Original Article

A Pharmacy Gatekeeping System for Admitting, Restricting, and Withdrawing Clinical Algorithms

Nora Schneider[1]*, Tobias Frank[1], Andreas Müller[2], Christoph Meier[3]

[1]Department of Pharmacy Algorithm Governance, Faculty of Pharmacy, University of Freiburg, Freiburg, Germany. [2]Department of Clinical Algorithm Lifecycle Management, Faculty of Pharmacy, Heidelberg University, Heidelberg, Germany. [3]Department of Gatekeeping and AI Withdrawal, Faculty of Pharmacy, University of Hohenheim, Stuttgart, Germany.

Abstract

Clinical algorithms increasingly influence medication surveillance, alert prioritization, risk estimation, prescription review, and other pharmacy-related decisions. Yet institutions commonly govern their acquisition, technical installation, clinical validation, and day-to-day use through separate processes that do not establish a coherent decision about whether an algorithm should be allowed to enter, remain within, or leave medication-use practice. This article proposes a Pharmacy Algorithm Gatekeeping Architecture that treats algorithmic admission as a continuing institutional governance decision rather than a one-time technical approval. The architecture comprises a Pharmacy Algorithm Dossier, multidisciplinary gatekeeping review, bounded Permission Envelope, monitored-use layer, documented state transitions, and re-entry review. Admission is conditioned on the algorithm's intended task, users, population, setting, data dependencies, permitted outputs, degree of autonomy, uncertainty handling, safety controls, equity implications, and monitoring readiness. Continuing permission may be narrowed through restriction, interrupted through suspension, or revoked through withdrawal when evidence, performance, workflow, safety, accountability, or implementation conditions become unacceptable. Remediation does not automatically restore permission; re-entry requires a renewed assessment capable of producing different boundaries or rejection. Evaluation must extend beyond model accuracy to calibration, human–AI joint performance, medication-safety consequences, workflow effects, subgroup performance, implementation fidelity, incident detection, and governance traceability. The architecture is an original non-empirical synthesis whose components and transition rules require prospective, comparative, and setting-specific validation. It does not replace regulatory authorization, professional judgment, local legal duties, or clinical accountability, and it does not establish that any admitted algorithm is universally safe, effective, or ready for unrestricted use.

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

Introduction

Clinical algorithms can influence medication-related work even when they do not directly prescribe, dispense, or administer a medicine. They may determine which prescriptions receive attention, which interactions are displayed, which patients are classified as high risk, which alerts are suppressed, or when a professional is prompted to intervene. Their effects therefore arise not only from predictions but also from how outputs redistribute attention, shape workflow, frame uncertainty, and alter the conditions under which medication decisions are made. Machine-learning systems may produce unintended clinical consequences that are not captured by conventional model-performance measures [1].

Responsible clinical artificial intelligence consequently requires safeguards across problem formulation, data selection, validation, implementation, stakeholder involvement, monitoring, and updating [2]. These responsibilities are especially important in pharmacy because apparently modest changes in prioritization or alert presentation can affect multiple professionals and medication-use stages. An algorithm may be technically accurate yet poorly aligned with a pharmacy task, insufficiently calibrated for the local population, incompatible with professional roles, or unable to communicate when its output should be questioned.

Address for correspondence: Nora Schneider, Department of Pharmacy Algorithm Governance, Faculty of Pharmacy, University of Freiburg, Freiburg, Germany.
nora.schneider@natfak.uni-freiburg.de
Received: 29 November 2025; **Accepted:** 03 March 2026

How to cite this article: Schneider N, Frank T, Müller A, Meier C. A Pharmacy Gatekeeping System for Admitting, Restricting, and Withdrawing Clinical Algorithms. Arch Pharm Pract. 2026;17(1):68-76. https://doi.org/10.51847/u7RohUC21I

Institutional controls are therefore needed before clinical influence is granted. Clinical decision support requires evaluation of observed care and workflow consequences because system availability alone does not establish net clinical value [3, 4]. The relevant institutional question is not simply whether an algorithm can run, has been purchased, or has obtained external authorization. It is whether a defined algorithm may influence a defined medication-related task under specified conditions, with sufficient evidence, accountable oversight, and feasible mechanisms for detecting and responding to failure.

This article calls that decision algorithmic admission. Algorithmic admission is proposed as the institutional authorization of a specified clinical algorithm to operate within a bounded medication-use context. It differs from procurement, installation, regulatory authorization, model validation, and individual clinician acceptance. Admission establishes neither permanent permission nor proof of clinical benefit. It creates a conditional relationship between evidence, permitted use, monitoring, and institutional responsibility.

