Original Article

Governing Artificial Intelligence after Deployment in the Medication-Use System

David Osei[1]*, Akua Afriyie[1], Kofi Adu[2]

[1]Department of AI Governance and Medication Safety, Faculty of Pharmacy, University of Ghana, Accra, Ghana. [2]Department of Post-Deployment AI Oversight, Faculty of Pharmacy, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.

Abstract

Artificial intelligence is increasingly embedded within medication-related decision support, prescription review, alerting, prioritization, pharmacovigilance, and operational workflows. Deployment, however, does not stabilize the data environment, clinical task, professional response, patient population, organizational context, or relationship between model output and medication decisions. Postdeployment governance must therefore extend beyond periodic technical performance checks. This article develops a proposed postdeployment-governance architecture for artificial intelligence used within medication-use systems. The architecture defines the governed object as the complete AI-enabled arrangement, including its model, data pipelines, interfaces, users, workflow position, organizational policies, external dependencies, and intended-use boundaries. It connects versioned baseline evidence with multidomain surveillance of performance, drift, equity, human–AI interaction, workflow consequences, medication incidents, pharmacovigilance signals, and patient experience. Detected signals enter structured triage and investigation before proportionate decisions concerning continued use, intensified monitoring, workflow correction, model modification, revalidation, restriction, suspension, rollback, or withdrawal. Accountability functions assign responsibility for evidence review, decision authority, communication, alternative workflow activation, and learning closure. Validation would require longitudinal and multisite assessment of technical, clinical, sociotechnical, safety, equity, organizational, and patient-relevant outcomes. The architecture does not establish universal thresholds, causal attribution rules, regulatory acceptability, or improved medication outcomes. Its original contribution is an integrated and testable governance structure that treats deployment as the beginning of continuing institutional responsibility rather than the endpoint of model development.

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

Introduction

Artificial intelligence can influence medication decisions after it has passed development testing, local validation, procurement review, and technical integration. At that point, responsibility does not end; it changes form. Developers and healthcare organizations must account for how a system behaves under real clinical conditions, how its outputs are interpreted, and whether its continued use remains consistent with patient benefit and avoidance of harm [1].

This obligation cannot be discharged by reporting model accuracy alone because clinical impact depends on workflow integration, available alternatives, professional behaviour, data quality, implementation context, and the consequences of acting or failing to act on an output [2].

The relevant unit of analysis is therefore not an isolated algorithm. Clinical decision support operates through interconnected data sources, inference processes, interfaces, users, organizational rules, and clinical workflows [3]. A medication-related model may retain its original software code while its meaning changes because formularies, prescribing patterns, patient populations, interfaces, staffing, treatment pathways, or documentation practices have changed. Evaluation must accordingly address transparency, reproducibility, ethics, effectiveness, and the conditions under which a model is used rather than treating technical discrimination as a complete account of performance [4].

Implementation also remains dynamic after adoption. Technologies may be locally modified, incompletely used, abandoned, scaled beyond their initial context, or sustained

Address for correspondence: David Osei, Department of AI Governance and Medication Safety, Faculty of Pharmacy, University of Ghana, Accra, Ghana.
david.osei@ug.edu.gh
Received: 05 February 2025; **Accepted:** 02 June 2025

How to cite this article: Osei D, Afriyie A, Adu K. Governing Artificial Intelligence after Deployment in the Medication-Use System. Arch Pharm Pract. 2025;16(2):62-71. https://doi.org/10.51847/hPZnbLGKmC

despite weakening evidence of benefit. These trajectories are shaped by the complexity of the technology, the condition or task, the organization, the wider system, and adaptation over time [5]. In medication-use systems, such complexity is amplified because a model output may affect prescribing, pharmacist verification, dispensing, administration, monitoring, patient counselling, or pharmacovigilance, with responsibility distributed across several professional and organizational boundaries.

This article proposes a postdeployment-governance architecture for medication-use AI. It distinguishes model performance from clinical-task performance, workflow consequence, medication safety, equity, and governance performance. The proposed contribution integrates system definition, multidomain surveillance, incident investigation, controlled modification, revalidation, restriction, withdrawal, accountability, and organizational learning. It is a conceptual architecture rather than a validated clinical standard. Its decision rules, relationships, and governance mechanisms require prospective evaluation within specific medication tasks and organizations.

Postdeployment Risks in Medication-Use Systems

Medication-related AI can support alert optimization and help distinguish potentially relevant warnings from large volumes of low-value notifications. Existing evaluations nevertheless remain heterogeneous and frequently provide limited external validation, transparency, or evidence connecting alert reduction with improved clinical outcomes [6]. Machine learning has also been used to prioritize prescriptions with a higher probability of medication error for pharmacist review [7]. Such prioritization may improve allocation of professional attention, but it does not establish that errors are prevented, that important prescriptions are never deprioritized, or that the same performance will persist after changes in patients, medicines, staffing, or workflow.

