Risk Management and Healthcare Policy

Dovepress
Taylor & Francis Group

Open Access Full Text Article

PERSPECTIVES

Feasibility-First Risk Management for Clinical AI: A Deterministic Systems Intelligence Framework for Release, Monitoring, Rollback and Withholding

Adegoke O Adefolalu

Practice of Medicine and Clinical Integrated Programmes, School of Medicine, Sefako Makgatho Health Sciences University, Pretoria, Gauteng, South Africa

Correspondence: Adegoke O Adefolalu, Email Adegoke.adefolalu@smu.ac.za

Abstract: Artificial intelligence (AI) is increasingly used to predict deterioration, classify risk, prioritise workload, support diagnosis, tailor communication and monitor patients across healthcare settings. Existing safety discussions rightly emphasise model performance, bias, explainability, regulatory approval and post-deployment monitoring. However, these domains do not fully answer a prior risk-management question: when should an AI output be released into clinical or organisational action under the active healthcare regime? This Perspective develops a Deterministic Systems Intelligence (DSI) framework for feasibility-first clinical AI risk management. The proposed admissible-state layer evaluates candidate AI outputs against data fitness, population fit, clinical actionability, workflow capacity, equity, authority, monitoring and reversibility before action is permitted. It classifies outputs into five governance states: release, restricted release, active monitoring, rollback and HOLD. HOLD denotes disciplined non-release when evidence, feasibility or safeguards are insufficient. A worked deterioration-alert example shows how the same technically plausible output may be released, restricted, monitored, rolled back or held depending on local capacity, equity and safety controls. The framework complements reporting, audit, regulatory and algorithmovigilance approaches by inserting an explicit admissibility step between AI output and healthcare action. Responsible clinical AI therefore requires not only prediction, but release readiness, monitoring and the capacity to withhold.

Keywords: artificial intelligence, healthcare risk management, deterministic systems intelligence, admissible-state reasoning, clinical decision support, algorithmovigilance

Introduction: Healthcare AI Risk is a Release Problem

Artificial intelligence (AI) is increasingly positioned as a risk-management instrument in healthcare. It is used or proposed for diagnostic support, deterioration prediction, triage, remote monitoring, workflow prioritisation, population segmentation, adverse-event detection and clinical communication. In each setting, AI promises earlier recognition of risk and more timely action. The attraction is clear: a system that recognises deterioration, diagnostic inconsistency or operational strain earlier than conventional workflows could reduce preventable harm and improve the allocation of scarce clinical resources.[1–3]

Yet the use of AI also creates a distinctive safety problem. The immediate question is not only whether an algorithm can predict, classify or recommend; it is whether the output should be released into clinical or organisational action under the current conditions of care.[4,5] Healthcare AI outputs do not operate in a vacuum. They enter institutions with variable data quality, uneven staffing, different clinical pathways, unequal patient populations, competing alerts, regulatory obligations, liability concerns and differing capacity to monitor performance after deployment.[6–8] Under such conditions, a technically plausible output may still be unsafe, inequitable or operationally non-implementable.

https://doi.org/10.2147/RMHP.S621924
Received: 4 May 2026
Accepted: 17 June 2026
Published: 16 July 2026

Current AI governance frameworks rightly emphasise ethics, transparency, human oversight, fairness, performance evaluation, post-deployment monitoring and regulatory preparedness.[1–3] Reporting, trial and evaluation guidance have improved the discipline of AI assessment.[4,5] However, implementation experience has repeatedly shown that a model can be technically credible yet difficult to translate into healthcare delivery when workflow, local context, safety ownership and organisational adoption are weak.[6–8] A practical risk-management gap remains at the point where AI output becomes an action signal. Many governance approaches assume that if a model has been validated, monitored and approved, its outputs can be operationalised whenever they are produced. In real healthcare systems, this assumption is too permissive.