The proposed Pharmacy Algorithm Gatekeeping Architecture addresses the full permission lifecycle. It specifies how an institution may examine an algorithm before admission, define its authorized scope, monitor its operation, narrow or interrupt permission, withdraw it when its justification no longer holds, and reconsider it after remediation. The architecture integrates technical, clinical, sociotechnical, safety, equity, implementation, and governance evidence without treating any single domain as sufficient. Its state categories and decision relationships are proposed constructs requiring empirical validation rather than established regulatory or clinical standards.

Algorithmic Admission as a Governance Decision

Algorithmic admission should be understood as a governance decision because it assigns permission, responsibility, and exposure to risk. Clinical-AI safety cannot be delegated entirely to developers, vendors, regulators, or frontline professionals; healthcare organizations retain responsibilities for how systems are selected, integrated, supervised, and used [5]. In pharmacy, this responsibility includes determining whether the algorithm's output is advisory or action-triggering, which professionals may rely on it, which medication decisions remain outside its scope, and who may halt its operation.

Admission is thus narrower than general organizational endorsement but broader than technical validation. A model may demonstrate discrimination on retrospective data while leaving unresolved questions about calibration, missing data, intended users, human responses, workflow consequences, fairness, and patient benefit. Evidence review must therefore examine transparency, reproducibility, ethical acceptability, effectiveness, and the connection between the technical task and the claimed clinical purpose [6]. The gatekeeping decision should record unresolved uncertainty rather than converting incomplete evidence into an implicit approval.

Regulatory authorization and institutional admission also answer different questions. External authorization may establish that a product has satisfied a particular legal pathway, but it does not automatically determine whether its evidence, intended use, population, data requirements, or controls fit a specific pharmacy service [7]. Institutional admission should not reproduce regulatory review or contradict legal requirements. It should evaluate local permission within those requirements.

This distinction is particularly important where public information about clinical evaluation, comparator selection, subgroup assessment, or external testing is incomplete. Analyses of authorized medical-AI products have identified limitations in the evidence available for evaluating clinical performance and generalizability [8]. A pharmacy institution may therefore need additional documentation, local evaluation, restricted use, or post-admission monitoring before permitting algorithmic influence.

Alternative explanations must remain visible. Poor outcomes associated with an algorithm may originate from data quality, interface design, staffing, training, workflow interruption, or inappropriate reliance rather than the model alone. Conversely, acceptable aggregate performance may conceal subgroup harm, unsafe automation, or alert burden. Gatekeeping should therefore evaluate the combined institutional arrangement without attributing every consequence exclusively to the algorithm.

Table 1 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for definitions, components, relationships, and intended uses for the article A Pharmacy Gatekeeping System for Admitting, Restricting, and Withdrawing Clinical Algorithms.

Table 1. Definitions, components, relationships, and intended uses for the article A Pharmacy Gatekeeping System for Admitting, Restricting, and Withdrawing Clinical Algorithms.

Component or Scientific Dimension	Problem Addressed	Inputs or Determinants	Proposed Mechanism or Relationship	Expected Contribution	Evidence Required	Failure or Uncertainty Risk	Boundary Statement
Algorithmic Admission	Installation or purchase may be mistaken for	Intended task, users, population, setting, outputs, autonomy, evidence, controls.	**Proposed:** a formal institutional decision creates bounded and	Makes authorization conditions explicit and reviewable.	Technical, clinical, workflow, safety, equity, and	Admission may become ceremonial or procurement-led.	Admission does not prove benefit, safety, or regulatory acceptance.

