Journal of Engineering, Management and Information Technology

ASPUR PUBLISHING

Vol. 05, No. 01 (2027) 57-70, doi:10.61552/JEMIT.2027.01.006 - http://jemit.aspur.rs

AI-DRIVEN ANOMALY DETECTION FOR VALIDATED GXP CLOUD INFRASTRUCTURE MONITORING

Rajashekar Reddy Gujjula [1]

Received 11.08.2026.
Revised 24.09.2026.
Accepted 17.11.2026.

Keywords:

GxP compliance, anomaly detection, cloud infrastructure monitoring, change control, audit trail, machine learning operations, computerised system validation, data integrity.

Original research

A B S T R A C T

Cloud platforms that host GxP-regulated operations increasingly use machine learning anomaly detectors to monitor validated infrastructure, yet these detectors are seldom brought under the change-control and qualification discipline applied to the systems they watch. This article argues that a detector whose output triggers or suppresses an alert leading to a change-control, incident or deviation record on a validated system is part of that system's control loop, and that its lifecycle must therefore produce reviewable evidence. We separate what the regulatory sources require (a validated state, controlled configuration change, and audit trails for GxP-relevant records) from what this article proposes as practice (training-data provenance records, alert-to-audit-trail binding, and a promotion gate for detector versions). We then present a validation-aware architecture of four layers: evidence capture, a change-control gate, a human qualification checkpoint, and alert-to-audit-trail binding. The architecture is assessed with an explicit evaluation framework: eight criteria, each traced to a regulatory source or flagged as proposed practice, scored on a four-level maturity rubric, and supplemented by three audit-query traceability scenarios and a failure-mode analysis. The assessment is analytical, not empirical. It shows which criteria the design can satisfy by technical enforcement, which remain procedural, and where residual risk persists, most notably the tension between continuous retraining and controlled change. A protocol for the quantitative validation that this study does not provide is specified for future work.

© 2027 Journal of Engineering, Management and Information Technology

1. INTRODUCTION

Machine learning anomaly detectors are now routine in cloud infrastructure monitoring, including in cloud platforms that host workloads subject to Good Practice (GxP) requirements, the umbrella term for good manufacturing, laboratory, clinical and related practice rules. When such a detector's output triggers an alert, suppresses one, or initiates a change-control, incident or deviation workflow against a validated system, the detector takes part in that system's control loop, whether or not it was documented as doing so. The question this article addresses is what evidence the detector's own lifecycle must then produce.

The scope is deliberately narrow. We consider computerised systems, and the cloud infrastructure that supports them, in regulated pharmaceutical manufacturing and related quality-system contexts, and we treat the anomaly detector as a software component. Clinical, laboratory and regulatory-submission matters are outside the scope. The perspective is that of platform engineering: the author's background is in cloud landing

[1] Corresponding author: Rajashekar Reddy Gujjula
Email: rajashekargujjular@gmail.com

zones, infrastructure-as-code, and policy-as-code and role-based access control (RBAC) guardrails for workloads that must remain demonstrably compliant. That background informs the design, but it is not offered as evidence that the design works; the claims made here are limited to what the evaluation in Sections 4 and 6 can support.

Regulation in this area is not silent, but it is not specific. Existing rules require GMP computerised systems to be validated, configuration changes to be controlled, and GMP-relevant record changes to be captured in audit trails (European Commission, 2011; U.S. Food and Drug Administration, n.d.a; U.S. Food and Drug Administration, n.d.b). None of these instruments mentions anomaly detectors. The first AI-specific EU GMP text, draft Annex 22, was released for public consultation in 2025 and should be read as draft guidance until it is adopted (European Commission, & Pharmaceutical Inspection Co-operation Scheme, 2025). Whether a given detector falls inside a validated system's boundary is therefore an interpretation to be settled by documented risk assessment (European Commission, 2011; International Council for Harmonisation, 2023); it cannot be read off a clause. This article makes that interpretation explicit (Section 2.3) and keeps three things strictly apart throughout: regulatory requirements, authoritative guidance, and the practices proposed here.