Concrete examples illustrate the problem. Widely implemented sepsis or deterioration prediction systems can generate plausible alerts while creating workflow burden, alert fatigue, uncertainty about escalation responsibility or poor transportability to local patient populations.[9] Diagnostic support systems may be useful in specialist pathways but unsafe where confirmatory testing, professional accountability or follow-up infrastructure are absent. Remote-monitoring alerts may be clinically meaningful but inadmissible if the response team cannot contact or review patients within the necessary timeframe. In each example, the risk-management failure is not only predictive error; it is premature release into an unsupported action regime.

This Perspective develops a Deterministic Systems Intelligence (DSI) framework for feasibility-first clinical AI risk management. The argument is deliberately narrow. DSI is not proposed as a replacement for statistical learning, validation, regulation, algorithmic audit or clinical judgement. Its contribution is to insert an admissible-state layer between AI output and healthcare action. This layer asks whether the output is actionable, authorised, equitable, auditable, monitorable and reversible under the active healthcare regime. The resulting decision is not binary. Outputs may be released, restricted, placed under active monitoring, rolled back or held.

Why Prediction is Not Risk Control

Healthcare AI risk is often discussed as a model-performance problem. Poor discrimination, poor calibration, biased training data, lack of external validation, distribution shift, inadequate reporting and limited transparency can all create harm.[10–14] These risks are real and must remain central to responsible AI development. Nevertheless, model performance is not equivalent to risk control. Risk control concerns what happens after an output is produced: whether it is acted on, by whom, under which authority, with which safeguards and through what monitoring and reversal pathway.

The difference is visible in ordinary clinical decision support. A deterioration score may be statistically sound but increase risk if it is released into a ward with no escalation team available, if it duplicates existing alerts and contributes to alarm fatigue, or if it is trained on populations that differ from those currently served. A diagnostic suggestion may be useful in a specialist setting but unsafe in a primary-care pathway without confirmatory testing or clear responsibility for follow-up. A remote-monitoring alert may be clinically meaningful but inadmissible if the response infrastructure cannot contact patients in time. In each example, the risk-management failure is not simply predictive error. It is premature release into an unsupported action regime.

This distinction matters because AI can convert uncertainty into apparent authority. A dashboard, alert or recommendation can make a probabilistic output appear operationally ready even when the system is not ready to use it safely. This is especially problematic in high-pressure clinical environments, where automation bias, alert fatigue, workflow fragmentation and institutional liability all shape how outputs are interpreted. A release-control framework is therefore needed to prevent AI systems from speaking with more authority than the active healthcare regime can justify.

Deterministic Systems Intelligence and Admissible-State Risk Management

Deterministic Systems Intelligence is introduced here as a feasibility-first reasoning framework for constrained sequential systems. In the present context, the sequential system is the pathway from AI output to healthcare action. The relevant question is not first, “What does the model predict?” but rather, “Is this output admissible for action under the active healthcare risk regime?” This change in logical order is the central DSI move. The framework is an original conceptual contribution in this Perspective; it is anchored to adjacent literatures in responsible machine learning, clinical AI implementation, algorithmovigilance, medical algorithmic audit, safety governance and risk management rather than to a previously validated DSI instrument.

The active healthcare risk regime includes at least seven constraint domains. The first is data fitness: timeliness, missingness, measurement consistency, provenance and whether inputs match the model's intended use. The second is population fit: whether the patient group, subgroup, disease mix or care setting is within the evidence base. The third is clinical actionability: whether a feasible intervention exists and can be delivered in time. The fourth is workflow capacity: whether clinicians, escalation teams, beds, diagnostics and communication channels can absorb the output. The fifth is equity and trust: whether release could create disparate harm, stigmatisation, exclusion or unequal response. The sixth is legal and professional authority: whether the output can be used within accepted scope, consent, regulatory and accountability rules. The seventh is monitoring and reversibility: whether performance drift, adverse incidents, overrides and rollback can be detected and acted upon.