	permission to influence care.		revocable permission.		implementation evidence.		
Clinical Algorithm	"Algorithm" may refer to different functions and levels of influence.	Version, inputs, processing logic, output, integration, update status.	The evaluated object includes its configuration and operational context.	Prevents evaluation of an abstract model disconnected from use.	Version-controlled documentation and intended-use description.	Undocumented updates or local modifications change the evaluated object.	A model evaluated outside its implementation context is not the complete clinical system [1].
Institutional Accountability	Responsibility may be dispersed across vendor, managers, and users.	Decision authority, safety roles, escalation routes, contractual duties.	**Proposed:** admission, monitoring, and withdrawal authorities are named separately.	Improves traceability of safety decisions.	Governance records and role allocation.	Shared participation may obscure final accountability.	Human involvement alone does not resolve institutional responsibility [5].
Evidence Sufficiency	Model accuracy may dominate the decision despite unresolved clinical questions.	Validity, calibration, usefulness, reproducibility, ethics, uncertainty.	Evidence is assessed across distinct domains rather than reduced to one score.	Exposes gaps that conditional permission or rejection must address.	Transparent and reproducible evaluation evidence [6].	Complete reporting may be confused with valid or beneficial performance.	Evidence completeness is necessary but not sufficient for admission.
Regulatory–Institutional Relationship	External authorization may be treated as automatic local approval.	Legal status, authorized use, local population, pharmacy workflow, additional controls.	Institutional review operates within, but is distinct from, regulatory requirements.	Connects external authorization to local use conditions.	Authorization records and locally relevant evidence [7].	Duplication, conflict, or false reassurance.	Gatekeeping neither replaces nor overrides regulation.
Permission Envelope	"Approved use" may remain vague and expand informally.	Task, users, setting, population, actions, data dependencies, duration, monitoring.	**Proposed:** admission generates an explicit boundary around permitted use.	Makes scope expansion and deviation detectable.	Locally testable intended-use and monitoring specifications.	Informal workarounds may enlarge the operational scope.	Permission applies only to the documented envelope.
Conditional Permission	Evidence may be promising but incomplete for unrestricted use.	Residual uncertainty, reversibility, monitoring capacity, fallback pathway.	**Proposed:** narrowed use or intensified oversight permits limited learning without implying readiness.	Supports proportionate rather than binary decisions.	Evidence that residual risk can be observed and contained.	"Conditional" status may persist indefinitely without reassessment.	Conditional permission is not provisional proof of effectiveness.
Revocability	Initial admission may be treated as permanent.	Drift, incidents, workflow change, new evidence, data change, organizational capacity.	**Proposed:** permission remains contingent and may be restricted, suspended, or withdrawn.	Aligns continued use with changing evidence and context.	Reliable monitoring and decision authority.	Signals may be delayed, ambiguous, or unavailable.	A monitoring signal initiates review; it does not by itself prove harm.

Proposed Pharmacy Gatekeeping Architecture

The proposed architecture treats gatekeeping as a connected system rather than a single committee decision. Technologies may be abandoned, normalized, adapted, or rendered unsustainable through interactions among the technology, intended value, users, organization, wider system, and changes over time [9]. Accordingly, the gatekeeping object is not merely the trained model. It is the model–data–interface–workflow–organization arrangement through which an output acquires clinical influence.

The architecture begins with a Pharmacy Algorithm Dossier. The dossier identifies the medication-use problem, existing alternatives, algorithm version, intended users, population, data provenance, output, proposed action, degree of autonomy, evaluation evidence, known limitations, uncertainty handling, human-review pathway, change-control process, and monitoring plan. A named institutional sponsor is responsible for completeness but should not possess unilateral admission authority.

The dossier enters a multidisciplinary Gatekeeping Review. Organizational AI governance requires defined oversight structures, responsibilities, lifecycle controls, and mechanisms for managing risk [10]. The proposed review therefore examines distinct domains: clinical need, technical and data validity, calibration and uncertainty, medication safety, human–AI interaction, equity, implementation readiness, accountability, and monitoring feasibility. Reviewers may request additional evidence, reject the proposal, defer a decision, or authorize use within a defined Permission Envelope.

The Permission Envelope is the architecture's central control. It defines the task the algorithm may influence, authorized users, eligible population, clinical setting, required inputs, permitted outputs or actions, degree of automation, fallback

process, monitoring intensity, and expiry or review condition. Trustworthy clinical AI depends on multiple qualities—including fairness, traceability, usability, robustness, and explainability—whose relevance continues across the lifecycle [11]. The Permission Envelope converts these concerns into use boundaries without claiming that the boundaries have already been validated.

Figure 1 presents the original conceptual synthesis for pharmacy algorithm gatekeeping architecture, showing how the article's principal components, evidence relationships, uncertainties, and decision boundaries are connected.