General-purpose anomaly detection is designed for accuracy and speed, whether it is statistical, supervised, or based on autoencoders and Long Short-Term Memory (LSTM) networks (Chandola et al., 2009; Hundman et al., 2018; Xu et al., 2018). Those goals differ from auditability, and a detector optimised for the former does not automatically produce the records a quality reviewer would need.

The contributions are as follows. First, a three-tier classification that separates regulatory requirements, authoritative guidance and proposed practice, applied to every element of the design (Section 2.3 and Table 3). Second, a validation-aware architecture of four layers, with an explicit applicability condition that states when the detector falls inside the validated boundary (Sections 2.3 and 5). Third, an evaluation framework of eight criteria with a four-level maturity rubric, three audit-query traceability scenarios and a failure-mode analysis (Sections 4 and 6). Fourth, a discussion of the tension between retraining and change control that takes draft Annex 22's treatment of dynamic models into account, together with a protocol for the quantitative validation that this analytical study does not provide (Section 7). Section 2 gives the background and regulatory landscape, Section 3 reviews related work, and Section 8 concludes.

2. BACKGROUND

2.1 Validated state and change control

Under GMP, a qualified computerised system is expected to remain in a validated state throughout its life, and changes to it, including changes to system configuration, are made in a controlled manner under a defined procedure (European Commission, 2011; European Commission, 2015). EU Annex 11 states this for configuration explicitly in its clause on change and configuration management, and requires periodic evaluation to confirm that the system remains valid (European Commission, 2011). US regulations likewise require validation of systems to ensure accuracy, reliability and consistent intended performance, controls over documentation changes, and controls ensuring that changes to production and control records are made only by authorised personnel (U.S. Food and Drug Administration., n.d.a; U.S. Food and Drug Administration, n.d.b).

Where a monitoring threshold defines the acceptable state of a validated system, that threshold is part of the system's configuration, and changing it is a configuration change, not a tuning exercise. It needs a documented request, an impact assessment, an approval record and evidence that the new value was verified before release. Whether a particular threshold belongs to the validated configuration is a risk-based determination (European Commission, 2011; International Council for Harmonisation, 2008; International Council for Harmonisation, 2023); we call the outcome the applicability condition and define it in Section 2.3. When an inspector or quality reviewer asks why an alert fired, or why it did not, the answer should trace to a specific approved configuration state and not to whatever model happened to be running that week (Pharmaceutical Inspection Co-operation Scheme, 2007; U.S. Food and Drug Administration, 2018).

2.2 Cloud monitoring and detection families

A modern landing zone generates telemetry at a volume and velocity that manual, rule-based monitoring struggles to match: utilisation metrics, network flow logs, policy-compliance signals and application traces arriving continuously across many services. Static thresholds, set by hand against expected baselines, degrade as workloads scale and legitimate operating patterns shift. This is the practical pressure that brings machine learning into infrastructure monitoring.

Statistical methods form the oldest layer of the response: they characterise a normal distribution from historical telemetry and flag departures beyond a bound, which is simple to reason about but weak against gradual or multivariate shifts (Chandola et al., 2009). Supervised approaches gain accuracy where labelled anomalies exist, but regulated environments rarely produce labelled datasets cleanly, since true anomalies are rare and often environment-specific. Unsupervised methods, including autoencoder-based approaches, learn a compressed representation of normal behaviour and flag inputs that reconstruct poorly (Xu et al., 2018); this sidesteps labelling but makes the model's notion of normal a function of its training data. LSTM-based sequence models extend this to temporal patterns, learning how a

metric should evolve and flagging deviations from the learned trajectory (Hundman et al., 2018).

