone, must therefore own implementation conditions. It should define the intended use, eligible prescriptions, required data, staffing, escalation routes, audit schedule, change control, incident response, suspension authority, and criteria for requalification.

Professional roles should remain explicit. The prescriber retains responsibility for treatment intent and authorization. The first-reading pharmacist or clinician performs the primary contextual review. An independent pharmacist, prescriber, or specialist resolves conflicts when escalation is required. Developers are responsible for documentation, provenance, version control, known limitations, and notification of material change. A governance body controls approval of use cases, monitoring, equity review, and suspension. The AI has no professional or moral accountability; it performs a technically bounded information-processing function.

Explainability must be matched to the user and task. Pharmacists may require the relevant prescription features, missing data, uncertainty, and reason for escalation; prescribers may require the clinical conflict and action needed; governance teams require provenance, performance, drift, subgroup results, and version history. Explainability consequently includes traceability, contestability, and access to limitations, not merely a visually persuasive rationale. Its requirements span technical, clinical, ethical, legal, and patient-facing concerns [26].

Explanations cannot establish that a model is valid, causal, unbiased, or clinically useful [27]. A plausible explanation may increase inappropriate trust in an incorrect output. Education should therefore address automation bias, AI abstention, disagreement handling, uncertainty, data limitations, override documentation, and the continuing duty to escalate human-only concerns. Competence should be assessed through realistic conflict scenarios rather than passive completion of system training.

Limitations

This contribution is conceptual and has not been prospectively evaluated. Machine-learning decision support can create unintended consequences that are not captured by technical performance measures [28]. A second reader may add delay, false reassurance, duplicated work, or diffusion of responsibility.

The evidence base includes medication administration and non-pharmacy diagnostic tasks because direct prescription-specific human–AI interaction evidence remains limited. Prescription models may also be vulnerable to distribution shift, malformed data, and adversarial or subtle input perturbations [29]. Human and AI readers may share the same incomplete records and institutional biases, limiting genuine independence.

Transparent reporting is necessary for appraisal but cannot replace external validation, clinical evaluation, or post-implementation surveillance [30]. High risk remains context-dependent, and some organizations may obtain greater benefit from strengthening human review than from introducing AI. The proposed trigger dimensions, reconciliation states, escalation ladder, and readiness boundaries require empirical validation.

Conclusion

Artificial intelligence should not become the first or final authority in high-risk prescribing merely because it can

process prescriptions rapidly or identify statistical patterns. The proposed second-reader model instead preserves a human first interpretation, introduces a separately bounded AI challenge, and returns every unresolved concern to accountable professional resolution.

Its central scientific claim is testable rather than established: independent human and AI reading may identify complementary error pathways when their roles, timing, uncertainty, abstention, and authority are explicitly controlled. Agreement is treated as corroboration rather than proof, while disagreement becomes a structured safety event requiring contextual verification and proportionate escalation.

The model also shifts evaluation from isolated algorithm performance to net work-system safety. Potential benefit must be assessed alongside false reassurance, inappropriate challenges, delay, workload, unequal performance, security risks, and responsibility diffusion. Professional accountability remains human, while organizations retain responsibility for validation, staffing, education, monitoring, change control, and suspension.

This architecture offers an original way to organize human–AI collaboration in digital pharmacy without assuming that automation is equivalent to independent review. It does not establish clinical superiority, universal applicability, regulatory acceptance, or deployment readiness. Its value will depend on whether prospective, externally validated, equity-sensitive evaluation shows that it intercepts clinically important prescription problems that current practice misses without creating greater harms elsewhere in the medication-use system.