Postdeployment risk can emerge even when the underlying model remains unchanged. Repeated alerts, high workload, and workflow complexity can reduce clinician responsiveness to decision support [8]. Conversely, changes in clinical practice may alter relationships learned during development. A model may therefore become unreliable because its inputs, target, intervention context, or surrounding care pathway has shifted; in some circumstances, this may justify disabling or decommissioning the system rather than attempting to preserve routine use [9].

Human oversight is not an automatic safeguard. Clinicians can be influenced by erroneous AI recommendations, particularly when the advice appears authoritative or is difficult to independently verify [10]. The relevant question is consequently not whether a professional remains nominally involved, but whether the workflow preserves meaningful opportunity, information, time, authority, and competence to challenge the output. Under-reliance may also forfeit useful support, while excessive verification demands may transfer work rather than reduce it.

The proposed risk model distinguishes six interacting classes. Input and environmental risk concerns changes in data, populations, medicines, coding, interfaces, and clinical practice. Model-behaviour risk concerns deterioration in calibration, discrimination, reliability, or uncertainty. Equity risk concerns differential performance, benefit, burden, or visibility across patient groups. Human–AI interaction risk includes automation bias, under-reliance, alert fatigue, workarounds, and loss of situation awareness. Medication-workflow risk concerns missed review, inappropriate prioritization, interruption burden, delayed intervention, or responsibility gaps. Governance risk concerns unclear ownership, inadequate escalation, undocumented modification, weak vendor control, or inability to restrict and withdraw the system. These categories are proposed as an organizing synthesis; they do not establish the prevalence or causal contribution of any risk in a particular organization.

Table 1 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for definitions, components, relationships, and intended uses for the article Governing Artificial Intelligence after Deployment in the Medication-Use System.

Table 1. Definitions, components, relationships, and intended uses for the article Governing Artificial Intelligence after Deployment in the Medication-Use System.

Component or scientific dimension	Problem addressed	Inputs or determinants	Proposed mechanism or relationship	Expected contribution	Evidence required	Failure or uncertainty risk	Boundary statement
Governed AI-enabled arrangement	Model-only oversight omits workflow and organizational determinants.	Model version, data pipeline, interface, users, policies, vendor dependencies, medication-use stage	**Proposed:** governance applies to the complete sociotechnical arrangement rather than only the algorithm. CDS evidence supports a systems interpretation [3].	Defines the object subject to surveillance, change control, and accountability.	Versioned technical and workflow documentation; named owners; intended-use statement	Components may change independently or remain undocumented.	A complete inventory does not establish safety or effectiveness.

Intended-use and exclusion boundary	Outputs may be applied to patients, tasks, or settings not represented in evaluation.	Population, medication task, setting, user, comparator, exclusions, required human verification	**Proposed:** surveillance and action thresholds are conditional on an explicit operational boundary.	Makes out-of-scope use observable and governable.	Local use records; population and task comparison; exception reporting	Informal expansion of use may occur without formal model change.	Intended use is an organizational control, not proof of validity.
Postdeployment signal	Isolated metrics may not reveal clinically consequential change.	Performance measures, drift indicators, overrides, incidents, complaints, pharmacovigilance data	**Proposed:** a signal is any credible deviation suggesting altered benefit, burden, safety, or equity. Medication-alert evidence demonstrates the need to interpret intermediate metrics cautiously [6].	Creates an entry point for triage and investigation.	Baseline comparator; exposure data; signal provenance; clinical context	False signals, delayed signals, and unmeasured harms are possible.	A signal does not establish causality or mandate one response.
Human–AI collaboration	Nominal human involvement may conceal automation bias or unmanageable verification demands.	Output presentation, explanation, workload, expertise, authority, time, alternative information	**Proposed:** human oversight is evaluated as an observable workflow capability. Susceptibility to erroneous AI advice supports this concern [10].	Connects model use with professional judgment and responsibility.	Observation of acceptance, override, verification, workarounds, and escalation	Excessive reliance, systematic rejection, or responsibility displacement	Human presence alone does not establish meaningful control.
Medication-safety consequence	Model and task metrics may not represent effects on medication processes or patients.	Errors, near misses, delays, omissions, monitoring failures, adverse-event signals	**Proposed:** technical and workflow evidence is linked to medication-safety investigation.	Prevents governance from ending at model-performance dashboards.	Incident review, process tracing, pharmacovigilance, patient-relevant evidence	Events may be underreported or incorrectly attributed.	Absence of detected incidents does not prove absence of harm.
Governance decision	Organizations may recognize a problem without having authority or criteria to act.	Severity, likelihood, scope, equity, reversibility, uncertainty, safe alternatives	**Proposed:** evidence is translated into proportionate continuation, correction, revalidation, restriction, suspension, rollback, or withdrawal.	Connects surveillance with accountable action.	Decision record; assigned authority; rationale; alternative workflow	Delayed action, inconsistent escalation, or commercially influenced decisions	Decision levels require local validation and lawful authority.
Learning closure	Corrective actions may be implemented without checking completion or residual risk.	Action status, communication, revalidation, follow-up monitoring, unresolved issues	**Proposed:** incidents close only after action, communication, residual-risk review, and retained learning are documented.	Supports institutional memory and future prevention.	Closure criteria; follow-up evidence; updated training, contracts, and baselines	Administrative closure may substitute for substantive resolution.	Process completion does not demonstrate improved patient outcomes.