Modern MLOps platforms already provide components such as model registries and metadata stores (Kreuzberger et al., 2023), so the gap addressed here is not an absence of lineage data. It is the absence of a GxP-oriented binding between that data, change control and the audit trail. As commonly configured, the techniques above are not designed to produce the three things a validated-state monitoring programme needs: a record of what data trained the model, a controlled process for approving changes to its thresholds or retraining schedule, and a durable link between a specific alert and the audit-trail entry that resulted from it.

2.3 Regulatory landscape: requirement, guidance and proposal

To keep the argument auditable, every control in this article is assigned to one of three tiers. Tier 1, regulatory requirement, covers provisions in binding regulation, or in the official GMP guidelines that inspectors apply as the inspection standard: 21 CFR Parts 11 and 211 (U.S. Food and Drug Administration, n.d.a; U.S. Food and Drug Administration, n.d.b) and EudraLex Volume 4 Annexes 11 and 15 (European Commission, 2011; European Commission, 2015). Tier 2, authoritative guidance, covers non-binding regulator, international or industry guidance that expresses accepted practice: ICH Q9(R1) and Q10 (International Council for Harmonisation, 2008; International Council for Harmonisation, 2023), PIC/S PI 011-3 (Pharmaceutical Inspection Co-operation Scheme, 2007), the FDA data-integrity guidance (U.S. Food and Drug Administration, 2018), GAMP 5 Second Edition (International Society for Pharmaceutical Engineering, 2022) and the GAMP Guide on Artificial Intelligence (International Society for Pharmaceutical Engineering, 2025). Draft regulatory text, namely draft Annex 22 (European Commission, & Pharmaceutical Inspection Co-operation Scheme, 2025), is placed in Tier 2 with an explicit draft marker because it signals direction but does not yet bind. Tier 3, proposed practice, covers design elements this article proposes as a reasonable way to apply Tier 1 and Tier 2 principles to machine learning detectors. Tier 3 elements are not regulatory requirements, and we do not claim that an inspector would demand them. Table 1 summarises the instruments and their status.

Three points follow from the Tier 1 sources. First, validation, controlled configuration change and periodic evaluation of GMP computerised systems are requirements, and thresholds that form part of the validated configuration fall within them (European Commission, 2011). Second, audit trails are required for GMP-relevant record creation, modification and deletion: Annex 11 frames this as a risk-based consideration (European Commission, 2011), and 21 CFR 11.10(e) requires secure, time-stamped audit trails for electronic records within Part 11's scope (U.S. Food and Drug Administration, n.d.a). An alert record that is not itself a GMP record is therefore outside the literal audit-trail requirement, although a ticket or deviation record that the alert creates is inside it, so binding the two is a Tier 3 proposal. Third, none of the Tier 1 sources specifies requirements for training-data provenance or for retraining machine learning models. Such requirements can only be inferred from the general expectation that systems are specified, controlled and traceable (Tier 2) and, in draft form, from Annex 22.

Table 1. Regulatory and guidance instruments referenced in this article, their status and tier.

Instrument	Issuer and status	Relevance to detectors	Tier
21 CFR Part 11 (U.S. Food and Drug Administration, n.d.a).	US FDA; binding regulation	Validation of systems (11.10(a)); secure, time-stamped audit trails for electronic records (11.10(e)); revision and change control of system documentation (11.10(k))	1
21 CFR Part 211, §211.68 (U.S. Food and Drug Administration, n.d.b).	US FDA; binding regulation	Controls over computer systems so that changes to master production and control records are made only by authorised personnel	1
EudraLex Vol. 4, Annex 11 (European Commission, 2011).	European Commission; EU GMP guideline applied by inspectors	Risk management; validation; audit trails (cl. 9); change and configuration management (cl. 10); periodic evaluation (cl. 11); data storage and archiving (cl. 7, 17)	1
EudraLex Vol. 4, Annex 15 (European Commission, 2015).	European Commission; EU GMP guideline applied by inspectors	Qualification and validation; change control of qualified systems	1
ICH Q9(R1), Q10 (International Council for Harmonisation, 2008; International Council for Harmonisation, 2023)	ICH; guidance adopted by regulators	Quality risk management; change management within the pharmaceutical quality system	2
PIC/S PI 011-3 (Pharmaceutical Inspection Co-operation Scheme, 2007).	PIC/S; inspectorate guidance	Good practices for computerised systems in GxP environments	2