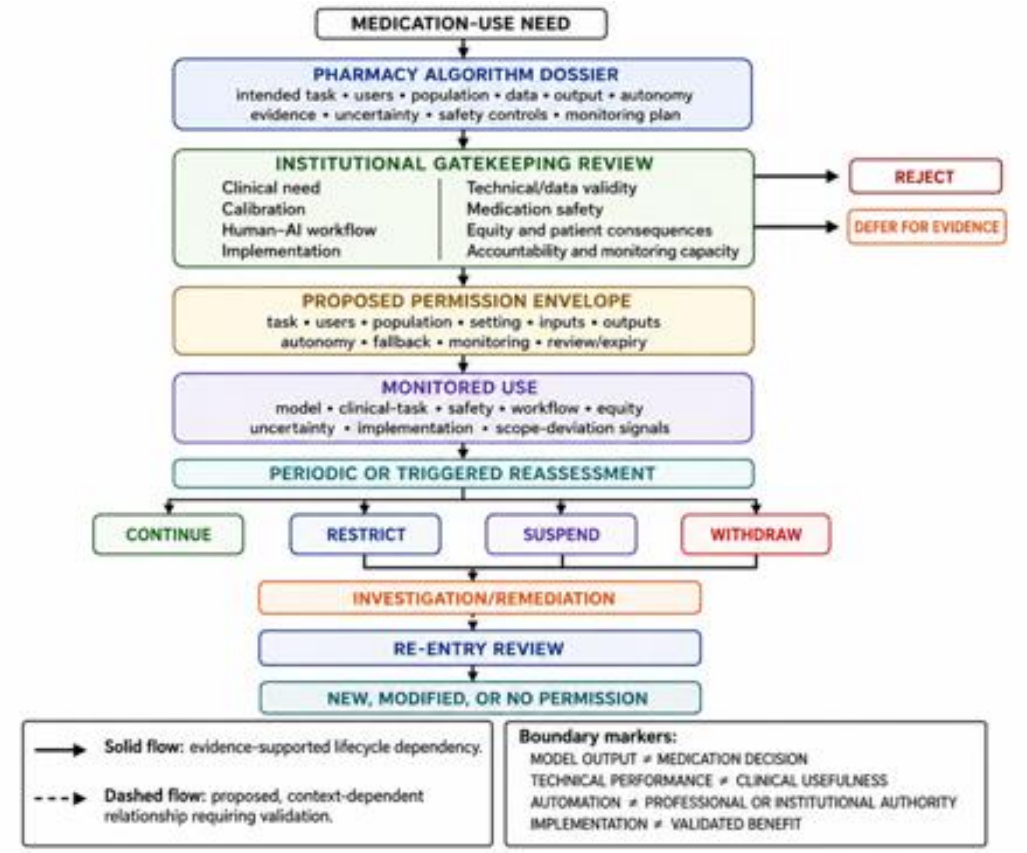

Figure 1. Pharmacy Algorithm Gatekeeping Architecture

After admission, the algorithm enters a Monitored-Use Layer. This layer receives technical, clinical-task, medication-safety, workflow, equity, uncertainty, and implementation signals. Monitoring does not presume that every adverse event was caused by the algorithm or that stable aggregate performance establishes safety. Signals are interpreted against the intended use, baseline workflow, exposure level, data quality, and Permission Envelope.

A State-Transition Function connects monitored use to continuation, restriction, suspension, or withdrawal. Human oversight is included as one control, but it is not treated as an automatic safeguard: clinicians may accept erroneous algorithmic advice, particularly when outputs appear authoritative or are embedded in routine workflow [12]. The architecture consequently requires documented fallback processes, authority to interrupt use, and evidence that users can identify when reliance should be reduced.

Finally, remediation does not return the algorithm directly to its previous state. A Re-entry Review reassesses the corrected algorithm, relevant evidence, changed context, and proposed permission boundaries. It may reject re-entry, request further evidence, or issue a new Permission Envelope. This separation prevents a technical patch from being treated as proof that the original clinical, organizational, safety, or equity concern has been resolved.

Admission Criteria and Conditional Permissions

Admission criteria should operate as structured questions rather than as a universal checklist whose completion guarantees readiness. The evidence package must specify intended use, data sources, model development, evaluation methods, users, workflow, errors, and foreseeable safety conditions [13, 14]. Reporting quality permits scrutiny, but it does not establish that the algorithm is clinically useful, equitable, or suitable for the local pharmacy environment.

The first criterion is clinical and medication-use relevance. The dossier should identify the decision or workflow problem, who currently performs the task, available non-algorithmic alternatives, and the consequence of false, delayed, or omitted outputs. An algorithm designed to suppress medication alerts, for example, requires a different evidentiary burden from one that merely organizes an administrative queue because its errors can directly redistribute safety attention. The claimed purpose must match the outcome that was evaluated.