Proposed Governance Architecture

Postdeployment governance should begin with an inventory of every AI-enabled arrangement that can materially influence medication use. Algorithmovigilance provides a basis for longitudinal examination of clinical effectiveness and equity rather than one-time technical assessment [11]. Model maintenance similarly requires a 360-degree view spanning technical behaviour, clinical relevance, and operational conditions [12]. The proposed architecture extends these principles by linking each system to its medication-use task, responsible owner, intended users, affected patient groups, local interfaces, required professional verification, safe alternative workflow, and authority for restriction or withdrawal.

The architecture contains seven connected functions. First, a governed inventory and baseline dossier records the operational system and its intended-use boundaries. Second, multidomain surveillance observes technical performance, drift, subgroup effects, workflow behaviour, medication incidents, pharmacovigilance signals, and patient feedback. Third, signal triage determines credibility, severity, scope, urgency, and whether immediate containment is necessary. Fourth, investigation examines model, data, workflow, human, medication-process, and organizational explanations. Fifth, change control and revalidation govern modifications before unrestricted return to use. Sixth, restriction and withdrawal authority enables bounded use, suspension, rollback, or retirement. Seventh, accountability and learning closure require documented decisions, communication,

follow-up, and incorporation of lessons into future monitoring and procurement. Recommendations for AI-enabled decision support support explicit validation, monitoring, incident reporting, documentation, and user preparation [13]. Trustworthiness also requires distinctions among explainability, transparency, usability, and reliability rather than treating any one property as sufficient [14].

Figure 1 presents the original conceptual synthesis for postdeployment governance architecture for medication-use AI, showing how the article's principal components, evidence relationships, uncertainties, and decision boundaries are connected.

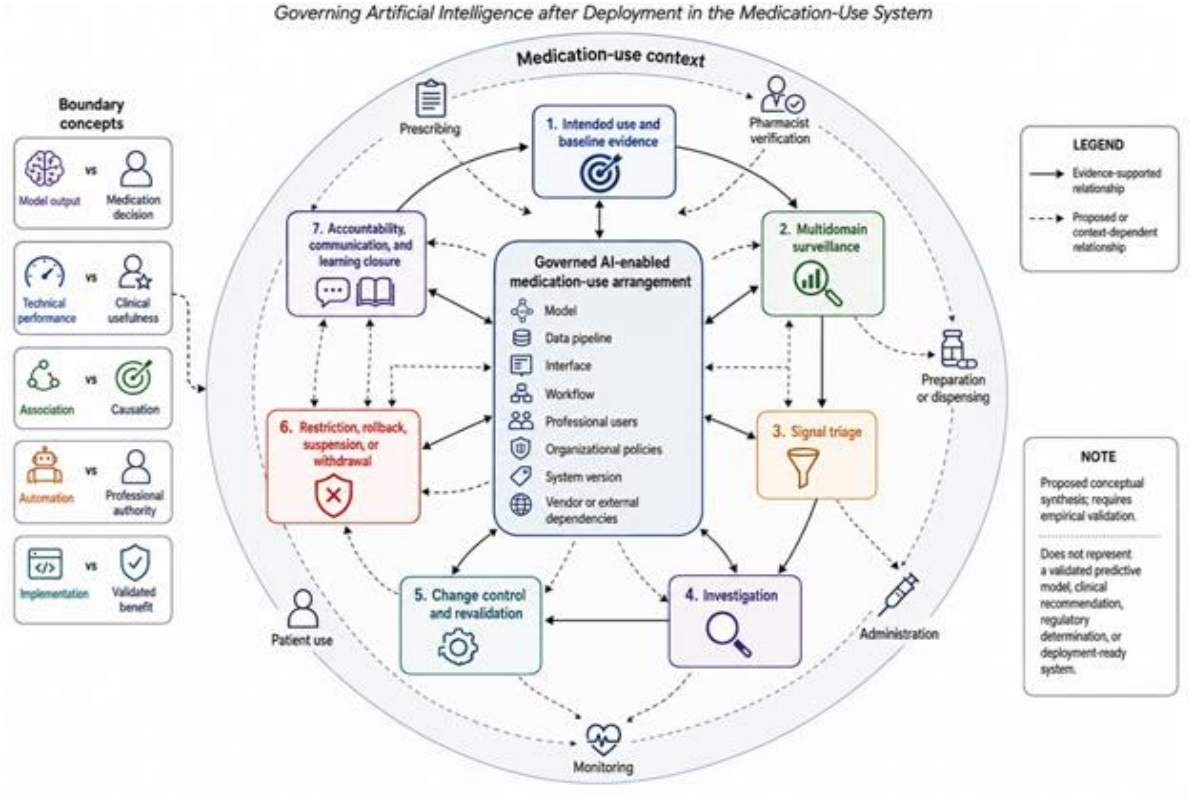

Figure 1. Postdeployment Governance Architecture for Medication-Use AI

A versioned baseline dossier is necessary because revalidation requires knowledge of what changed and what remained constant. Standardized modelling frameworks demonstrate the value of reproducible data definitions, analytical procedures, model specifications, and evaluation records [15]. The proposed dossier would additionally document medication-task purpose, clinical comparator, user responsibilities, subgroup expectations, interface design, upstream dependencies, known exclusions, surveillance plan, incident pathway, modification history, and rollback arrangements. Its purpose is not administrative completeness for its own sake; it is to preserve the evidence needed to determine whether an observed problem arises from the model, data, workflow, environment, or interaction among them.