Instrument	Issuer and status	Relevance to detectors	Tier
FDA data-integrity Q&A (U.S. Food and Drug Administration, 2018)	US FDA; non-binding guidance	Data-integrity expectations for CGMP records	2
GAMP 5 (2nd ed.) (International Society for Pharmaceutical Engineering, 2022); GAMP AI Guide (International Society for Pharmaceutical Engineering, 2025).	ISPE; industry guidance	Risk-based lifecycle for computerised systems and for AI-enabled systems	2
Draft Annex 22 and draft revised Annex 11 (European Commission, & Pharmaceutical Inspection Co-operation Scheme, 2025).	European Commission with PIC/S; draft released for consultation July–October 2025; current status to be confirmed by the reader	Expectations for AI in GMP applications; dynamic models not for critical applications	2 (draft)
FDA PCCP guidance (U.S. Food and Drug Administration, 2024)	US FDA; final guidance for AI-enabled medical device software	Analogy only: pre-specified bounds for changes to adaptive models; not a GxP infrastructure requirement	Analogy
Controls proposed in this article	Author's proposal	Training-data provenance records; alert-to-audit-trail binding; promotion gate and qualification checkpoint for detector versions	3

AI-specific texts are emerging. The GAMP Guide on Artificial Intelligence applies lifecycle thinking to AI-enabled GxP systems (International Society for Pharmaceutical Engineering, 2025). Draft Annex 22 would exclude dynamic, continuously learning models from critical GMP applications and set expectations for static models (European Commission, & Pharmaceutical Inspection Co-operation Scheme, 2025). Both are directed at AI used in GMP applications in general. Whether a detector that monitors supporting infrastructure is a critical GMP application is a classification to be made by risk assessment, and we do not assume the answer; Section 7.2 discusses both outcomes.

Applicability condition. A detector is treated as inside the validated boundary, and the architecture of Section 5 applies, when a documented risk assessment finds that at least one of the following holds: (a) its output triggers or suppresses an alert that initiates a change-control, incident, deviation or corrective-and-preventive-action record for a GxP system; (b) its output is used as evidence that a validated state is being maintained, for example in periodic review; or (c) it can change the configuration of a validated system automatically. Where none holds, the detector is operational tooling, and the architecture is optional good practice.

3. LITERATURE REVIEW

The anomaly detection literature and the GxP compliance literature have developed largely without reference to each other, and the seams between them are where the gap described here becomes visible. Three tensions organise this review: general-purpose cloud anomaly detection against validated-infrastructure requirements, static thresholds against adaptive retrained models, and detector alert volume against audit-trail obligations.

Fit for purpose. Foundational surveys treat detection effectiveness and applicability across domains as the central concerns of the field (Chandola et al., 2009), and a systematic review of machine learning for anomaly detection organises the literature by technique and application domain (Nassif et al., 2021; U.S. Food and Drug Administration, 2024); to our reading, neither treats regulatory auditability as an organising dimension. AIOps research extends these concerns to cloud scale. Work on node failure prediction in a very large cloud platform shows that machine learning can anticipate failures before they manifest, and reports that adoption requires more than a top-performing model, since the model must also be trustable, interpretable, maintainable, scalable and evaluated in context (Li et al., 2020). A survey of AIOps methods for failure management catalogues techniques without, to our reading, treating regulatory qualification as a design constraint (Notaro et al., 2021). This is not a flaw in that literature, because AIOps was not built for regulated environments, but practitioners importing these techniques into GxP infrastructure inherit a design philosophy that never considered qualification evidence. Work on trustworthy AI synthesises requirements such as explainability, accountability and safety for AI systems in general (Kaur et al., 2023), and a systematic review of validation methods for AI systems catalogues techniques for verifying AI behaviour before deployment (Myllyaho et al., 2021). Neither was written with pharmaceutical change control in mind, but both supply vocabulary and method that transfer. Industry guidance meets GxP most directly: the GAMP AI Guide and draft Annex 22 address AI-enabled GxP systems in general (European Commission, & Pharmaceutical Inspection Co-operation Scheme, 2025; International Society for Pharmaceutical Engineering, 2025), but to our reading neither specifies

