

ISO 15189-ALIGNED GOVERNANCE FRAMEWORK FOR CLINICAL AI IN MEDICAL LABORATORIES: A PROPOSED CONTROL MATRIX, EVIDENCE PACKAGE, AND RISK-TIERED MATURITY ROADMAP FOR VERIFICATION, MONITORING, AND CHANGE CONTROL

OKVIR UPRAVLJANJA KLINIČKOM VEŠTAČKOM INTELIGENCIJOM U MEDICINSKIM LABORATORIJAMA USKLAĐEN SA STANDARDOM ISO 15189: PREDLOG MATRICE KONTROLE, PAKETA DOKAZA I FAZNE MAPE RAZVOJA ZASNOVANE NA NIVOIMA RIZIKA ZA VERIFIKACIJU, PRAĆENJE I KONTROLU PROMENA

Bandar Saad Alshreef

Department of Clinical Laboratory Sciences, College of Applied Medical Sciences,
Shaqra University, Shaqra, Saudi Arabia

Summary

Background: Artificial intelligence is increasingly used in medical laboratories, yet ISO 15189:2022 does not specify artificial-intelligence-specific thresholds, lifecycle monitoring, or change control. This study developed a proposed ISO 15189-aligned governance framework that distinguishes externally supplied systems from in-house or materially modified systems.

Methods: A design science approach used a bounded purposive corpus of 31 standards, regulations, professional guidance documents, reporting guidelines, and laboratory artificial-intelligence publications. Sources were coded across intended use and regulatory route, data and measurement inputs, supplier or developer evidence, local verification or validation, deployment qualification, monitoring, incident management, and change control. Evaluation was author-led and formative; no independent expert panel, prospective pilot, or accreditation-audit validation was performed.

Kratak sadržaj

Uvod: Veštačka inteligencija se sve više koristi u medicinskim laboratorijama, ali ISO 15189:2022 ne definiše pragove specifične za veštačku inteligenciju, praćenje tokomivotnog ciklusa niti kontrolu promena. U ovoj studiji razvijen je predloženi okvir upravljanja usklađen sa standardom ISO 15189, kojim se razlikuju sistemi nabavljeni od eksternih dobavljača od sistema razvijenih interno ili značajno modifikovanih sistema.

Metode: Primenjen je pristup nauke o dizajnu na ograničenom, namenski odabranom korpusu od 31 standarda, propisa, dokumenata sa stručnim smernicama, smernica za izveštavanje i publikacija o veštačkoj inteligenciji u medicinskim laboratorijama. Izvori su kodirani prema sledećim oblastima: predviđena namena i regulatorni put, podaci i merni ulazni parametri, dokazi dobavljača ili programera, lokalna verifikacija ili validacija, kvalifikacija za primenu, praćenje, upravljanje incidentima i kontrola promena. Evaluaciju su sprovedi autori i bila je

Address for correspondence:

Bandar Saad Alshreef
e-mail: bsalshreef@su.edu.sa

List of abbreviations: AI, artificial intelligence; CAP, College of American Pathologists; CAPA, corrective and preventive action; DSR, design science research; EHR, electronic health record; EQA, external quality assessment; HCI, human-computer interaction; IFCC C-AILM, International Federation of Clinical Chemistry and Laboratory Medicine Committee on Artificial Intelligence in Laboratory Medicine; IQ, installation qualification; IVDR, in vitro diagnostic medical devices regulation; LIS, laboratory information system; ML, machine learning; MU, measurement uncertainty; NPV, negative predictive value; OQ, operational qualification; PBRTQC, patient-based real-time quality control; PEP, Proposed Evidence Package; PPV, positive predictive value; PQ, performance qualification; QC, quality control; QMS, quality management system; RACI, responsible, accountable, consulted, informed; SAAC, Saudi Accreditation Centre; SFDA, Saudi Food and Drug Authority; URS, user requirements specification.

Results: Five linked artefacts were produced: a qualification and regulatory-route gate; a clause-linked control matrix separating paraphrased normative ISO requirements from author-derived operationalisation; a proposed evidence package incorporating user requirements and installation, operational, and performance qualification; a risk-tiered maturity roadmap; and a worked autoverification monitoring artefact with explicit metrics, proposed local thresholds, log fields, review cadence, and escalation actions. The framework also addresses European regulatory roles, measurement uncertainty, external quality assessment, input harmonisation, and clinical utility.

Conclusions: The framework is a proposed implementation aid, not a validated accreditation instrument. It organises ISO 15189 requirements and clinical artificial-intelligence lifecycle controls while requiring local regulatory classification, risk assessment, threshold calibration, and prospective evaluation before routine use.

Keywords: artificial intelligence, governance, ISO 15189, medical laboratories, Saudi Arabia, verification

Introduction

The integration of artificial intelligence (AI) and machine learning (ML) into ISO 15189:2022 quality management systems (1) is reshaping laboratory medicine, supporting applications ranging from pre-analytical decision support and ML-enabled autoverification to advanced image analysis and patient-based real-time quality control (PBRTQC) (2–5). However, AI models operate within socio-technical workflows; outputs can be sensitive to data quality, shifting populations, instrument/reagent changes, and local processes, creating risks of silent performance degradation (drift) and human-factor failures such as automation bias (2, 5).