Information should flow bidirectionally. Operational use generates performance, workflow, safety, equity, and patient signals. Investigation may revise the interpretation of those signals, identify previously unrecognized scope conditions, and require changes to training, interfaces, contracts, or monitoring. Modification returns the system to a revalidation gate rather than directly to unrestricted use. Restriction or withdrawal activates communication and a safe alternative medication workflow. Every pathway ends in documented learning closure. This architecture is proposed as a coordination mechanism; it does not establish the effectiveness of any component or the sufficiency of the combined system.

Table 2 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for evaluation criteria, evidence requirements, and failure modes for the article Governing Artificial Intelligence after Deployment in the Medication-Use System.

Table 2. Evaluation criteria, evidence requirements, and failure modes for the article Governing Artificial Intelligence after Deployment in the Medication-Use System.

Architecture layer or process stage	Core function	Information flow	Interaction with other components	Evaluation criterion	Potential failure mode	Human or organizational responsibility	Readiness boundary
Governed inventory and baseline	Define the operational system, version, intended use, exclusions, users, and dependencies.	Development and local-validation evidence enter a versioned dossier.	Configures monitoring, investigation, and revalidation. Reproducible model documentation supports this function [15].	Completeness, currency, traceability, and correspondence with actual use	Undocumented systems, hidden dependencies, or divergence between recorded and operational versions	Model owner and medication-process owner jointly verify the record.	Unrestricted use is not supported when the operational system cannot be identified.
Multidomain surveillance	Detect technical, equity, workflow, safety, and patient-relevant change.	Operational data, incidents, overrides, reports, and feedback enter monitored domains.	Supplies signals to triage; receives revised baselines after investigation. Algorithmovigilance supports longitudinal effectiveness and equity review [11].	Coverage, timeliness, stratification, interpretability, and linkage to exposure	Stable aggregate metrics conceal subgroup or workflow deterioration.	Technical teams, pharmacy services, safety functions, and patient-representation mechanisms contribute evidence.	Monitoring readiness does not exist when consequential domains cannot be observed.
Signal triage	Determine credibility, urgency, potential scope, and need for containment.	Signals are combined with exposure, severity, reversibility, and uncertainty information.	Directs routine review or immediate containment and investigation.	Documented rationale, response time, and consistency	Dismissal of weak early signals or escalation of non-consequential variation	Designated triage authority with clinical, technical, safety, and equity competence	A signal alone does not justify causal attribution or a predetermined action.
Investigation	Examine technical, clinical, sociotechnical, and organizational explanations.	Version, data, subgroup, workflow, user, incident, and contextual evidence are integrated.	Informs change classification, restriction, or continuation.	Scope adequacy, competing explanations, affected-population identification, and uncertainty disclosure	Investigation focuses on model code while omitting workflow or environmental causes.	Multidisciplinary investigation lead with access to vendor and local evidence	Closure is premature when material explanations or affected groups remain unevaluated.
Change control and revalidation	Govern modifications before unrestricted return to use.	Proposed changes are linked to rationale, version control, testing, and approval evidence.	Receives investigation findings and returns revised baselines to surveillance. Lifecycle maintenance requires technical, clinical, and operational review [12].	Traceability, proportional testing, independent challenge, and residual-risk assessment	Silent updates, inadequate local testing, or evaluation limited to the changed code	Authorized change body; model owner supplies evidence but does not unilaterally approve high-risk changes.	Reporting and testing completeness do not by themselves establish acceptable decision use.
Restriction, rollback, suspension, or withdrawal	Contain risk when continued unrestricted use is not supportable.	Decision rationale and affected versions are communicated to users and system operators.	Activates alternative workflow, residual-risk monitoring, and learning closure.	Timeliness, enforceability, communication reach, and safe alternative activation	System remains accessible, obsolete outputs persist, or frontline responsibility is unclear	An organization must preassign authority to restrict use independently of commercial ownership.	No universal threshold is proposed; action remains context- and risk-dependent.
Accountability and learning closure	Confirm action completion, communication, follow-up, and institutional learning.	Decision records and follow-up evidence update baselines, training, contracts, and procurement.	Closes the loop while preserving unresolved uncertainty for continued monitoring. Governance recommendations support explicit reporting and accountability functions [13].	Completion, unresolved-signal age, residual-risk review, and documented transfer of learning	Administrative closure without verification of operational change	Governing body confirms closure; process owners implement and report corrective actions.	Process performance does not demonstrate improved medication or patient outcomes.