The DSI layer treats these domains as admissibility conditions rather than as secondary implementation details. An AI output that fails these conditions is not merely low confidence; it is not yet entitled to become an action signal. This is why the framework distinguishes HOLD from model failure. HOLD means that the output is withheld because the evidence, context or governance conditions do not support safe release. In a mature risk-management system, the ability to withhold is as important as the ability to predict.

Figure 1 summarises the proposed release pathway. AI output is treated as a candidate signal. The active healthcare risk regime supplies the constraints. The admissible-state layer evaluates actionability, proportionality, authorisation, auditability, reversibility and safety case. The output is then assigned to a risk-management state, and the decision is logged as part of a minimum audit record.

Five Risk-Management States

A practical clinical AI risk-management system should not be limited to "use" versus "do not use". Healthcare action is more granular than that. Some outputs are ready for routine use; others are useful only under restricted conditions; others require enhanced monitoring; some should be rolled back after safety signals; and some should be held before release. Table 1 defines the five proposed states.

Release denotes routine action under the specified conditions of use. It should be reserved for outputs that are technically supported, clinically actionable, authorised, equitable, auditable and monitorable under the active regime. Release does not imply perfection. It means that remaining uncertainty is acceptable within the safety case and response pathway.

Restricted release is appropriate when the output is useful but should be narrowed by setting, population, purpose, user group or mandatory human review. For example, a prediction tool may be admissible for prioritising case review but

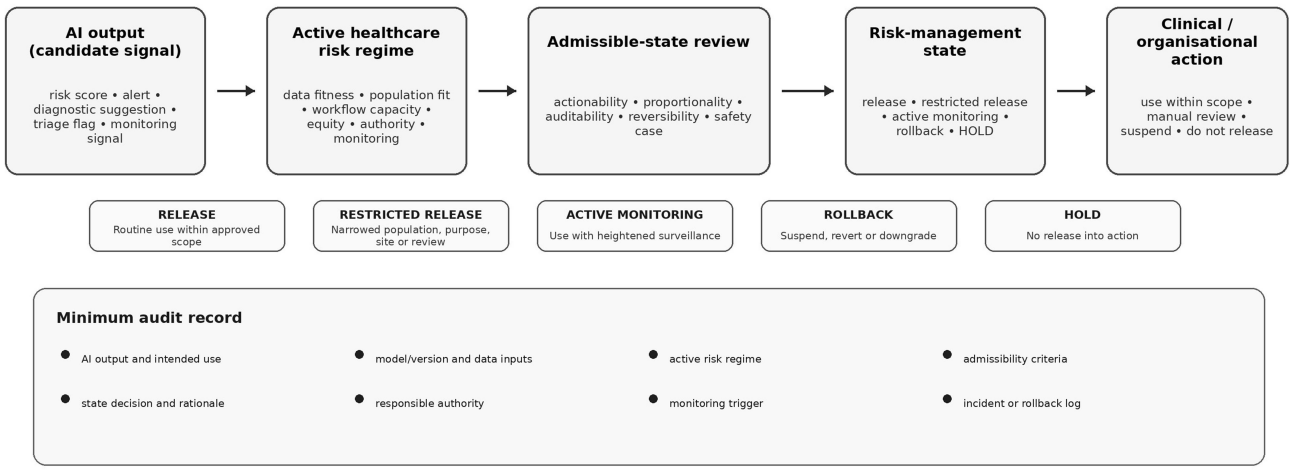

Figure 1 Feasibility-first risk management layer for clinical AI. AI output is treated as a candidate signal and evaluated against the active healthcare risk regime before it is assigned to a risk-management state. The minimum audit record preserves the AI output and intended use, model/version and data inputs, active risk regime, admissibility criteria, state decision and rationale, responsible authority, monitoring trigger, and incident or rollback log.