Although ISO 15189:2022 strengthens risk-based thinking and management-system expectations, it provides limited explicit guidance on AI-specific lifecycle risks (e.g., dataset shift, drift, automation bias, and predetermined change control). The 2026 IFCC C-AILM survey and recommendations (6) substantially advance implementation by specifying provider information, local verification, version-integrity controls, input/output monitoring, application-specific quality control, external quality assessment, and risk-adapted surveillance. The remaining gap addressed here is narrower: how to integrate those practices with clause-level ISO 15189 evidence, regulatory-route decisions, and locally auditable records.

formativnog karaktera; nisu sprovedeni nezavisna ekspertiska procena, prospektivno pilot-ispitivanje, niti validacija putem akreditacione provere.

Rezultati: Izrađeno je pet međusobno povezanih elemenata: kriterijum za kvalifikaciju i određivanje regulatornog puta; matrica kontrole povezana sa odredbama standarda, koja razdvaja parafrazirane normativne zahteve ISO standarda od operacionalizacije koju su definisali autori; predloženi paket dokaza koji obuhvata zahteve korisnika i kvalifikaciju instalacije, rada i performansi; fazna mapa razvoja zasnovana na nivoima rizika; i praktični model praćenja autoverifikacije sa eksplicitnim pokazateljima, predloženim lokalnim pragovima, poljima evidencije, učestalošću pregleda i merama eskalacije. Okvir takođe obuhvata regulatorne uloge u Evropi, mernu nesigurnost, eksternu procenu kvaliteta, harmonizaciju ulaznih podataka i kliničku korisnost.

Zaključak: Ovaj okvir predstavlja predloženo sredstvo za implementaciju, a ne validirani instrument za akreditaciju. Njime se sistematizuju zahtevi standarda ISO 15189 i kontrole životnog ciklusa kliničke veštačke inteligencije, uz obaveznu lokalnu regulatornu klasifikaciju, procenu rizika, kalibraciju pragova i prospektivnu evaluaciju pre uvođenja u rutinsku primenu.

Ključne reči: veštačka inteligencija, upravljanje, ISO 15189, medicinske laboratorije, Saudijska Arabija, verifikacija

Published frameworks and standards

Published guidance relevant to AI/ML in clinical laboratories spans consensus recommendations, accreditation-oriented concept papers, laboratory-specific lifecycle architectures, and regulatory analyses. Box 1 summarises six frequently cited sources and indicates how explicitly they engage ISO 15189 or laboratory quality systems.

Gaps and revised scope

The gap addressed here is narrower than in the original submission. The contribution is not that operational laboratory-AI guidance is absent; reference (6) already provides detailed best-practice recommendations. The remaining implementation problem is how to assemble those recommendations with ISO 15189 clause ownership, evidence records, regulatory-route decisions, risk-tiered control depth, and a reproducible local monitoring log that can be embedded in an existing laboratory QMS. Clause traceability: AI controls should identify the relevant ISO 15189 clause and clearly distinguish the standard's normative requirement from the author's AI-specific operational interpretation.

- Regulatory route: a CE-marked commercial system used as intended is not governed identically to an in-house model, an off-label use, or a retrained/customised system.
- Operational surveillance: metrics require local baselines, thresholds, log fields, ownership, review cadence, and escalation actions, not just a list of drift indices.

Box 1 Published guidance and frameworks relevant to AI/ML in clinical laboratories and their QMS alignment.

Framework or guidance	Scope and audience	Key elements described	ISO 15189 or QMS alignment
IFCC working group recommendations (2)	Clinical laboratory medicine practitioners and researchers	Consensus best practices across the ML lifecycle (problem formulation → development → deployment), emphasising reproducibility and quality in laboratory ML workflows	Quality-system relevant practices are described but without clause-level ISO 15189 crosswalk
CAP performance-evaluation concept paper (7)	Pathology laboratories evaluating ML systems for clinical use	Risk assessment; data sourcing/curation; verification/validation; change control and revalidation; human-computer interaction; training and competency	Explicitly maps accreditation-relevant activities (validation, revalidation, quality management) to ML evaluation needs
CARE lifecycle series introduction (laboratory-specific) (3)	Clinical laboratory and pathology audience	Comprehensive laboratory AI lifecycle and readiness milestones (CARE) to bridge the gap between model development and clinical implementation	Explicitly targets clinical laboratory test development and adoption within accreditation contexts
Spies et al. mini-review on validation and monitoring (5)	Clinical laboratories implementing ML solutions	Practical validation study design considerations, multidisciplinary implementation roles, and routine monitoring for performance degradation	Emphasises operational roles and monitoring consistent with QMS expectations
European regulatory landscape analysis (8)	Manufacturers, laboratories, and regulators in Europe	IVDR and AI-MDSW obligations; clinical evidence expectations; post-market surveillance challenges; delineation of responsibilities	Discusses how ISO 15189 and IVDR obligations intersect for AI-driven software used in labs
IFCC C-AILM survey/recommendations (6)	Clinical laboratories and AI providers	Provider specification, local verification, version-integrity checks, input/output monitoring, application-specific QC, EQA, and risk-adapted surveillance	Direct operational benchmark for this revision; the present framework adds ISO 15189 clause ownership, regulatory-route separation, and auditable local records

- Metrological and interlaboratory validity: model inputs must account for measurement uncertainty, harmonisation, and EQA/interlaboratory comparison where interpretation or model output depends on laboratory measurements.
- Risk proportionality: monitoring intensity and maturity expectations should scale with clinical consequence and regulatory risk rather than require Level 5 controls for low-risk software.