Performance, Drift, Equity, and Incident Surveillance

Postdeployment surveillance should begin by separating constructs that are often collapsed into a single performance score. Discrimination concerns ordering or classification, whereas calibration concerns correspondence between predicted probabilities and observed outcomes. A model can retain acceptable discrimination while producing probabilities that are unsuitable for clinical decisions [16].

Surveillance should also examine missingness, input validity, pipeline availability, latency, uncertainty, and changes in data distributions. Empirical drift-detection work shows that real-world medical data can change over time and that different detection methods identify different forms of change [17]. A detected distributional change is therefore a reason for review, not proof that clinical performance has deteriorated.

Equity surveillance requires an independent analytical pathway. Population-average performance may remain stable while error rates, calibration, access to benefit, or exposure to burden change for particular groups. Fairness can drift longitudinally, and updating a model for overall performance may reduce, preserve, or worsen subgroup disparities [18]. Inequity can also originate from the selected target or proxy. A model may reproduce unequal access to care when expenditure, utilization, or another structurally patterned variable is treated as a substitute for clinical need [19]. Postdeployment equity review should consequently examine data availability, target meaning, subgroup calibration, false-positive and false-negative consequences, workflow access, professional response, and whether affected patients can challenge or understand the resulting decision. No single fairness metric can determine whether a medication-use arrangement is equitable.

Medication-safety surveillance should connect model and workflow evidence with incidents, near misses, pharmacovigilance signals, professional concerns, and patient reports. Artificial intelligence can assist pharmacovigilance processes, but representative data, validation, quality controls, expert review, and retained human responsibility remain necessary [20]. Automated signal generation may increase sensitivity while also increasing false-positive burden or creating new opportunities for unrecognized failure. Conversely, incident-reporting systems may underrepresent harms that are difficult to detect, distributed across several medication-use stages, or normalized within routine workflow.

The proposed surveillance model therefore contains five linked domains. Technical surveillance examines availability, data integrity, discrimination, calibration, uncertainty, and drift. Equity surveillance examines subgroup performance, proxy validity, differential access, and distribution of burdens and benefits. Human–AI surveillance examines acceptance, override, independent verification, workarounds, alert burden, and professional understanding of scope. Medication-process surveillance examines prioritization, delays, omitted review, inappropriate intervention, and effects across prescribing, verification, dispensing, administration, monitoring, and patient use. Incident surveillance integrates medication incidents, near misses, pharmacovigilance, complaints, patient feedback, and contextual investigation.

Signals should be triangulated rather than interpreted independently. A calibration change accompanied by stable workflow and safety evidence may justify investigation without immediate restriction. A modest technical change accompanied by concentrated subgroup harm, unsafe clinician reliance, or a serious medication incident may require containment despite apparently acceptable aggregate performance. Conversely, an increase in incident reports may reflect improved reporting rather than worsening system performance. The proposed approach therefore requires exposure data, baseline comparators, competing explanations, affected-version identification, and explicit uncertainty.

Surveillance is successful only when it can support action. Measures should include not merely whether a metric is available, but whether relevant populations and medication stages are covered, whether signals reach an accountable reviewer, whether investigations are completed, and whether unresolved concerns remain visible. The architecture does not prescribe universal monitoring intervals or numerical alarm thresholds. Such parameters must be validated for the medication task, severity of plausible harm, reversibility of decisions, availability of alternatives, organizational capacity, and consequences of false reassurance or unnecessary interruption.

Change Control, Revalidation, Restriction, and Withdrawal

Postdeployment change control begins when surveillance, an incident, a workflow concern, or an external change creates reasonable doubt about continued use. Continual quality improvement for clinical AI requires repeated measurement and controlled updating rather than indefinite reliance on the original validation state [21]. Statistical surveillance can identify calibration deterioration and inform whether review or updating is warranted [22]. A detected change, however, does not independently determine whether the model should be modified. The organization must first establish the affected version, exposed patients and medication processes, likely mechanism, clinical significance, subgroup consequences, and uncertainty.

Revalidation should extend beyond retraining or technical testing. A medical algorithmic audit encompasses data, model behaviour, intended task, users, human–AI interaction, and the deployment environment [23]. For medication-use AI, the proposed revalidation scope additionally includes effects on pharmacist prioritization, prescribing or verification burden, override behaviour, escalation pathways, safe alternatives, and downstream medication processes. The extent of revalidation should be proportionate to the change, but seemingly minor modifications may require broader review when they alter output interpretation, patient eligibility, workflow position, or professional responsibility.

A versioned change dossier should document the reason for modification, affected components, intended-use implications, development data, subgroup assessment, comparison with the operational baseline, local testing,

unresolved uncertainty, authorization, and monitoring plan. Transparent reporting of intended use, model development, evaluation, updating, and limitations provides an evidentiary foundation for this review [24]. Documentation completeness nevertheless does not establish acceptable performance or justify unrestricted use.