Table 1 Five Admissible Risk-Management States for Clinical AI Output

State	Meaning	Typical Release Condition	Risk-Management Implication
Release	Routine use is permitted under specified conditions of use.	Evidence, workflow, authority, monitoring and auditability are adequate.	Output may guide clinical or organisational action within scope.
Restricted release	Use is narrowed by population, setting, purpose, user group or mandatory review.	Output is useful but general release would over-extend evidence or safeguards.	Prevents scope creep while preserving bounded value.
Active monitoring	Use continues only with heightened performance, equity, drift and incident surveillance.	Uncertainty is acceptable only if monitoring is active and responsive.	Connects deployment to algorithmovigilance and predefined safety triggers.
Rollback	The model version, threshold, integration or use case is suspended, reverted or downgraded.	Safety, performance, equity or governance assumptions have been breached.	Turns post-deployment monitoring into corrective action.
HOLD	Output is not released into clinical or organisational action.	Evidence, feasibility, authorisation, equity safeguards, monitoring or auditability are insufficient.	Recognises disciplined non-release as a patient-safety function.

not for autonomous triage. A diagnostic aid may be released for specialist use but not for unsupervised primary-care deployment. Restricted release preserves value while preventing scope creep.

Active monitoring is a heightened state for tools whose use is allowed but whose risk profile requires intensified surveillance. Triggers may include a new site, a new population, a changed electronic health record workflow, borderline subgroup performance, early drift signals or uncertainty about downstream behaviour. Active monitoring connects release control to algorithmovigilance by making surveillance a condition of use rather than an afterthought.

Rollback is the state in which a model version, workflow integration, threshold or use case is suspended, reverted or downgraded because safety, equity, performance or governance assumptions have been breached. Rollback should not be improvised only after a major incident. It should be pre-specified through thresholds, authority lines and communication procedures. Predetermined change-control planning and postmarket monitoring are particularly relevant here.[15,16]

HOLD is disciplined non-release. It applies when the output is not sufficiently supported for action because the population is outside scope, the data are unstable, the response pathway is absent, authorisation is unclear, equity safeguards are missing, monitoring is inadequate or the audit trail is insufficient. HOLD is not anti-innovation. It is a patient-safety and organisational-risk function.

Operational Checklist for Healthcare Organisations

For the framework to be useful, admissibility must be operational rather than rhetorical. Healthcare organisations need a checklist that can be used before deployment, during local release decisions and after performance concerns arise. Table 2 proposes a compact checklist. It is intended to complement, not replace, existing model reporting, regulatory, quality-improvement and clinical governance requirements.

The checklist begins with evidence and data, but it does not end there. It asks whether the current inputs resemble the development and validation environment; whether subgroup performance and equity risks are known; whether the output is linked to a feasible intervention; whether the clinical team has authority and capacity to act; whether the workflow burden is acceptable; and whether there is an incident-reporting, monitoring and rollback pathway. In doing so, it shifts AI risk management from abstract policy language to a concrete release grammar.

This approach is particularly important in resource-constrained settings. A tool can be accurate but unsafe if it generates recommendations that the health system cannot deliver. Conversely, a modest model may be useful if released only for a narrow, auditable purpose with clear human review and monitoring. Feasibility-first reasoning therefore prevents both over-release and under-use. It helps institutions deploy AI in ways that are proportionate to the safeguards they can actually sustain.