Accordingly, this article presents a proposed implementation framework. It does not claim that the artefacts are accreditation-approved, empirically validated, or universally minimal. The Saudi context is retained as an implementation overlay through ISO 15189 accreditation transition considerations and SFDA medical-device guidance (9–11), while the European regulatory route is added because peer review specifically identified it as a major omission (12).

Materials and Methods

Study design and evidence boundary

We retained a design science research (DSR) approach because the output is an implementation artefact rather than an effectiveness study (12). The final evidence corpus was bounded to the 31 documents cited in this manuscript. Selection was purposive and targeted, not systematic. The corpus comprised laboratory and conformity standards (13–15), AI risk-management frameworks (14–16), national and international health-AI/device guidance and regulation (10, 11, 17, 18, 26, 27), professional and peer-reviewed laboratory-AI guidance and implementation literature (2–9), PBRTQC literature (23, 24), AI reporting/evaluation guidance (19–22), and regulatory lifecycle literature (25). No PRISMA-style completeness claim is made.

The DSR stages implemented in this manuscript were problem definition, artefact design, analytical demonstration, and formative author review.

Evaluation included traceability checks, internal completeness checks, and a worked use-case application. We did not perform independent expert review, a live accreditation audit, a prospective clinical pilot, or inter-rater validation. Consequently, we use terms such as »proposed«, »illustrative«, and »implementation aid« throughout.

Coding frame and clause mapping

Sources were coded using a predefined lifecycle frame: (1) qualification, intended use, and regulatory route; (2) data and measurement inputs; (3) supplier/developer evidence; (4) local verification or validation; (5) installation, operational, and performance qualification; (6) deployment and human oversight; (7) post-deployment monitoring and quality control; (8) incidents, nonconforming work, and CAPA; and (9) modification, retraining, rollback, and retirement. Each proposed control was then linked to one or more ISO 15189:2022 clauses.

To avoid implying that the author's operational choices are themselves ISO requirements, each mapping is labelled either »Normative ISO requirement–paraphrased« or »Author-derived AI operationalisation« . ISO 15189 text is copyrighted; the tables therefore provide clause numbers and concise requirement-focused paraphrases rather than reproducing extensive normative text. The licensed ISO 15189:2022 publication remains the controlling source (1).

Regulatory and terminology rules

The revised framework separates externally supplied systems used within the manufacturer's intended purpose from in-house, off-label, or materially modified systems. For externally supplied systems, the laboratory's principal local task is verification that the system performs acceptably in the intended environment, consistent with the distinction between verification and validation in conformity-assessment terminology and with the local-verification emphasis in reference (6). In-house or materially modified systems require a development and validation pathway, with regulatory responsibilities determined by jurisdiction and intended use; ISO 5649:2024 provides additional laboratory-developed-test concepts relevant to internal development (28).

For European use, IVDR Article 5(5) provides an in-house health-institution route only when its stated conditions are met, including QMS and ISO 15189-related conditions. AI systems that are products or safety components under Annex I legislation and require third-party conformity assessment can meet the AI Act Article 6(1) high-risk rule (26). For such systems, Article 26 includes deployer duties

concerning competent human oversight, sufficiently representative input data within deployer control, monitoring and incident escalation, and retention of automatically generated system logs under deployer control for at least six months, unless otherwise provided by applicable Union or national law. The consolidated AI Act text current at revision (27 July 2026) applies Chapter III Sections 1–3 to Article 6(1)/Annex I systems from 2 August 2028 (26).

A further gate is required before retraining, fine-tuning, changing intended purpose, or otherwise customising a commercial high-risk AI system. Under AI Act Article 25, a deployer making a substantial modification can become the provider for that system; under IVDR Article 16, a person modifying a device in a way that may affect compliance can assume manufacturer obligations (26, 27). Therefore, local adaptation is not treated as a routine Level-5 maturity activity; it is a regulatory-change event requiring prior classification and legal review.

Demonstration use cases and risk tiers

We retained three representative use cases: ML-based autoverification, PBRTQC, and AI-assisted morphology review. These were chosen to expose different control needs: automated release and automation bias; continuous numeric-data drift; and image representativeness, interpretation, and EQA. A three-tier risk lens was added: Tier A (low-impact operational support with easy human reversibility), Tier B (moderate clinical influence or workflow prioritisation), and Tier C (high-impact diagnostic, critical-result, or autonomous/semiautonomous action). Risk tier is a local governance construct, not an ISO or regulatory classification, and should not replace jurisdiction-specific device classification.

Saudi implementation overlay

For Saudi laboratories, the framework is intended to sit inside the laboratory's ISO 15189 QMS and existing accreditation transition work rather than create a parallel system. Saudi Accreditation Centre (SAAC) transition materials and Saudi Food and Drug Authority (SFDA) AI/ML medical-device guidance are therefore treated as local implementation references (10, 11). The proposed route gate requires the laboratory to document whether the AI is an externally supplied regulated medical device, an internally developed tool, or a modified use, and to confirm applicable SFDA requirements before clinical activation. The framework does not infer a Saudi device classification from the use case alone.

Results

Qualification and classification gate

The first artefact is a pre-go-live gate that prevents technical validation work from starting until the intended use, regulatory route, modification status, and clinical risk have been defined (*Table I*).

Clause-linked control matrix by regulatory route

Table II links ISO 15189 clauses to AI-specific controls. The »ISO requirement focus« is a concise paraphrase, not a substitute for the licensed standard. The »AI operationalisation« column is the author's proposed interpretation and is not presented as an ISO »shall« statement.