The proposed response pathway contains seven possible states: continued use with documented rationale; enhanced monitoring; workflow correction without model modification; bounded technical change followed by revalidation; restricted use; temporary suspension or rollback; and withdrawal. A credible postdeployment signal should enter a controlled response process that combines continual monitoring, drift-informed review, and multidimensional audit before unrestricted use is restored [21-23]. Selection among states should consider plausible harm, affected scope, urgency, reversibility, detectability, subgroup impact, uncertainty, and availability of a safe alternative.

Restriction may limit the system to specified users, populations, medicines, settings, or advisory functions. Suspension temporarily removes operational influence while investigation or correction proceeds. Rollback restores a previously governed version when its data and workflow dependencies remain valid. Withdrawal ends authorized use when the intended function is no longer supportable, risk cannot be adequately controlled, or organizational capacity for safe oversight is absent. These actions require preassigned authority and technical means to prevent obsolete outputs from remaining accessible. The proposed pathway structures deliberation but does not supply universal thresholds or replace legal, regulatory, professional, or organizational decision authority.

Accountability and Governance Performance Measures

Postdeployment accountability concerns who must observe, interpret, decide, act, communicate, and verify closure. Early clinical evaluation frameworks emphasize examination of actual use, workflow integration, human–AI interaction, and emerging safety concerns rather than model performance in isolation [25]. Organizational governance must therefore distribute responsibility while preserving an identifiable decision authority. Governance models for healthcare AI support explicit oversight, transparency, accountability, and monitoring functions [26].

The proposed architecture assigns joint responsibility to a model owner and a medication-process owner. The model owner maintains version identity, technical evidence, dependencies, and change documentation. The medication-process owner determines whether outputs remain appropriate for the clinical task and workflow. Pharmacy leadership, safety and quality functions, information-technology teams, frontline professionals, patient-representation functions, vendors, and independent reviewers contribute distinct evidence and authority. Ethical concerns may arise across data acquisition, system design, professional practice, access, patient autonomy, and distribution of benefit and harm [27]. Principles alone are insufficient when organizations lack procedures, incentives, enforcement mechanisms, and authority to resolve conflicts [28].

Governance performance should therefore be measured independently from model performance. Proposed indicators include surveillance coverage, signal-detection latency, triage latency, unresolved-signal age, completion of investigations, subgroup-monitoring coverage, change-dossier completeness, implementation of restrictive decisions, activation of alternative workflows, communication reach, and confirmation of learning closure. These are process indicators rather than validated safety surrogates. Improvement in governance-process measures would not independently demonstrate fewer medication errors or better patient outcomes.

Accountability should follow authority. Frontline pharmacists and clinicians should report concerns, verify outputs within feasible professional limits, and use escalation pathways, but they should not bear sole responsibility for defects in data pipelines, model design, interface configuration, or organizational controls. Vendors should provide version, update, limitation, and corrective-action information, while the deploying organization retains responsibility for local use. Independent review becomes especially important when commercial, operational, or reputational incentives may favour continued deployment despite unresolved uncertainty.

Implementation across Medication-Use Organizations

Implementation should be adapted to organizational context rather than treated as installation of a finished technical product. The updated Consolidated Framework for Implementation Research identifies interacting influences related to the intervention, outer setting, inner setting, individuals, and implementation process [29]. Empirical implementation research also identifies evidence quality, resources, leadership, professional acceptance, technical integration, and workflow fit as potential facilitators or barriers [30]. These factors determine whether surveillance information can be generated, interpreted, and converted into action.

An end-to-end implementation pathway should cover selection, procurement, local assessment, workflow design, user preparation, controlled activation, postdeployment monitoring, incident management, modification, and retirement [31]. **Figure 2** presents the original conceptual synthesis for drift, incident, change-control, restriction, and withdrawal pathway, showing how the article's principal

components, evidence relationships, uncertainties, and decision boundaries are connected.