Table 2 Practical Admissible-State Checklist for Clinical AI Risk Management

Admissibility Domain	Core Question	Minimum Evidence or Control
Data fitness	Are current inputs valid, timely, complete, correctly mapped and within intended provenance?	Data-quality checks, missingness thresholds, input-schema validation, provenance record and local data-readiness sign-off.
Population fit and equity	Does the current patient group, subgroup mix or service setting match the development and validation evidence?	Subgroup performance review, local case-mix assessment, equity-risk screen and monitoring for differential impact.
Model uncertainty and evidence boundary	Is uncertainty sufficiently characterised for the proposed use?	Calibration, confidence or uncertainty reporting; explicit intended-use boundaries; evidence limits and contraindicated-use statement.
Clinical actionability	Is there a feasible and beneficial action linked to the output?	Defined intervention, response time, escalation pathway and accountable clinician or team.
Workflow burden and alert fatigue	Can the service absorb the alert or recommendation without unsafe cognitive or operational burden?	Workflow simulation or review, staffing assessment, alert-volume threshold, prioritisation logic and usability testing.
Authority, legal and regulatory status	Who is authorised to use, override, restrict or suspend the output, and under what regulatory status?	Named governance owner, clinical lead, approval pathway, regulatory classification, liability boundaries and escalation rules.
Monitoring plan and drift detection	Can performance decay, data drift, subgroup harm and unsafe use be detected early?	Monitoring metrics, baseline performance, drift triggers, subgroup surveillance, review cadence and reporting route.
Version control and change management	Can the organisation identify which model, threshold, dataset mapping or workflow version was active?	Version log, change-control procedure, validation requirement for updates and deployment history.
Incident reporting and learning	Can safety events, near-misses and unexpected consequences be captured and acted upon?	Incident-reporting channel, taxonomy, root-cause review, feedback to users and governance committee review.
Rollback and decommissioning threshold	Are criteria defined for suspending, reverting, downgrading or withdrawing the tool?	Pre-specified rollback triggers, authority to suspend, communication plan and fallback workflow.
Human override and contestability	Can clinicians question, override or escalate disagreement with the AI output?	Override mechanism, documentation field, human-review pathway and protection against automation bias.
Audit trail	Can the release decision and downstream action be reconstructed?	Minimum audit record: output, model version, user, patient or service context, regime conditions, decision state, action taken and review trail.

Applied Clinical Example: Deterioration and Sepsis-Alert Release States

Because this Perspective does not report a new empirical dataset, the practical value of the framework is demonstrated through an applied clinical scenario. Consider a hospital deterioration or sepsis-alert model that produces a high-risk alert for an admitted patient. The model has previously shown acceptable retrospective performance. A conventional implementation pathway might treat the high-risk output as sufficient to notify clinicians or trigger escalation. A feasibility-first pathway asks a prior question: is this output admissible for the proposed action under the current local care regime?

The same alert can move through different states. If current inputs are valid, the patient population is within scope, an escalation team is available, antibiotics and diagnostics can be delivered, equity risks are monitored and a clear audit trail exists, the alert may be released for routine action. If the tool is newly introduced in a ward with unfamiliar staff or uncertain subgroup performance, it may be restricted to mandatory clinician review rather than automatic escalation. If the electronic health record mapping has recently changed or alert volume has risen sharply, use may continue only under active monitoring. If the alert is associated with missed escalations, unsafe alert burden or differential harm in a subgroup, the workflow or threshold may require rollback. If data feeds are stale, the response team is unavailable, authorisation is unclear or monitoring cannot be performed, the appropriate state is HOLD. Table 3 summarises the deterioration/sepsis-alert scenario across the five release states.

This example shows why HOLD is necessary. Without an explicit non-release state, organisations may feel compelled to act on every technically plausible output or to remove the tool entirely. HOLD creates a third option: the output is not released into clinical action until minimum conditions are restored. It protects patients from unsupported action while

Table 3 Applied Deterioration/Sepsis-Alert Example Across Release States

Clinical Scenario	DSI State	Practical Implication
Routine staffed ward; current data feed valid; escalation team available; local subgroup checks acceptable; alert volume manageable.	Release	Alert may trigger standard escalation pathway; decision, action and response are logged.
New deployment site; model within intended use but local subgroup performance and alert burden are uncertain.	Restricted release	Alert supports mandatory clinician review only; no autonomous escalation; early review period required.
Recent electronic health record mapping change or rising alert frequency; no safety incident yet.	Active monitoring	Use continues with enhanced data-quality checks, alert-volume monitoring and weekly governance review.
Safety incident, missed escalation, serious workflow overload, subgroup harm or unacceptable drift signal.	Rollback	Suspend or revert threshold/version/workflow; communicate fallback pathway and investigate cause.
Stale data feed, unavailable response team, unclear authority, absent audit trail or no monitoring capacity.	HOLD	Do not release alert into clinical action until minimum admissibility conditions are restored.

preserving the possibility of future release once data fitness, workflow capacity, authorisation, monitoring or equity safeguards have been corrected.