Table I Qualification and regulatory-route gate for clinical laboratory AI.

Gate	Question	Required documented decision	Route
1. Intended use	What decision, population, specimen/data type, instrument, site and user role are within scope?	Document bounded intended use, exclusions, clinical consequence and final decision authority.	All routes
2. Supplier/developer route	Is the system externally supplied and used within the approved/labelled intended purpose?	If yes: commercial/deployer route → local verification. If no: in-house/off-label/modified route → development/validation and regulatory reassessment.	Commercial vs in-house
3. European in-house route	If used in the EU, does an internally developed device rely on IVDR Art. 5(5)?	Document health-institution conditions, QMS basis, internal-use limitation and justification; confirm national implementation.	In-house EU only
4. High-risk AI screen	Is the AI a product/safety component under Annex I legislation and subject to third-party conformity assessment?	If yes, treat as Article 6(1) high-risk when applicable; map deployer/provider duties and application date.	EU regulated systems
5. Modification gate	Will the laboratory retrain, fine-tune, change intended purpose, thresholds, model components, or other behaviour-affecting elements?	Stop before change; assess substantial modification, provider/manufacture role, conformity implications, re-verification/re-validation and rollback.	All modified systems
6. Local risk tier	What is the consequence of failure, autonomy, detectability and reversibility?	Assign Tier A/B/C to scale evidence depth, audit frequency and escalation.	Local QMS

Table II ISO 15189–AI control matrix with normative/interpretive distinction and route-specific evidence.

ISO 15189:2022 clause(s)	ISO requirement focus/status	Author-derived AI operationalisation	Illustrative evidence	Route
5.2; 5.4; 5.6	Normative ISO requirement–paraphrased: accountable laboratory leadership, defined authority and risk management.	Define AI owner, clinical decision authority, escalation authority and regulatory-route owner.	Governance charter; RACI; intended-use statement; route decision; risk tier.	Both
6.2	Normative ISO requirement–paraphrased: personnel competence, authorisation and continuing competence.	Train users on limitations, automation bias, override rules, incident recognition and safe fallback.	Training records; competency assessment; authorisation list; simulation scenarios.	Both
6.4; 7.6	Normative ISO requirement–paraphrased: equipment/software and data/information systems are controlled and fit for use.	Lock version; verify interfaces; record checksum/hash; control access; verify LIS/EHR field mapping and rollback.	URS; IQ/OQ records; model/version register; hash; interface test; release notes; rollback evidence.	Commercial: supplier + local verification; in-house: developer + local validation

7.2	Normative ISO requirement–paraphrased: pre-examination inputs are controlled.	Specify admissible specimen/ data inputs, units, missingness, quality flags and harmonisation requirements.	Data dictionary; input QC SOP; instrument/ reagent mapping; missingness rules; rejection rules.	Both
7.3.3; 7.3.4(f)	Normative ISO requirement–paraphrased: examination procedures are fit for intended use; input measurement uncertainty that can affect software/ model output is considered.	Quantify or clinically assess sensitivity of AI output to measurement uncertainty and non-harmonised measurands; define site/platform limits.	MU impact analysis; perturbation/ sensitivity study; assay/platform comparability; site-specific thresholds if needed.	Both
7.3.7; 7.3.7.3(a)	Normative ISO requirement–paraphrased: validity of examination results is monitored; relevant EQA/ interlaboratory comparison includes interpretation where applicable.	Include model interpretation/output in EQA or equivalent interlaboratory comparison when an appropriate scheme exists; otherwise, document justified alternatives.	EQA participation/ results; standardised-case comparison; blinded image review; alternative-EQA rationale.	Both
7.4	Normative ISO requirement–paraphrased: results are reviewed, released and communicated appropriately.	Define confidence display, interpretive limitations, human confirmation and critical-result exclusions.	Reporting template; override SOP; critical-result policy; uncertainty display.	Both
7.5; 8.7	Normative ISO requirement–paraphrased: nonconforming work is controlled, and corrective action is documented.	Treat unsafe AI behaviour, drift alarms, and interface failures as QMS events with containment, CAPA, and re-verification.	Incident log; CAPA; affected-result review; suspension/ rollback decision.	Both
8.3; 8.4	Normative ISO requirement–paraphrased: documents and records are controlled and retained.	Retain version-linked evidence, monitoring logs, approvals, release notes and decision records for defined periods; where EU AI Act Article 26(6) applies, retain automatically generated system logs under deployer control for at least six months, unless otherwise provided by applicable Union or national law.	Document-control index; versioned evidence package; monitoring log; retention schedule.	Both; EU high-risk overlay where applicable
8.5; 8.8; 8.9	Normative ISO requirement–paraphrased: risks/opportunities are addressed, performance is evaluated, and management reviews the QMS.	Trend AI KPIs, subgroup performance, drift, incidents, overrides, vendor changes and unresolved risks at risk-proportionate intervals.	Dashboard; quality indicators; internal-audit evidence; management-review minutes; risk register.	Both

Note: ISO requirement-focus wording in Table II is deliberately paraphrased. Check the exact normative wording against the laboratory's licensed ISO 15189:2022 copy. »Both« indicates that the QMS control applies to both externally supplied and internally developed systems, although verification/validation responsibility differs.

Proposed Evidence Package (PEP)

The former »minimum evidence package« has been renamed the Proposed Evidence Package (PEP) because no empirical study has established a universal minimum. The PEP is a modular evidence set that should be expanded or reduced according to risk, regulatory route, and local QMS needs (Table III).