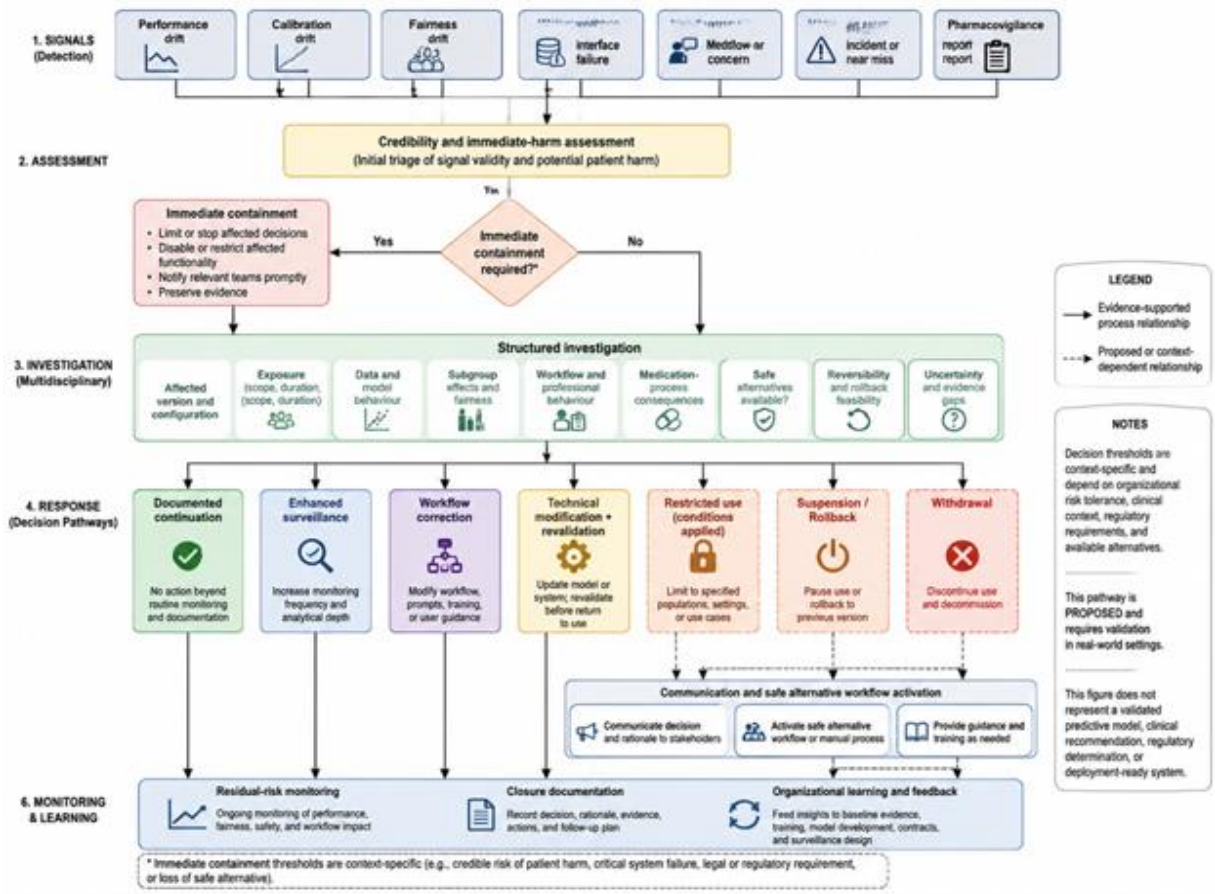

Figure 2. Drift, Incident, Change-Control, Restriction, and Withdrawal Pathway

Sustainable implementation also requires digital infrastructure, leadership, workforce capability, evaluation capacity, and feedback mechanisms compatible with a learning health system [32]. Implementation capacity depends jointly on organizational context, an end-to-end implementation process, and infrastructure for continuing learning [29, 31, 32]. A small organization may rely on shared or external technical resources, but it must still preserve local authority over medication workflow, incident escalation, restriction, and safe alternative processes.

Figure 2 should therefore be operationalized before activation rather than constructed after an incident. Organizations should identify how the AI function can be technically disabled, how affected users will be informed, how medication work will continue, and how patients exposed to an identified failure will be assessed. Implementation should not proceed when the organization cannot observe consequential risks, investigate credible signals, control system versions, or suspend use.

Limitations

The architecture is constrained by the quality of the underlying clinical-AI evidence. Comparisons between AI and clinicians have frequently contained design, reporting, and interpretive weaknesses that limit strong clinical claims [33]. Machine-learning decision support also does not consistently improve clinician performance and may help or hinder depending on the task, user, interface, and evaluation setting [34].

Generalisability cannot be presumed across organizations, populations, medication tasks, professional roles, or time [35]. Explanation methods may assist inspection or communication but do not establish causality, safety, fairness, or appropriate decision use [36]. Validation must therefore address variable human–AI effects, limited transferability, and the inability of explanation outputs alone to establish safe decision use [34-36].

The proposed architecture draws partly on evidence from diagnostic, imaging, and general clinical-AI settings because medication-specific postdeployment research remains limited. It does not provide validated thresholds, causal attribution methods, legal interpretations, economic

estimates, or proof that governance processes improve medication outcomes. Local empirical evaluation remains necessary.

Conclusion

Deployment begins a new evidentiary and organizational phase for artificial intelligence in medication-use systems. The model, its data, interface, users, workflow position, intended-use boundary, and institutional dependencies continue to change or acquire new meaning after implementation. Governance must therefore examine more than technical accuracy and more than the model in isolation.

The proposed architecture connects a governed system inventory and baseline dossier with surveillance of performance, drift, equity, human–AI interaction, medication processes, incidents, pharmacovigilance, and patient experience. It then connects credible signals to investigation, controlled modification, revalidation, restriction, suspension, rollback, or withdrawal. Accountability is treated as an operational distribution of information, authority, action, communication, and learning rather than as a general ethical commitment.

The architecture also preserves several boundaries. Stable aggregate performance does not establish stable safety or equity. Human involvement does not guarantee meaningful oversight. Documentation does not prove clinical validity. Monitoring does not establish causality. Implementation does not demonstrate benefit, and explainability does not authorize decision use. Restriction and withdrawal must remain available when the evidence supporting continued use is inadequate.