The second criterion is technical and data validity. Review should examine data provenance, eligibility rules, missingness, label construction, leakage, comparator selection, internal validation, external evaluation, and model-version stability. Performance measures must correspond to the clinical decision. Discrimination alone is insufficient when outputs are used as probabilities or thresholds because calibration describes whether predicted risks correspond to observed risks [15]. No universal calibration threshold is proposed; acceptable error depends on the task, consequence, decision threshold, and available fallback.

The third criterion is human–AI and workflow validity. Institutions should determine who receives the output, when it appears, what action it invites, how uncertainty is represented, and whether professional review is practically possible. A stated human-in-the-loop requirement is inadequate when workload, interface design, time pressure, or authority structures make review superficial. Admission should specify whether the algorithm is silent, advisory, prioritizing, action-triggering, or partially autonomous.

The fourth criterion is medication safety. The review should consider missed hazards, unnecessary interruption, alert burden, correlated failure across patients, delayed recognition, and interactions with existing safeguards. The relevant comparison is not an idealized algorithm versus unaided cognition, but the proposed algorithm-enabled workflow versus current practice and feasible alternatives.

The fifth criterion is equity and patient consequence. Review must examine whether targets, labels, proxies, access patterns, and missingness differ across groups. An algorithm can appear technically successful while producing inequitable decisions if the predicted target is a poor proxy for clinical need [16]. Conditional permission may therefore limit the eligible population, require subgroup monitoring, or prohibit an output from determining access to care without additional review.

The sixth criterion is implementation and governance readiness. Admission requires trained users, technical support, incident pathways, update notification, monitoring capacity, and authority to restrict or stop use. Evidence should show that required inputs will be available with adequate quality and that failure can be contained. A technically strong model should be deferred where the institution cannot observe its operation or act on identified problems.

The proposed decision categories are reject, defer, admit, and admit conditionally. Conditional permission may authorize silent evaluation, limited users, selected populations, advisory-only output, intensified monitoring, shorter review intervals, or mandatory second review. Conditions should correspond to identified uncertainty and have an owner, review point, and possible exit decision. Conditional status must not become indefinite permission by default.

The resulting Permission Envelope should record the authorized algorithm version, task, users, population, setting, input dependencies, output, permissible action, autonomy level, fallback pathway, monitoring indicators, escalation process, and review or expiry condition. Scope changes, material data changes, retraining, interface redesign, new automation, or expansion to another population should trigger review rather than being absorbed as routine implementation.

Admission remains an institutional judgment under uncertainty. Satisfying a criterion means that a relevant question has been addressed to an acceptable local standard; it does not prove guaranteed safety, clinical benefit, universal portability, or regulatory sufficiency. The proposed criteria and conditional-permission categories require empirical evaluation for consistency, proportionality, feasibility, and unintended effects.

Restriction, Suspension, Withdrawal, and Re-entry

Admission does not end institutional evaluation. Clinical algorithms operate within changing populations, data pipelines, workflows, formularies, staffing arrangements, interfaces, and professional practices. Continued permission therefore depends on whether the conditions supporting the original admission remain credible. Postdeployment quality improvement requires continuing surveillance and controlled processes for reviewing or updating clinical algorithms [17].

The proposed architecture distinguishes four post-admission states. Restriction narrows one or more elements of the Permission Envelope while some use continues. An institution might limit an algorithm to advisory output, remove an affected patient subgroup, require additional professional review, disable an automated action, or reduce the number of authorized settings. Restriction is appropriate when a concern appears containable and continuing use remains justifiable within narrower boundaries.

Suspension temporarily interrupts admitted use while a potentially consequential concern is investigated. Suspension may be warranted when the institution cannot promptly determine whether an observed change arises from the model, data pipeline, interface, workflow, or surrounding clinical environment. Temporal deterioration may occur even without intentional modification of the algorithm; calibration drift has

been demonstrated in clinical prediction models as patient and practice conditions change [18]. Suspension should therefore remain available when uncertainty itself creates unacceptable exposure.

Withdrawal revokes institutional permission. It may follow confirmed harm, repeated boundary violations, loss of monitoring capability, unacceptable subgroup performance, unresolved data dependence, vendor non-disclosure, material workflow incompatibility, or disappearance of the clinical justification for use. Withdrawal is not a universal declaration that the algorithm has no value. It is a local determination that the evidentiary and operational basis for its admitted use is no longer sufficient.