Risk-tiered maturity roadmap

The five maturity levels remain as organisational capability descriptors, but they are no longer presented as a universal staircase every AI system must climb. A low-risk Tier A tool can remain appropriately controlled at a lower organisational maturity if its risks are demonstrably managed. Higher-impact Tier C systems require more intensive monitor-

Table III Proposed Evidence Package for verification/validation, qualification, monitoring, and change control.

PEP module	Core evidence	Metrics/acceptance evidence	Cadence/ trigger	ISO clauses
Scope & governance	Intended use; exclusions; workflow map; benefit-risk analysis; route decision; risk tier; accountable owner.	Scope coverage; unresolved hazards; decision authority.	Pre-go-live; review after scope change.	5.2, 5.4, 5.6, 8.5
Supplier/ developer specification	Model specification sheet; training/test population; inputs; operating range; known limitations; failure modes; external performance; version history.	Completeness against local use; clinically relevant metrics.	Pre-procurement/ pre-development; each major release.	6.4, 7.6, 8.3
Data & measurement inputs	Data dictionary; provenance; units; missingness; assay/platform mapping; MU impact; harmonisation assessment.	Representativeness; missingness; feature-positive rates; platform comparability; perturbation sensitivity.	Pre-go-live; after instrument/reagent/ interface change.	7.2, 7.3.4(f)
Verification/ validation	Commercial: local verification report. In-house/modified: analytical/clinical validation plan and report as appropriate.	Discrimination: AUC/C-index, sensitivity, specificity, PPV/NPV. Calibration: plot, intercept/slope, Brier score. Overall: log loss/Brier or task-appropriate global metric. Classification: confusion matrix/F1 where relevant. Clinical utility: decision-curve/net benefit or prespecified workflow utility. Subgroup performance and uncertainty intervals.	Pre-go-live; after defined triggers.	7.3.3, 7.3.4, 8.8
Qualification (URS/IQ/OQ/ PQ)	User requirements specification; installation qualification; operational qualification; performance qualification; fail-safe/rollback test; user competency sign-off.	Interface mapping; latency/uptime where relevant; regression-test pass rate; workflow conformance; acceptance criteria.	Before clinical activation and after major infrastructure change.	6.4, 7.6
QC/EQA	Application-specific QC; anonymous expert review for image models; synthetic/curated known-output regression cases for numeric models; EQA/interlaboratory comparison where available.	QC pass rate; inter-reader/inter-site agreement; EQA score; regression mismatches.	Risk-tiered: continuous to quarterly; EQA per programme cycle.	7.3.7, 7.3.7.3
Monitoring & incidents	Provider release notes and integrity verification; local input/output distribution monitoring; adjudicated performance audit; incident/CAPA linkage.	Feature-positive rate; class frequency; drift statistic; override/disagreement; false auto-release; calibration; subgroup error; time to detect/respond.	Continuous technical checks; weekly/monthly/quarterly governance cadence by risk.	7.5, 8.4, 8.7, 8.8, 8.9
Change control & retirement	Change request; regulatory-role reassessment; impact analysis; re-verification/re-validation; rollback; retirement and archive.	Change classification; affected functions; comparative performance; unresolved risk.	Before any behaviour-affecting change; post-change review.	6.4, 7.6, 8.3, 8.5

Table IV Risk-tiered AI governance maturity roadmap.

Level	Governance & QMS integration	Verification/monitoring	Risk-proportional interpretation
1. Basic	Named owner; intended use; version control; basic local check.	Manual/ad hoc checks; documented fallback.	Reasonable starting point for Tier A only; not sufficient for high-impact clinical autonomy.
2. Standardised	Risk register; formal training; PEP baseline; CAPA linkage.	Prespecified acceptance criteria; baseline dashboard.	Expected baseline for most Tier B deployments.
3. Managed	Management review includes AI KPIs; supplier/change protocols; formal route reassessment.	Automated or scheduled drift checks; periodic adjudicated audit; tested rollback.	Preferred minimum organisational capability for many Tier C systems.
4. Optimised	AI management-system practices integrated with QMS; cross-site governance; internal audit.	Continuous signals plus risk-triggered performance review; human-factors audits.	Appropriate for complex or multi-model Tier C environments.
5. Learning	External benchmarking; evidence-driven improvement; independent review; controlled innovation.	Adaptive improvement only within legally permitted change-control boundaries; no assumption that local retraining is permissible.	Optional advanced capability; not required for low-risk tools and not a license to modify regulated systems.

ing and governance regardless of the laboratory's nominal maturity level.

Complete worked artefact: ML-based autoverification surveillance log

Table V provides a complete illustrative artefact for a Tier C ML-assisted autoverification system. Thresholds are deliberately labelled »proposed local starting thresholds«; they are not ISO requirements and must be calibrated to local prevalence, risk tolerance, analytical performance, and historical workflow. The system is assumed to recommend release/hold decisions while existing critical-result, analytical-quality, delta-check, and hard-stop rules remain authoritative.

The same monitoring logic can be adapted to PBRTQC and morphology. For image models, regular anonymous expert review of representative cases and EQA/standardised case comparison are emphasised; for numeric models, known-output regression cases and input/output distribution monitoring are particularly useful (6).

Novelty and item-level delta versus existing guidance

The novelty claim is deliberately narrowed. The framework does not claim to originate supplier specification sheets, local verification, drift monitoring, QC, or lifecycle governance. Its proposed contribution is the integration shown in Table VI.