Relationship to Algorithmovigilance and Post-Deployment Monitoring

Algorithmovigilance has been defined as the systematic monitoring of AI-driven healthcare for effectiveness, equity and unintended consequences.[17] Recent work has drawn an explicit analogy with pharmacovigilance, arguing that healthcare AI requires post-approval evaluation, case reporting, data standardisation, causality assessment and dissemination of adverse-event signals.[18] Medical algorithmic audit similarly emphasises evaluation of AI systems for safety, quality, fairness, robustness and real-world impact.[19] The DSI framework supports these developments but places them within a broader release-control sequence.

The distinction is temporal and operational. Algorithmovigilance often begins after a tool is deployed. Admissible-state reasoning begins before release, continues during restricted release and active monitoring, and provides explicit criteria for rollback and HOLD. It therefore links pre-release risk assessment, live surveillance and post-incident learning into a single governance sequence. The framework also clarifies that monitoring is not a passive dashboard function. Monitoring is an admissibility condition: a model that cannot be monitored adequately may not be eligible for routine release, even if it performs well in retrospective evaluation.

This is also where audit records become essential. Each release decision should preserve the model output or use case, model version, local regime context, admissibility criteria, responsible authority, allowed scope, monitoring triggers and rollback thresholds. Without such records, healthcare organisations cannot reconstruct why an AI output was acted upon, why it was restricted or why it was withheld. Auditability is therefore not merely documentation; it is a core risk-control mechanism.

What DSI Adds and What It Does Not Claim

The proposed framework is intentionally modest. It does not claim that DSI can determine clinical truth, replace prospective validation, remove bias, solve regulatory uncertainty or automate accountability. It also does not claim that every AI system requires the same level of restriction. Some low-risk administrative tools may need only light release controls, while high-risk diagnostic or treatment-support tools require formal governance. The point is not to slow all AI, but to match release authority to the active risk regime.

The distinctive contribution of DSI is sequencing. Conventional AI governance often moves from model development to validation, then to deployment, monitoring and incident response. DSI inserts a feasibility-first admissibility step before outputs are allowed to become action signals. This step asks whether the output belongs to the set of actions that the current healthcare system can safely support. In other words, DSI does not replace prediction; it constrains prediction by feasibility before risk is transferred to patients, clinicians or organisations.

This sequencing also creates a useful language for decision makers. Instead of asking only whether a model is good enough, governance committees can ask: good enough for what state? Routine release, restricted release, active monitoring, rollback or HOLD? That language is practical because healthcare organisations already make graded risk decisions. DSI makes the grading explicit and auditable.

This has practical implications for procurement and governance committees. Contracts and deployment approvals should not ask only whether a vendor model has acceptable retrospective performance. They should also require an admissibility dossier: intended-use boundaries, local data-fit checks, subgroup risk review, workflow-load assessment, monitoring metrics, incident-reporting mechanism, rollback authority and a record of conditions under which outputs must be held. In this sense, feasibility-first risk management transforms AI governance from a static approval event into a continuing release-control process.

The framework also aligns with recent healthcare-risk discussions on just culture, cybersecurity and AI-enabled medical-device risk management. A just culture requires learning from adverse events without concealing accountability.[20] Cybersecurity risks show that AI deployment can introduce hazards beyond prediction error, including data integrity, system access and infrastructure vulnerability.[21] Medical-device risk-management approaches show the value of formal knowledge structures and documentation for standard-risk identification.[22] DSI adds a release-state grammar to these concerns by requiring that clinical AI outputs be mapped to release, restriction, active monitoring, rollback or HOLD before risk is transferred to patients, clinicians or institutions.