State transitions should not depend exclusively on deterioration in aggregate accuracy. Drift-detection methods vary in their sensitivity to the type of shift, sample size, monitored features, and available labels [19]. Relevant triggers may therefore include changes in input distributions, missingness, calibration, override patterns, alert burden, medication incidents, subgroup outcomes, user behavior, workflow configuration, or deviations from the authorized scope. A signal should initiate proportionate review without being treated automatically as proof that the algorithm caused harm.

The review pathway is proposed as: signal verification, immediate containment where necessary, investigation of alternative explanations, assessment of clinical exposure, and a documented decision to continue, restrict, suspend, or withdraw. Emergency containment should not require complete causal attribution when delay could expose patients or professionals to avoidable risk. Conversely, transient statistical variation should not automatically produce withdrawal without considering data quality, clinical relevance, and reproducibility.

Technical remediation may include recalibration, retraining, input correction, interface redesign, revised thresholds, or workflow modification. Corrective approaches can restore performance under some acquisition shifts, but their effectiveness remains dependent on the shift mechanism and evaluation context [20]. Remediation consequently does not create automatic reinstatement.

Re-entry is proposed as a renewed admission decision. The revised dossier must identify the initiating concern, root-cause assessment, corrective action, affected algorithm version, new validation evidence, unresolved uncertainty, and proposed Permission Envelope. Re-entry may be rejected, deferred, or granted with conditions different from those previously applied. This design preserves institutional memory and prevents a corrected technical component from obscuring unresolved medication-safety, workflow, equity, or accountability concerns.

No universal numerical transition threshold is proposed. Thresholds must reflect the clinical task, severity and reversibility of potential consequences, monitoring reliability, uncertainty, exposure, and available fallback. The distinctions among restriction, suspension, withdrawal, and re-entry are original governance constructs whose consistency and effects require empirical assessment.

Gatekeeping Evidence and Performance Measures

Gatekeeping evaluation should begin with the quality and clinical meaning of the available data. Health data may contain missingness, confounding, label error, documentation artifacts, selection effects, and leakage that allow technically persuasive results without establishing transportable clinical value [21]. The dossier should therefore describe how data were generated, which patients or encounters were absent, how outcomes were defined, and whether the model may learn institutional processes rather than the intended clinical phenomenon.

Model evaluation remains necessary but should be disaggregated. Relevant measures may include discrimination, calibration, threshold-specific error, robustness to missing inputs, uncertainty behavior, subgroup performance, and stability across time and settings. Metric selection must follow the intended decision. A model that prioritizes pharmacist review, suppresses alerts, recommends dose reassessment, or predicts medication harm creates different consequences for false-positive, false-negative, delayed, and uncertain outputs.

The evaluated object must also include the human–algorithm system. Clinical assistance can improve or worsen task performance, and its effect may depend on user expertise, interface presentation, confidence, and the quality of the algorithmic recommendation [22, 23]. Gatekeeping should therefore measure whether professionals appropriately accept, reject, verify, or defer outputs rather than assuming that individual model performance will translate directly into safer decisions.

Uncertainty should be operational rather than merely statistical. It should indicate when an output requires additional data, pharmacist checking, a second opinion, delayed action, or escalation. Communicating uncertainty is important because apparently precise predictions may otherwise encourage unjustified confidence [24]. Institutions should evaluate whether uncertainty information is understandable, actionable, and compatible with time-sensitive medication work.

The proposed evaluation portfolio contains seven levels:

1. Model level: discrimination, calibration, robustness, missingness sensitivity, and drift.
2. Clinical-task level: accuracy and timeliness of the medication-related task.

3. Human–AI level: reliance, override, verification, deferral, and joint performance.
4. Medication-safety level: missed hazards, unnecessary interruption, correlated errors, and incident consequences.
5. Workflow level: workload, alert burden, task redistribution, interruptions, and workarounds.
6. Equity level: subgroup performance, access consequences, proxy validity, and unequal exposure to errors.
7. Governance level: review timeliness, traceability, adherence to the Permission Envelope, escalation performance, and closure of corrective actions.

These measures should not be collapsed into an unexplained readiness score. A favorable result in one domain cannot compensate automatically for an unacceptable failure in another. Evaluation should instead produce a structured evidence profile showing which conditions support permission, which require restriction, and which remain unresolved.

Gatekeeping itself also requires evaluation. Relevant research questions include whether reviewers identify material evidence gaps, whether comparable cases receive consistent decisions, whether permission boundaries are followed, how quickly credible safety signals are escalated, whether restrictions are enforceable, and whether re-entry decisions address the initiating failure. These outcomes would test the architecture without presuming its effectiveness.