Use-case implications

Autoverification: Externally supplied models should be locally verified against the intended workflow and hard-stop rules. In-house or retrained models require validation and a regulatory-role assessment. Monitoring should emphasise false auto-release, class frequency, feature-positive rates, interface regression, overrides, and known-output cases.

PBRTQC: The framework emphasises input harmonisation, instrument/reagent change detection, continuous distribution monitoring, pre-defined control limits, and CAPA linkage. A change in analytical system or data interface can require renewed verification even when the model code is unchanged.

Morphology review: The framework emphasises local staining/scanner representativeness,

Table V Illustrative autoverification monitoring and escalation artefact.

Signal/control	Metric/log field	Proposed local starting threshold	Cadence	Escalation/evidence
Model/version integrity	SHA-256 model hash; software/container version; release-note ID	Exact match to approved baseline. Any unexplained mismatch = RED.	At startup, daily automated check, and after any update.	Immediate hold; verify release; rollback or re-verify before use.
Input interface integrity	LIS/EHR field names, units, null mapping, reference-range encoding	100% pass on curated interface regression set. Any schema mismatch affecting a model input = RED.	After interface/configuration change; monthly regression test.	Suspend affected pathway; correct mapping; repeat OQ/PQ.
Feature-positive rate	Proportion of records with each prespecified binary/derived feature	YELLOW if absolute shift >5 percentage points or >3 baseline SD for 2 consecutive windows; RED if >10 points or clinically unexplained abrupt shift.	Rolling 1,000 eligible results; reviewed weekly.	Investigate case mix/instrument/interface; increase adjudication; re-baseline only with approval.
Predicted class frequency	Auto-release/hold/manual-review distribution	YELLOW if relative change >20% from locked baseline for 2 windows; RED if >30% or associated with safety signal.	Rolling 1,000 results; weekly dashboard.	Check prevalence and threshold behaviour; restrict automation if unresolved.
Safety-critical outcome	Incorrect auto-release of a critical or explicitly excluded result	Target = 0. Any confirmed event = BLACK/STOP.	Continuous incident capture; monthly adjudicated audit.	Immediate suspension of automatic release; patient/result impact review; CAPA; re-verification before restart.
Adjudicated discrimination/classification	Sensitivity for cases that require manual review; specificity; PPV/NPV; confusion matrix	Sensitivity must remain $\geq 95\%$ and not decline by more than 5 points from the locked verified baseline; review other metrics against prespecified local acceptance limits.	Monthly stratified audit; all serious incidents included.	Escalate if limit breached; increase sample; restrict use; investigate subgroup/interface causes.
Calibration/overall performance	Calibration plot; calibration intercept/slope; Brier score	YELLOW if Brier worsens >0.02 from baseline or calibration slope leaves 0.8–1.2; RED if accompanied by safety-performance breach.	Quarterly or after material case-mix/model change.	Recalibration only after regulatory/change-control review; otherwise adjust workflow or suspend.
Clinical utility	Decision-curve/net benefit at operational threshold; manual-review burden	Net benefit must not fall below the locked rule-based comparator at the chosen operating threshold; >20% unexplained rise in manual-review burden triggers review.	Quarterly.	Governance review of threshold/workflow; do not change threshold without change control.
Known-output regression set	Curated/synthetic cases with expected model outputs	100% exact agreement for deterministic expected outputs; any safety-relevant mismatch = RED.	Monthly and after any software/model/interface update.	Hold release; compare version/hash; rollback or re-verify.
Management review log	Timestamp; model/version/hash; data window; metric; baseline; current value; threshold; status; action owner; CAPA/change ID; disposition	Assign all alerts an owner within 1 business day; document RED/BLACK actions before restart.	Weekly operational review; monthly clinical review; quarterly management review.	Close only with documented evidence, approval, and re-verification where triggered.

Table VI Item-level delta between the proposed framework and key prior guidance.

Source	What it already provides	Proposed incremental contribution here
IFCC C-AILM (6)	Provider specification; local verification; integrity checks; release notes; input/output drift monitoring; application-specific QC; EQA; risk-adapted surveillance.	Embeds these practices in a numbered ISO 15189 control/evidence structure, regulatory-route gate, risk-tiered maturity model, and worked QMS monitoring log.
CAP (7)	Pathology-focused ML performance evaluation, HCI, risk and change-control considerations.	Extends to multiple laboratory use cases and explicitly separates commercial verification from in-house/modified validation routes.
CARE (3)	Comprehensive laboratory/pathology AI lifecycle and readiness milestones.	Adds clause-level ISO 15189 evidence ownership, regulatory modification gate, PEP modules, and risk-proportional monitoring cadence.
Spies et al. (5)	Practical validation, implementation roles and routine monitoring.	Adds operational log schema, proposed thresholds, IQ/OQ/PQ structure, measurement-input uncertainty and EQA linkage.
ISO/IEC 42001 (15)	Generic AI management-system requirements.	Translates generic AI governance into laboratory-specific ISO 15189 records and process controls.
ISO/DIS 24051-1 and ISO/CD 24051-2 (29, 30)	Emerging AI principles for medical laboratories and digital pathology/image analysis.	Positions the framework as a non-normative implementation bridge while these projects remain under development, not as a competing standard.

anonymous expert review, confidence/uncertainty communication, EQA or standardised interlaboratory cases where available, and explicit handling of interpretive disagreement. Clause 7.3.7.3 is therefore directly relevant to AI-assisted interpretation.