Current Scenario, Future Direction and Author Perspective

The current healthcare AI landscape is moving from proof-of-concept modelling toward procurement, integration, postmarket surveillance and institutional accountability. This transition exposes a mismatch between model-centric assurance and service-level risk management. Many tools are evaluated as predictive systems but deployed as organisational interventions. A model may satisfy retrospective performance expectations while the clinical service lacks the staffing, escalation pathway, monitoring infrastructure or rollback authority needed for safe operational use.

Future research should therefore evaluate release governance empirically. Useful study designs include simulation of AI release decisions, prospective shadow-mode evaluation, implementation pilots, governance-panel concordance studies and post-deployment audit of rollback or HOLD triggers. Research should also test whether documented release-state decisions improve safety, equity, clinician trust, incident learning and institutional accountability compared with ordinary deployment approval alone.

The author's perspective is that clinical AI risk management should mature from general assurance language to auditable release control. Innovation should not be slowed unnecessarily, but neither should technical plausibility be allowed to masquerade as readiness for clinical action. The decisive governance question is whether the current healthcare regime can support the action implied by the AI output. If it can, release may be justified. If it can do so only within boundaries, restricted release or active monitoring is appropriate. If assumptions fail, rollback is necessary. If minimum conditions are absent, HOLD is the responsible state.

Healthcare Policy Implications

At policy level, feasibility-first risk management shifts attention from one-off model approval to lifecycle release control. Health ministries, regulators, payers and organisational boards should not assume that a validated AI product is automatically admissible in every clinical pathway or institution. Local release should depend on whether data quality, staffing, response capacity, legal authority, equity safeguards, monitoring and rollback arrangements are sufficient for the intended use.

This has direct implications for procurement and contracting. Purchasers should require vendors and implementing organisations to provide an admissibility dossier before routine deployment: intended-use boundaries, model and data provenance, local data-fit checks, subgroup and equity-risk review, workflow-load assessment, monitoring metrics, incident-reporting channels, version-control procedures and pre-specified rollback thresholds. These requirements convert AI risk management from general assurance language into auditable operational conditions.

The framework is also relevant for resource-constrained health systems. A technically strong AI system can become unsafe if it recommends actions that the service cannot deliver, such as rapid diagnostics, specialist review, bed escalation, remote-monitoring follow-up or medication adjustment without the necessary workforce and supply chain. Feasibility-first policy therefore protects against importing AI tools whose performance claims are detached from local response capacity.

Finally, policy should legitimise HOLD as a safety-preserving governance state. Organisations should not be penalised for withholding an AI output when the evidence base, action pathway, safeguards or audit trail are inadequate. A mature policy environment should treat release, restricted release, active monitoring, rollback and HOLD as normal components of responsible clinical AI governance, rather than as signs of either innovation success or failure.

Conclusion

Clinical AI risk management must extend beyond model performance. A healthcare AI output may be accurate, explainable and well reported, yet still unsafe to release if the active healthcare regime cannot support the implied action. Feasibility-first reasoning provides a disciplined way to identify that boundary.

The proposed DSI framework classifies AI outputs into release, restricted release, active monitoring, rollback and HOLD. It treats withholding as a legitimate safety state rather than as a failure of innovation. It also ties monitoring, incident reporting and rollback to pre-specified admissibility conditions. This makes the framework directly relevant to healthcare risk management, patient safety and algorithmovigilance.

In practical implementation, the framework can be housed within existing clinical AI governance, quality and safety, digital health or medical-device oversight structures. Its purpose is to make the release decision explicit: who authorised the output, under which local regime, for what action, with which safeguards, and with which trigger for review, rollback or HOLD. Responsible clinical AI is therefore not only a matter of asking whether a system predicts correctly. It is also a matter of knowing when the output may act, when it must be monitored, when it must be rolled back and when it should be withheld.

Data Sharing Statement

Data sharing is not applicable because no datasets were generated or analysed for this manuscript.

Ethics Approval and Consent to Participate

Not applicable. This is a conceptual Perspective article and did not involve human participants, patient recruitment, identifiable human data, human tissue or animal data.