the evidence trail that connects a monitoring detector's alerts to infrastructure change-control records.

Static thresholds against adaptive models. Concept-drift research shows that models trained on historical data degrade as the data distribution shifts, so learning systems must adapt to stay accurate in non-stationary environments (Lu et al., 2019). MLOps research describes automated retraining and continuous delivery pipelines as typical components of production ML systems (Kreuzberger et al., 2023) and organises MLOps activities into a taxonomy and methodology covering the model lifecycle (Testi et al., 2022). This is where the tension bites: automatic redeployment of an updated detector, common in MLOps practice, conflicts with a change-control regime that requires each threshold or detection-logic change to be reviewed and approved before it reaches a validated system. In medical imaging, data drift denotes differences between the data used to train a model and the data it meets in operation, and the literature discusses its monitoring and remedies (Sahiner et al., 2023). A scoping review of FDA-approved AI medical devices found that only 46.1% provided comprehensive performance study results and only 9.0% included a prospective post-market surveillance study (Muralidharan et al., 2024). That domain lies outside cloud infrastructure, but it suggests, by analogy, that post-deployment drift monitoring of adaptive models is under-developed even where regulatory scrutiny is high. In the same domain, regulators have formalised one way of reconciling adaptive models with approval gates, the predetermined change control plan, in which permissible modifications and the protocol for validating them are specified in advance (U.S. Food and Drug Administration, 2024); a published case applies such a plan to an AI-based diagnostic device (Wall et al., 2023). These sources address medical devices, not cloud infrastructure, and we cite them as an analogy for bounding change, not as a requirement for GxP infrastructure.

Alert volume against audit-trail obligations. Autoencoder-based unsupervised detection for seasonal key performance indicators in web applications reports strong detection results on production-scale data (Xu et al., 2018). A comprehensive evaluation of 71 time-series anomaly detection algorithms on 976 datasets reports that effectiveness varies by dataset and identifies a simple wavelet-based method, DWT-MLEAD, as having the best cost-benefit ratio, with an average AUC-ROC of 83% and a runtime of 2.2 ms per data point (Schmidl et al., 2022). What the alert-fatigue literature disputes is whether high-volume detection is useful once it reaches a human analyst: work on security operations centres reports that high alert rates and volumes increase analyst cognitive burden and contribute to burnout (Tariq et al., 2025), and root-cause-analysis research in cloud data centres warns that inadequate modelling of problem patterns can produce alert or event storms (Poghosyan et al., 2021). Alert-reduction research responds with correlation-based filtering before alerts reach a reviewer (Yu & Zhang, 2023), and deep-learning methods for log anomaly detection are surveyed separately (Landauer et al., 2023). None of this work asks the question GxP audit readiness raises: for the alerts that remain, can we produce a durable record of what data and threshold state produced this specific alert, and can that record survive an audit.

An industrial account from IBM Cloud comes closest to operational reality. It describes a deep-learning detector for near-real-time monitoring of cloud components on a platform of more than 60 data centres and over 170 managed services, reports 99.95% accuracy on a VM database-server dataset while noting that accuracy is a weak measure for anomaly detection, and reports, on the basis of DevOps team feedback, that the detector can capture anomalies up to 20 minutes earlier than the system it replaced (Islam et al., 2021). The account is framed around detection and operations and does not address qualification evidence. We make no claim that these figures would transfer to a GxP environment; the point is that if detection of this kind is feasible in production, the open problem for regulated deployment is qualification evidence, not detection capability.