The contribution is a proposed and testable postdeployment-governance structure for medication-use AI. Its value should be examined through longitudinal, multisite, and task-specific research that evaluates technical behaviour, professional work, organizational response, equity, medication safety, patient consequences, and governance performance. Until such evidence is available, the architecture should be interpreted as a framework for disciplined inquiry and accountable decision-making rather than a validated standard or deployment-ready system.

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

References

1. 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
2. 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
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(1):17. doi:10.1038/s41746-020-0221-y
4. 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. doi:10.1136/bmj.l6927
5. 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). doi:10.2196/jmir.8775
6. 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
7. 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
8. Ancker JS, Edwards A, Nosal S, Hauser D, Mauer E, Kaushal R. 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
9. Finlayson SG, Subbaswamy A, Singh K, Bowers J, Kupke A, Zittrain J, et al. The clinician and dataset shift in artificial intelligence. N Engl J Med. 2021;385(3):283-6. doi:10.1056/NEJMc2104626
10. 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
11. Embi PJ. Algorithmovigilance—advancing methods to analyze and monitor artificial intelligence-driven health care for effectiveness and equity. JAMA Netw Open. 2021;4(4). doi:10.1001/jamanetworkopen.2021.4622
12. Davis SE, Embi PJ, Matheny ME. Sustainable deployment of clinical prediction tools—a 360° approach to model maintenance. J Am Med Inform Assoc. 2024;31(5):1195-8. doi:10.1093/jamia/ocae036
13. 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
14. Cutillo CM, Sharma KR, Foschini L, Kundu S, Mackintosh M, Mandl KD, et al. Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency. NPJ Digit Med. 2020;3(1):47. doi:10.1038/s41746-020-0254-2
15. Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. J Am Med Inform Assoc. 2018;25(8):969-75. doi:10.1093/jamia/ocy032
16. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: The Achilles heel of predictive analytics. BMC Med. 2019;17(1):230. doi:10.1186/s12916-019-1466-7
17. 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
18. Davis SE, Dorn C, Park DJ, Matheny ME. Emerging algorithmic bias: Fairness drift as the next dimension of model maintenance and sustainability. J Am Med Inform Assoc. 2025;32(5):845-54. doi:10.1093/jamia/ocaf039
19. 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
20. Kassekert R, Grabowski N, Lorenz D, Schaffer C, Kempf D, Roy P, et al. Industry perspective on artificial intelligence/machine learning in pharmacovigilance. Drug Saf. 2022;45(5):439-48. doi:10.1007/s40264-022-01164-5
21. 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
22. Davis SE, Greevy RA Jr, Lasko TA, Walsh CG, Matheny ME. Detection of calibration drift in clinical prediction models to inform model updating. J Biomed Inform. 2020;112:103611. doi:10.1016/j.jbi.2020.103611
23. Liu X, Glocker B, McCradden MM, Ghassemi M, Denniston AK, Oakden-Rayner L. The medical algorithmic audit. Lancet Digit Health. 2022;4(5). doi:10.1016/S2589-7500(22)00003-6
24. 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
25. 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
26. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. J Am Med Inform Assoc. 2020;27(3):491-7. doi:10.1093/jamia/ocz192
27. Morley J, Machado CCV, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: A mapping review. Soc Sci Med. 2020;260:113172. doi:10.1016/j.socscimed.2020.113172
28. Mittelstadt B. Principles alone cannot guarantee ethical AI. Nat Mach Intell. 2019;1(11):501-7. doi:10.1038/s42256-019-0114-4
29. 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
30. Strohm L, Hehakaya C, Ranschaert ER, Boon WPC, Moors EHM. Implementation of artificial intelligence applications in radiology: Hindering and facilitating factors. Eur Radiol. 2020;30(10):5525-32. doi:10.1007/s00330-020-06946-y
31. van der Vegt AH, Scott IA, Dermawan K, Schnetler RJ, Kalke VR, Lane T. Implementation frameworks for end-to-end clinical AI: Derivation of the SALIENT framework. J Am Med Inform Assoc. 2023;30(9):1503-15. doi:10.1093/jamia/ocad088
32. Kamel Rahimi A, Pienaar O, Ghadimi M, Canfell OJ, Pole JD, Shrapnel S, et al. Implementing AI in hospitals to achieve a learning health system: Systematic review of current enablers and barriers. J Med Internet Res. 2024;26. doi:10.2196/49655
33. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: Systematic review of design, reporting standards, and claims of deep learning studies. BMJ. 2020;368. doi:10.1136/bmj.m689
34. Vasey B, Ursprung S, Beddoe B, Taylor EH, Maruthappu M, Nagendran M, et al. Association of clinician diagnostic performance with machine learning-based decision support systems: A systematic review. JAMA Netw Open. 2021;4(3). doi:10.1001/jamanetworkopen.2021.1276
35. Futoma J, Simons M, Panch T, Doshi-Velez F, Celi LA. The myth of generalisability in clinical research and machine learning in health care. Lancet Digit Health. 2020;2(9). doi:10.1016/S2589-7500(20)30186-2
36. 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