Author Contributions

The author made a significant contribution to the work reported in the conception, study design, execution, analysis and interpretation; took part in drafting, revising and critically reviewing the article; gave final approval of the version to be published; has agreed on the journal to which the article has been submitted; and agrees to be accountable for all aspects of the work.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Disclosure

The author declares that intellectual property protection has been sought in relation to elements of the broader Deterministic Systems Intelligence research programme. This manuscript is conceptual, does not disclose proprietary implementation details and does not depend on any protected mechanism. The author reports no conflicts of interest in this work.

References

1. World Health Organization. *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*. Geneva: World Health Organization; 2021.
2. World Health Organization. *Regulatory Considerations on Artificial Intelligence for Health*. Geneva: World Health Organization; 2023.
3. World Health Organization. *Ethics and Governance of Artificial Intelligence for Health: Guidance on Large Multi-Modal Models*. Geneva: World Health Organization; 2024.
4. Wiens J, Saria S, Sendak M, et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med*. 2019;25(9):1337–1340. doi:10.1038/s41591-019-0548-6
5. 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):231–237. doi:10.1136/bmjqs-2018-008370
6. Sendak MP, D'Arcy J, Kashyap S, et al. A path for translation of machine learning products into healthcare delivery. *EMJ Innov*. 2020;4(1):84–91.
7. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorising and evaluating non-adoption, 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
8. 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):e489–e492. doi:10.1016/S2589-7500(20)30186-2
9. Wong A, Otles E, Donnelly JP, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern Med*. 2021;181(8):1065–1070. doi:10.1001/jamainternmed.2021.2626
10. Vasey B, Nagendran M, Campbell B, 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
11. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378. doi:10.1136/bmj-2023-078378
12. Liu X, Rivera SC, 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):1364–1374. doi:10.1038/s41591-020-1034-x
13. Rivera SC, Liu X, Chan AW, Denniston AK, Calvert MJ; SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med*. 2020;26(9):1351–1363. doi:10.1038/s41591-020-1037-0
14. 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):447–453. doi:10.1126/science.aax2342
15. International Medical Device Regulators Forum. *Good Machine Learning Practice for Medical Device Development: Guiding Principles. IMDRF/AIML WG/N88 FINAL: 2025*. Ottawa: IMDRF; 2025.
16. US Food and Drug Administration. *Marketing Submission Recommendations for a Predetermined change Control Plan for Artificial Intelligence/Machine Learning-Enabled Device Software Functions*. Silver Spring (MD): FDA; 2025.
17. Embi PJ. Algorithmovigilance - advancing methods to analyze and monitor artificial intelligence-driven health care for effectiveness and equity. *JAMA Netw Open*. 2021;4(4):e214622. doi:10.1001/jamanetworkopen.2021.4622
18. Balendran A, Benchoufi M, Evgeniou T, Ravaud P. Algorithmovigilance, lessons from pharmacovigilance. *NPJ Digit Med*. 2024;7:270. doi:10.1038/s41746-024-01237-y
19. Liu X, Glocker B, McCradden MM, Ghassemi M, Denniston AK, Oakden-Rayner L. The medical algorithmic audit. *Lancet Digit Health*. 2022;4(5):e384–e397. doi:10.1016/S2589-7500(22)00003-6
20. Glarcher M, Vaismoradi M. Healthcare systems at the intersection of just culture and artificial intelligence: emerging challenges for nursing management. *Risk Manag Healthc Policy*. 2026;19:1–6. doi:10.2147/RMHP.S572893
21. Di Palma G, Scendoni R, Ferorelli D, De Benedictis A, Tambone V, De Micco F. AI-induced cybersecurity risks in healthcare: a narrative review of blockchain-based solutions within a clinical risk management framework. *Risk Manag Healthc Policy*. 2025;18:3479–3497. doi:10.2147/RMHP.S544523
22. Zhu W, Zhang P, Xia W, et al. AI-driven medical device risk management: a new paradigm integrating large language models and prompt engineering for standard-risk knowledge graph construction and application. *Risk Manag Healthc Policy*. 2026;19:1–17.