Read together, the three tensions point to the same conclusion. The field has optimised for detection accuracy, adaptive retraining and alert-volume reduction, each a legitimate engineering goal, without building the qualification scaffolding that would let these techniques operate inside a validated-state control loop. Table 2 positions the approaches reviewed here against that scaffolding.

Table 2. Positioning of reviewed approaches against qualification evidence (criteria C1–C8, defined in Section 4). Entries on the reviewed work reflect the authors' reading of what each source addresses.

Approach	Primary purpose	Qualification evidence (C1–C8)	Alert-to-audit binding (C6)
General-purpose AIOps anomaly detection (Islam et al., 2021; Li et al., 2020; Notaro et al., 2021)	Detect or predict infrastructure failures	Not addressed as a design constraint	Not addressed
Detection algorithms and benchmarks (Chandola et al., 2009; Hundman et al., 2018; Schmidl et al., 2022; Xu et al., 2018)	Detection accuracy, efficiency	None	None

Approach	Primary purpose	Qualification evidence (C1–C8)	Alert-to-audit binding (C6)
Alert reduction and log-based detection (Landauer et al., 2023; Poghosyan et al., 2021; Tariq et al., 2025; Yu & Zhang 2023)	Reduce alert volume; detect anomalies in logs	None	Not addressed; reduces volume and can add unrecorded suppression (Table 7, F5)
MLOps frameworks (Kreuzberger et al., 2023; Testi et al., 2022)	Automate the ML lifecycle; version models	Partial: lineage and registry components can support C1 and C2	None
Medical-device change-control plans (U.S. Food and Drug Administration, 2024; Wall et al., 2023)	Bound changes to adaptive models	Partial analogue of C3 and C4, in a different domain	Not applicable
GAMP AI Guide; draft Annex 22 (European Commission, & Pharmaceutical Inspection Co-operation Scheme, 2025; International Society for Pharmaceutical Engineering, 2025)	Lifecycle and GMP expectations for AI-enabled systems	General lifecycle control of AI in GxP; detector-specific evidence trail not specified	Not specified
Proposed architecture (this article)	Evidence for detectors inside the validated boundary	C1–C8 at Levels 2–3, design-level (Table 4)	By design, subject to reconciliation (Table 7, F1)

4. METHODOLOGY: EVALUATION FRAMEWORK

4.1 Study design and epistemic status

This is a design-science style analytical study. Its artefact is the architecture of Section 5. Its evaluation is analytical: a criterion-by-criterion conformance analysis against a defined rubric, scenario walkthroughs, and a failure-mode analysis of the architecture's own controls. No dataset, experiment or regulatory audit was involved, and no detection-performance claim is made. We state this so that the results in Section 6 are not read as measurements.

4.2 Derivation of criteria

Criteria were derived in three steps. First, provisions in the Tier 1 and Tier 2 sources of Table 1 that relate to validation, configuration change control, audit trails and data integrity, and periodic review were extracted and mapped to the stages of the detector lifecycle: training and retraining, threshold governance, alert consequence, and qualification and review. Second, each provision was restated as a criterion testable on lifecycle evidence, that is, as a question that a specific record can answer. Third, each criterion was labelled with the strongest source supporting it. Where a criterion applies a Tier 1 principle to a detector-specific artefact such as a model file, the label is *Requirement (applied)*, meaning that the requirement applies once the applicability condition of Section 2.3 is met and that its application to the detector is by interpretation. Criteria resting on Tier 2 sources are labelled *Guidance*, and criteria with no direct source text are labelled *Proposed*. This yields eight criteria (Table 3).