Discussion

This work converts fragmented AI governance guidance into a proposed ISO 15189 implementation structure by providing a clause-linked crosswalk of controls and evidence. Unlike purely conceptual governance models, the proposed matrix and evidence package specify what to document and how to operate AI over time, without claiming the artefacts are already audit-ready or empirically validated. No live accreditation audit, prospective clinical study, or independent expert panel has evaluated the framework.

Harmonisation with ISO/IEC 23894, ISO/IEC 42001, and NIST AI RMF helps ensure that laboratory AI governance is compatible with internationally recognised AI risk management practices (14–16). Regulatory alignment with IMDRF GMLP and SFDA MDS-G010 supports laboratories that deploy commercial AI/IVD software and those developing in-house models that may be treated as medical device software (11, 17). The strongest practical benchmark for the present revision is reference (6), whose provider-side integrity checks,

release documentation, local input/output monitoring, application-specific QC, EQA, and risk-adapted surveillance are incorporated into the proposed ISO 15189-oriented governance architecture.

The regulatory split is equally important. A laboratory that purchases a CE-marked system and uses it as intended is primarily a deployer/user and should focus on local verification, competent oversight, monitoring, records, and supplier change control. An in-house system or a substantially modified commercial system can trigger a different regulatory role. Under the current EU framework, IVDR Article 5(5) sets conditions for in-house devices, IVDR Article 16 can transfer manufacturer obligations when modifications may affect compliance, and AI Act Article 25 can make a deployer the provider after a substantial modification (26, 27). These distinctions make uncontrolled local retraining an inappropriate marker of »maturity«.

The AI Act timing also requires date-specific interpretation. The reviewer correctly identified the importance of Article 26 deployer duties. Still, the consolidated EU text current at this revision applies Chapter III Sections 1–3 to Article 6(1)/Annex I high-risk systems from 2 August 2028 (26). Laboratories should therefore cite and monitor the current consolidated legislation rather than rely on a fixed historical implementation date.

Measurement-system context is another major addition. Clinical AI can be mathematically

stable while its inputs are not comparable across instruments, assays, units, sites, or reagent lots. The revised matrix therefore links AI verification to measurement uncertainty, harmonisation, and EQA/interlaboratory comparison. This is especially important for calculated scores and models that combine several laboratory measurands: identical software does not guarantee equivalent output if upstream measurements are not commutable or harmonised (28).

For Saudi implementation, the framework can be embedded in existing ISO 15189 QMS structures and mapped to SAAC transition activities and SFDA device governance (10, 11). The article does not claim that the European regulatory route applies in Saudi Arabia; instead, the EU section shows why any jurisdiction must explicitly define the regulatory role. Saudi laboratories should document local SFDA classification and authorisation requirements as part of the qualification gate.

Emerging ISO work further narrows the novelty claim. As of this revision, ISO/DIS 24051-1 is in the Draft International Standard process and ISO/CD 24051-2 remains under development (29, 30). The proposed framework should therefore be treated as an interim implementation architecture that may need revision when those projects advance or are published.

Monitoring and change control are emphasised because AI models may drift as patient populations, instruments, reagents, or workflows change. Recent laboratory-focused implementation guidance stresses that successful deployment requires not only initial validation but also structured post-deployment surveillance (2, 5, 6). PBRTQC literature provides a natural bridge for laboratories to operationalise drift detection and embed it into occurrence-management processes (23, 24). In parallel, regulatory analyses continue to highlight the need for lifecycle-integrated evidence and post-market surveillance approaches that can be translated into audit records (8, 25). The worked auto-verification artefact extends these principles by adding explicit proposed local thresholds, review cadence, log fields, ownership, and escalation actions.

Limitations: First, the source corpus was purposive rather than systematic. Second, the author developed the clause mappings and operational thresholds, and they have not undergone independent consensus validation. Third, the worked thresholds are illustrative starting points, not universal performance specifications. Fourth, legal and regulatory obligations vary by jurisdiction and can change over time. Fifth, risk tiers and maturity levels

require prospective testing for usability, inter-rater agreement, workload, and ability to reduce nonconformities or patient-safety events (31).

Conclusion

Clinical AI in medical laboratories requires more than a one-time accuracy study. It requires a documented decision about intended use and regulatory route, locally appropriate verification or validation, controlled installation and operation, measurement-aware input governance, risk-proportionate monitoring, incident/CAPA integration, and disciplined change control. This article provides a proposed ISO 15189-aligned control matrix, evidence package, risk-tiered maturity roadmap, and worked monitoring artefact to support that process. The framework is not a substitute for ISO 15189, device regulation, or prospective validation; it is intended to make those obligations easier to organise and review within an existing laboratory QMS.

Funding

No external funding was received.

Ethical approval

Not applicable (standards-based framework development; no patient data).

Authors' contributions

B.S.A. conceived the study, developed the framework, and drafted the manuscript.

Acknowledgments

The author would like to thank the Deanship of Scientific Research at Shaqra University for supporting this work.

This acknowledgment does not indicate financial support and does not change the funding statement of the article.

Conflict of interest statement

All the authors declare that they have no conflict of interest in this work.