Institutional Implementation

Implementation should begin with a limited number of high-consequence governance functions rather than an expansive administrative structure. Pharmacy institutions need identifiable responsibility for dossier submission, technical review, medication-safety assessment, workflow and human-factors evaluation, equity review, implementation readiness, monitoring, incident response, and final admission or withdrawal decisions.

Medication-alert optimization illustrates why implementation requires pharmacy-specific scrutiny. Artificial intelligence may help prioritize or suppress medication alerts, but the available literature remains limited by heterogeneous methods, incomplete reporting, and restricted external or prospective validation [25]. Institutions should therefore resist treating a promising model category as evidence that a particular local implementation is ready for unrestricted use.

Implementation conditions include the characteristics of the intervention, the individuals involved, the inner organizational setting, the external environment, and the processes used to plan, engage, execute, and reflect on change [26]. Gatekeeping must consequently be integrated with existing medication-safety, formulary, digital-governance, quality-improvement, informatics, procurement, and incident-management structures rather than operating as an isolated AI committee.

Interaction design and professional-role tailoring may influence medication-alert burden and user response [27]. Pharmacy implementation should therefore identify which professional receives the output, what authority accompanies it, whether it interrupts work, and how its presentation differs across pharmacists, technicians, prescribers, nurses, and other users. Local adaptation should be documented because interface or workflow changes can alter the admitted system even when the underlying model remains unchanged.

Implementation may use an existing committee or a newly constituted body, but the required functions should remain visible. Smaller organizations may share regional expertise, use external technical review, or adopt proportional processes for lower-influence algorithms. Proportionality should reduce unnecessary burden without removing the ability to identify the accountable decision maker, operational boundaries, monitoring owner, and emergency-stop pathway.

Educational preparation is also necessary. Users should understand what the algorithm does, what it does not establish, how uncertainty is represented, when outputs require verification, how to report concerns, and who may alter its scope. Gatekeepers require sufficient literacy to interrogate evidence without assuming that technical complexity is equivalent to validity. Education supports governance but cannot compensate for inaccessible evidence or unsafe design.

Limitations

The proposed architecture is constrained by the limited amount of pharmacy-specific evidence on institutional admission, withdrawal, and re-entry. Bias, dataset mismatch, and unsafe generalization remain clinical-safety concerns that a governance architecture may identify without necessarily resolving [28]. Local thresholds and responsibilities may also vary across jurisdictions, pharmacy sectors, organizational sizes, and available expertise.

Explainability cannot substitute for evidence of safety, causal validity, usefulness, or equitable performance [29]. Similarly, human oversight, regulatory authorization, and monitoring should not be treated as independent guarantees.

External portability is uncertain because algorithms may learn institution-specific signals and lose performance when transferred to another environment [30]. The architecture itself may create administrative burden, inconsistent decisions, delayed innovation, or false reassurance if implemented ceremonially. Its proposed constructs, state transitions, responsibility allocations, and evaluation measures require prospective and comparative validation before claims can be made about feasibility, effectiveness, proportionality, or patient benefit.

Conclusion

Clinical algorithms should not acquire influence over medication-use decisions merely because they are available, technically functional, externally authorized, or locally installed. Institutional permission requires a distinct governance decision connecting evidence to a defined clinical task, professional role, population, setting, action, and monitoring obligation.

The proposed Pharmacy Algorithm Gatekeeping Architecture organizes this decision through a Pharmacy Algorithm Dossier, multidisciplinary review, bounded Permission Envelope, monitored-use layer, state-transition process, and re-entry review. Its central contribution is to replace undifferentiated approval with revocable and context-specific permission. Restriction, suspension, and withdrawal become legitimate governance responses rather than exceptional technical failures, while re-entry becomes a renewed institutional judgment rather than automatic restoration.

The architecture also separates model performance from clinical usefulness, algorithmic output from medication decisions, human participation from effective oversight, and implementation from validated benefit. It requires evaluation across technical, clinical-task, human–AI, medication-safety, workflow, equity, implementation, and governance domains.

These components constitute an original conceptual synthesis, not a validated clinical standard. Their value depends on transparent evidence, identifiable accountability, enforceable boundaries, monitoring capacity, professional participation, and responsiveness to patients and local context. Future research should test whether gatekeeping improves the visibility of evidence gaps, consistency of institutional decisions, containment of emerging risks, traceability of responsibility, and appropriateness of algorithm withdrawal and re-entry. Until such evidence is available, the architecture should be treated as a structured and testable proposal for governing algorithmic influence in pharmacy practice.