4.3 Maturity rubric

Each criterion is scored on a four-level rubric. Level 0, absent: no record or control exists. Level 1, ad hoc: records or controls exist for some cases but are not systematically produced or are not linked to the audit trail. Level 2, defined: a procedure or record schema exists and is applied by people, but nothing technically prevents or detects an omission. Level 3, enforced: the control is technically enforced or the record is generated automatically, and omissions are detectable by an independent check. A criterion is treated as satisfied only at Level 3. We set this threshold because the failure the criteria guard against, an unnoticed omission, is by definition not prevented by procedure alone.

4.4 Conditions compared

Two conditions are compared. The *baseline* is a reference condition defined for this analysis: a detector deployed with the default capabilities of common monitoring and MLOps tooling, which may include a model registry and metadata store (Kreuzberger et al., 2023), and with no GxP-specific controls added. It is not an empirical sample of organisations. Baseline levels are stated as explicit assumptions in Table 4 so that readers can substitute their own. The *proposed* condition is the architecture of Section 5 assuming correct implementation, so its levels are design-level targets, which the failure-mode analysis of Section 7 then qualifies.

4.5 Traceability scenarios and failure-mode analysis

Three audit-query scenarios test whether an auditor's question can be answered from records alone, without relying on personnel recollection. S1 asks why a specific alert fired. S2 asks why an expected alert did not fire, that

is, whether it was suppressed and under which model and threshold state. S3 asks which threshold was in effect at a given time, who approved it and on what evidence. A scenario is scored *pass* when the answer can be reconstructed from the evidence store and the linked systems of record with every link resolvable, *partial* when part of the answer is recoverable, and *fail* otherwise. Following the risk-based approach of ICH Q9(R1) (International Council for Harmonisation, 2023), we also identify failure modes of the proposed controls themselves and assess them qualitatively for consequence, detection and residual risk. No numerical risk priority numbers are assigned, because no failure-rate data exist.

4.6 Threats to validity

Construct validity is limited because the criteria are the author's interpretation of the sources, and several sources are industry guidance, not regulation; the tier labels exist to make that visible. Internal validity is limited because a single analyst assigned all levels, without an inter-rater check. External validity is limited because the sources are US and EU texts and the baseline is a defined reference condition, not an observed population. Conclusion validity is limited because the study is analytical and offers no quantitative support. Section 7.4 specifies a protocol that addresses each of these

Table 3. Evaluation criteria, their regulatory basis and status. Detector lifecycle stages: A training and retraining, B threshold governance, C alert consequence, D qualification and review.

ID	Criterion (evidence question)	Basis	Status
C1 (A)	Is the training-data source, extraction window and filtering recorded for each deployed model version?	Traceability and data-integrity expectations (International Society for Pharmaceutical Engineering, 2022; Pharmaceutical Inspection Co-operation Scheme, 2007; U.S. Food and Drug Administration, 2018); AI lifecycle guidance (International Society for Pharmaceutical Engineering, 2025); draft Annex 22 (European Commission, & Pharmaceutical Inspection Co-operation Scheme, 2025).	Guidance
C2 (A)	Can the deployed model and its code be uniquely identified (version, hash) and tied to its training run?	Configuration identification and control (European Commission, 2011); control of system documentation (U.S. Food and Drug Administration, n.d.a)	Requirement (applied)
C3 (B)	Does every threshold, model-version or retraining-trigger change pass an approval gate before reaching production?	Change and configuration management (European Commission, 2011); authorised changes to records (U.S. Food and Drug Administration, n.d.b); change control (European Commission, 2015); change management (International Council for Harmonisation, 2008).	Requirement (applied)
C4 (B)	Is an impact assessment performed and the change verified before release?	Validation and qualification (European Commission, 2011; European Commission. 2015), risk-based assessment (International Council for Harmonisation, 2023).	Requirement (applied)
C5 (B)	Does an emergency path exist that is flagged and retrospectively reviewed, not silent?	Emergency change practice (International Society for Pharmaceutical Engineering, 2022); change management (International Council for Harmonisation, 2008)	Guidance
C6 (C)	Does a durable record bind each alert, or suppressed alert, to the audit-trail entry it produced?	Audit trails apply to the GMP-relevant records the alert creates (European Commission, 2011; U.S. Food and Drug Administration, n.d.a); the cross-system binding itself has no direct source text	Proposed
C7 (C)	Are detection records tamper-evident, retained and retrievable?	Audit trails and record integrity (U.S. Food and Drug Administration, n.d.a); data storage and archiving (European Commission, 2011)	Requirement (applied) for GMP records; proposed for other alert records
C8 (D)	Is a documented qualification review completed before promotion, and is the model periodically evaluated in operation?	Qualification (European Commission, 2015); validation and periodic evaluation (European Commission, 2011); drift-specific criteria are proposed	Requirement (applied) for review; proposed for drift criteria