References

1. International Organization for Standardization. ISO 15189: 2022 Medical laboratories – Requirements for quality and competence. Geneva: ISO; 2022.
2. Master SR, Badrick T, Bietenbeck A, Cervinski MA, Haymond S. Machine learning in laboratory medicine: recommendations of the IFCC Working Group. *Clin Chem* 2023; 69(7): 690–8. doi:10.1093/clinchem/hvad055.
3. Garcia CA, Reed KA, Lantz E, Day P, Zarella MD, Hart SN, et al. Establishing a comprehensive artificial intelligence lifecycle framework for laboratory medicine and pathology: a series introduction. *Am J Clin Pathol* 2025; 164(3): 424–37. doi:10.1093/ajcp/aqaf069.
4. Lennerz JK, Marquardt A, Genzen JR, Haymond S. Diagnostic quality model (DQM): an integrated framework for the assessment of diagnostic quality when using artificial intelligence/machine learning. *Clin Chem Lab Med* 2023; 61(9): 1390–402. doi:10.1515/cclm-2022-1151.
5. Spies NC, Farnsworth CW, Wheeler S, McCudden CR. Validating, implementing, and monitoring machine learning solutions in the clinical laboratory safely and effectively. *Clin Chem* 2024; 70(11): 1334–43. doi:10.1093/clinchem/hvae126.
6. Yang HS, Çubukçu HC, Ben FD, et al. Implementation of artificial intelligence systems in the clinical laboratory: insights from an expert survey and recommendations for best practice. *Clin Chem Lab Med* 2026. doi:10.1515/cclm-2025-1322.
7. Hanna MG, et al. Recommendations for performance evaluation of machine learning in pathology: a concept paper from the College of American Pathologists. *Arch Pathol Lab Med* 2024; 148(10): e335–e361. doi:10.5858/arpa.2023-0042-CP.
8. Çubukçu HC, Ben FD, et al. Regulating the future of laboratory medicine: European regulatory landscape of AI-driven medical device software. *Clin Chem Lab Med* 2025. doi:10.1515/cclm-2025-0482.
9. Owoyemi O, AlBarakati A, Alsaeed S, Shriee A, Alhilal A. Sociotechnical checklist for clinical AI implementation in healthcare and hospital settings. *JMIRx Med* 2023; 4: e46954. doi:10.2196/46954.
10. Saudi Accreditation Center (SAAC). SAAC transition plan for ISO 15189:2022. Riyadh; 2023.
11. Saudi Food and Drug Authority (SFDA). Guidance on AI and ML technologies based medical devices (MDS-G-010). Riyadh; 2022.
12. Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S. A design science research methodology for information systems research. *J Manag Inf Syst* 2007; 24(3): 45–77.
13. International Organization for Standardization. ISO 22367:2026 Medical laboratories – Application of risk management to medical laboratories. Geneva: ISO; 2026.
14. International Organization for Standardization. ISO/IEC 23894:2023 Information technology – Artificial intelligence – Guidance on risk management. Geneva: ISO; 2023.
15. International Organization for Standardization. ISO/IEC 42001:2023 Information technology – Artificial intelligence – Management system. Geneva: ISO; 2023.
16. National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. Gaithersburg (MD): NIST; 2023.
17. International Medical Device Regulators Forum (IMDRF). Good machine learning practice for medical device development: guiding principles. IMDRF/AIML WG/N88. 2025.
18. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: WHO; 2021.
19. Liu X, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: CONSORT-AI extension. *Nat Med* 2020; 26: 1364–74. doi:10.1038/s41591-020-1034-x.
20. Cruz Rivera S, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: SPIRIT-AI extension. *Nat Med* 2020; 26: 1351–63. doi:10.1038/s41591-020-1037-7.
21. Vasey B, et al. DECIDE-AI: an evaluation guideline for early-stage clinical evaluation of decision support systems driven by artificial intelligence. *Nat Med* 2022; 28: 924–33. doi:10.1038/s41591-022-01773-1.
22. Collins GS, et al. TRIPOD+AI: updated guidance for reporting clinical prediction models that use artificial intelligence. *BMJ* 2024; 385: e078378. doi:10.1136/bmj-2023-078378.
23. Lorde N, Mahapatra S, Kalaria T. Machine learning for patient-based real-time quality control: analytical stability assessment for clinical chemistry. *Clin Chim Acta* 2021; 517: 132–9. doi:10.1016/j.cca.2021.02.027.
24. Badrick T, St John A, Sikaris K, et al. Patient-based real-time quality control: a review. *Clin Biochem Rev* 2018; 39(2): 45–52.
25. Granlund T, et al. Toward a regulatory-compliant lifecycle for AI-based medical devices in the EU. *IEEE Comput* 2024; 57(2): 28–36.
26. European Parliament and Council. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). OJ L 2024/1689, 12 July 2024; consolidated text current 27 July 2026.

27. European Parliament and Council. Regulation (EU) 2017/746 on in vitro diagnostic medical devices (IVDR). OJ L 117, 5 May 2017: 176–332, as amended.
28. International Organization for Standardization. ISO 5649:2024 Medical laboratories – Concepts and specifications for the design, development, implementation and use of laboratory-developed tests. Geneva: ISO; 2024.
29. International Organization for Standardization. ISO/DIS 24051-1 Medical laboratories – Part 1: General principles for the application of artificial intelligence in medical laboratories. Draft International Standard, under development; 2026.
30. International Organization for Standardization. ISO/CD 24051-2 Medical laboratories – Part 2: Digital pathology and artificial intelligence (AI)-based image analysis. Committee Draft, under development; 2026.
31. International Organization for Standardization/International Electrotechnical Commission. ISO/IEC 17000:2020 Conformity assessment – Vocabulary and general principles. Geneva: ISO; 2020.

Received: July 18, 2026

Accepted: August 17, 2026

ONLINE FIRST