ACKNOWLEDGMENTS: None
CONFLICT OF INTEREST: None
FINANCIAL SUPPORT: None
ETHICS STATEMENT: None

References

1. Panagioti M, Khan K, Keers RN, Abuzour A, Phipps D, Kontopantelis E, et al. Prevalence, severity, and nature of preventable patient harm across medical care settings: Systematic review and meta-analysis. BMJ. 2019;366. doi:10.1136/bmj.l4185
2. Assiri GA, Shebl NA, Mahmoud MA, Aloudah N, Grant E, Aljadhey H, et al. What is the epidemiology of medication errors, error-related adverse events and risk factors for errors in adults managed in community care contexts? A systematic review of the international literature. BMJ Open. 2018;8(5). doi:10.1136/bmjopen-2017-019101
3. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: Benefits, risks, and strategies for success. NPJ Digit Med. 2020;3:17. doi:10.1038/s41746-020-0221-y
4. Chalasani SH, Syed J, Ramesh M, Patil V, Pramod Kumar TM. Artificial intelligence in the field of pharmacy practice: A literature review. Explor Res Clin Soc Pharm. 2023;12:100346. doi:10.1016/j.rcsop.2023.100346
5. Koyama AK, Maddox CSS, Li L, Bucknall T, Westbrook JI. Effectiveness of double checking to reduce medication administration errors: A systematic review. BMJ Qual Saf. 2020;29(7):595-603. doi:10.1136/bmjqs-2019-009552
6. Westbrook JI, Li L, Raban MZ, Woods A, Koyama AK, Baysari MT, et al. Associations between double-checking and medication administration errors: A direct observational study of paediatric inpatients. BMJ Qual Saf. 2021;30(4):320-30. doi:10.1136/bmjqs-2020-011473
7. Gurwitz JH, Kapoor A, Garber L, Mazor KM, Wagner J, Cutrona SL, et al. Effect of a multifaceted clinical pharmacist intervention on medication safety after hospitalization in persons prescribed high-risk medications: A randomized clinical trial. JAMA Intern Med. 2021;181(5):610-8. doi:10.1001/jamainternmed.2020.9285
8. Kuitunen S, Saksa M, Tuomisto J, Holmström AR. Medication errors related to high-alert medications in a paediatric university hospital - a cross-sectional study analysing error reporting system data. BMC Pediatr. 2023;23(1):548. doi:10.1186/s12887-023-04333-2
9. 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
10. Levivien C, Cavagna P, Grah A, Buronfosse A, Courseau R, Bézie Y, et al. Assessment of a hybrid decision support system using machine learning with artificial intelligence to safely rule out prescriptions from medication review in daily practice. Int J Clin Pharm. 2022;44(2):459-65. doi:10.1007/s11096-021-01366-4
11. Hogue SC, Chen F, Brassard G, Lebel D, Bussières 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
12. Segal G, Segev A, Brom A, Lifshitz Y, Wasserstrum Y, Zimlichman E. Reducing drug prescription errors and adverse drug events by application of a probabilistic, machine-learning based clinical decision support system in an inpatient setting. J Am Med Inform Assoc. 2019;26(12):1560-5. doi:10.1093/jamia/ocz135
13. Lyell D, Coiera E. Automation bias and verification complexity: A systematic review. J Am Med Inform Assoc. 2017;24(2):423-31. doi:10.1093/jamia/ocw105
14. Gaube S, Suresh H, Raue M, Merritt A, Berkowitz SJ, Lermer E, et al. Do as AI say: Susceptibility in deployment of clinical decision-aids. NPJ Digit Med. 2021;4(1):31. doi:10.1038/s41746-021-00385-9
15. Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, et al. Human-computer collaboration for skin cancer recognition. Nat Med. 2020;26(8):1229-34. doi:10.1038/s41591-020-0942-0
16. Kiani A, Uyumazturk B, Rajpurkar P, Wang A, Gao R, Jones E, et al. Impact of a deep learning assistant on the histopathologic classification of liver cancer. NPJ Digit Med. 2020;3:23. doi:10.1038/s41746-020-0232-8
17. Ancker JS, Edwards A, Nosal S, Hauser D, Mauer E, Kaushal R; with the HITEC Investigators. Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. BMC Med Inform Decis Mak. 2017;17(1):36. doi:10.1186/s12911-017-0430-8
18. Liu S, Kawamoto K, Del Fiol G, Weir C, Miao Y, Rupper R, et al. The potential for leveraging machine learning to filter medication alerts. J Am Med Inform Assoc. 2022;29(5):891-9. doi:10.1093/jamia/ocab292
19. Lee S, Shin J, Kim HS, Lee MJ, Yoon JM, Lee S, et al. Hybrid method incorporating a rule-based approach and deep learning for prescription error prediction. Drug Saf. 2022;45(1):27-35. doi:10.1007/s40264-021-01123-6
20. Graafsma J, Murphy RM, van de Garde EMW, Karapinar-Çarkit 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
21. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: A roadmap for responsible machine learning for health care. Nat Med. 2019;25(9):1337-40. doi:10.1038/s41591-019-0548-6
22. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. Nat Med. 2020;26(9):1364-74. doi:10.1038/s41591-020-1034-x

23. 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
24. Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nat Med. 2021;27(12):2176-82. doi:10.1038/s41591-021-01595-0
25. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17(1):195. doi:10.1186/s12916-019-1426-2
26. Amann J, Blasimme A, Vayena E, Frey D, Madai VI; Precise4Q consortium. Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20(1):310. doi:10.1186/s12911-020-01332-6
27. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Health. 2021;3(11). doi:10.1016/S2589-7500(21)00208-9
28. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. JAMA. 2017;318(6):517-8. doi:10.1001/jama.2017.7797
29. Finlayson SG, Bowers JD, Ito J, Zittrain JL, Beam AL, Kohane IS. Adversarial attacks on medical machine learning. Science. 2019;363(6433):1287-9. doi:10.1126/science.aaw4399
30. 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. doi:10.1136/bmj-2023-078378