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

References

1. 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
2. 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
3. Kwan JL, Lo L, Ferguson J, Goldberg H, Diaz-Martinez JP, Tomlinson G, et al. Computerised clinical decision support systems and absolute improvements in care: Meta-analysis of controlled clinical trials. BMJ. 2020;370:m3216. doi:10.1136/bmj.m3216
4. Co Z, Holmgren AJ, Classen DC, Newmark L, Seger DL, Danforth M, et al. The tradeoffs between safety and alert fatigue: Data from a national evaluation of hospital medication-related clinical decision support. J Am Med Inform Assoc. 2020;27(8):1252-8. doi:10.1093/jamia/ocaa098
5. Habli I, Lawton T, Porter Z. Artificial intelligence in health care: Accountability and safety. Bull World Health Organ. 2020;98(4):251-6. doi:10.2471/BLT.19.237487
6. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ. 2020;368:l6927. doi:10.1136/bmj.l6927
7. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: An online database. NPJ Digit Med. 2020;3:118. doi:10.1038/s41746-020-00324-0
8. Wu E, Wu K, Daneshjou R, Ouyang D, Ho DE, Zou J. How medical AI devices are evaluated: Limitations and recommendations from an analysis of FDA approvals. Nat Med. 2021;27(4):582-4. doi:10.1038/s41591-021-01312-x
9. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: A new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res. 2017;19(11):e367. doi:10.2196/jmir.8775
10. Wang A, Freeman S, Magrabi F. Governance for safe and responsible AI in healthcare organisations: A scoping review of frameworks. NPJ Digit Med. 2026;9(1):516. doi:10.1038/s41746-026-02679-2
11. Lekadir K, Frangi AF, Porras AR, Glocker B, Cintas C, Langlotz CP, et al. FUTURE-AI: International consensus guideline for trustworthy and deployable artificial intelligence in healthcare. BMJ. 2025;388:e081554. doi:10.1136/bmj-2024-081554
12. 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
13. 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. BMJ. 2022;377:e070904. doi:10.1136/bmj-2022-070904
14. 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
15. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Bossuyt PMM, et al. Calibration: The Achilles heel of predictive analytics. BMC Med. 2019;17(1):230. doi:10.1186/s12916-019-1466-7
16. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447-53. doi:10.1126/science.aax2342
17. Feng J, Phillips RV, Malenica I, Bishara A, Hubbard AE, Celi LA, et al. Clinical artificial intelligence quality improvement: Towards continual monitoring and updating of AI algorithms in healthcare. NPJ Digit Med. 2022;5(1):66. doi:10.1038/s41746-022-00611-y
18. Davis SE, Lasko TA, Chen G, Siew ED, Matheny ME. Calibration drift in regression and machine learning models for acute kidney injury. J Am Med Inform Assoc. 2017;24(6):1052-61. doi:10.1093/jamia/ocx030
19. Kore A, Abbasi Bavil E, Subasri V, Abdalla M, Fine B, Dolatabadi E, et al. Empirical data drift detection experiments on real-world medical imaging data. Nat Commun. 2024;15(1):1887. doi:10.1038/s41467-024-46142-w
20. Roschewitz M, Khara G, Yearsley J, Sharma N, James JJ, Ambrózay É, et al. Automatic correction of performance drift under acquisition shift in medical image classification. Nat Commun. 2023;14(1):6608. doi:10.1038/s41467-023-42396-y
21. Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R. Practical guidance on artificial intelligence for health-care data. Lancet Digit Health. 2019;1(4):e157-9. doi:10.1016/S2589-7500(19)30084-6
22. 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(1):23. doi:10.1038/s41746-020-0232-8
23. 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
24. Kompa B, Snoek J, Beam AL. Second opinion needed: Communicating uncertainty in medical machine learning. NPJ Digit Med. 2021;4(1):4. doi:10.1038/s41746-020-00367-3
25. 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
26. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated Consolidated Framework for Implementation Research based on user feedback. Implement Sci. 2022;17(1):75. doi:10.1186/s13012-022-01245-0
27. 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
28. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Qual Saf. 2019;28(3):231-7. doi:10.1136/bmjqs-2018-008370
29. 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):e745-50. doi:10.1016/S2589-7500(21)00208-9
30. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. PLoS Med. 2018;15(11):e1002683. doi:10.1371/journal.pmed.1002683