5. PROPOSED ARCHITECTURE

The architecture treats the anomaly detector as a governed component of the validated system and not as an external tool observing it. It layers four capabilities onto an existing detection pipeline (Figure 1), in the order in which a detector version moves through its lifecycle. Each layer is designed to sit alongside standard cloud-

native tooling and not to replace it, and each is mapped to the evaluation criteria and to its regulatory anchors in Table 4. The layers are Tier 3 designs: the requirements they help to satisfy are in Tier 1 and Tier 2, but the layers themselves are proposals.

Layer 1: evidence capture. Every model version deployed to production carries a provenance record written automatically by the training pipeline, not reconstructed afterwards. The record contains the training dataset's source identifier, extraction window and filter definitions; a summary of the training data's coverage; the training run's configuration and code version, tracked as infrastructure-as-code modules are; a cryptographic hash of the resulting model artefact; the identity of the pipeline or operator; and a timestamp. Where retraining is triggered automatically in response to detected drift, as is common in MLOps practice (Kreuzberger et al., 2023), the layer logs the trigger condition and the difference between the new and previous model alongside the provenance record, so that a reviewer can see why the model changed as well as what changed. Promotion is blocked when a mandatory field is empty.

Layer 2: change-control gate. Any change to a detection threshold, whether a manually tuned static bound or a parameter inside a retrained model, must pass an approval workflow before promotion. The gate is a policy-as-code check enforced at deployment time: a proposed change is submitted with a justification and an impact assessment, routed to an approver other than the change's author, and released to the production monitoring pipeline only after approval and after the verification evidence is attached. Access controls remove direct deployment paths around the gate. Emergency changes are permitted but flagged and automatically placed in a queue for retrospective review, and an overdue review raises an alert; they do not bypass the gate silently.

Layer 3: human qualification checkpoint. No model version moves from staging to production monitoring of validated infrastructure without a documented qualification review. The review confirms that the provenance record is complete, that the change-control record is closed, that the alert-binding mechanism has been verified against a test dataset, and that detection performance on a held-out dataset independent of the training data meets acceptance criteria specified in advance and approved through Layer 2 (European Commission, & Pharmaceutical Inspection Co-operation Scheme, 2025; International Society for Pharmaceutical Engineering, 2025). The checkpoint is a gate, not a dashboard: dashboards can be ignored and gates cannot. It is analogous to the evidence package required before any other validated-system change goes live (European Commission, 2015), adapted to the artefacts a machine learning detector produces.

Layer 4: alert-to-audit-trail binding. Every alert generated by the detector, and every case in which the detector's output suppressed an alert that a static threshold would otherwise have raised, is written as a structured, append-only record. The record contains a unique alert identifier; a UTC timestamp from a synchronised time source; the model artefact hash; the identifier of the approved threshold state; the identifiers of the input data window; the decision (fired or suppressed); a reference to the downstream audit-trail entry, whether a change-control ticket, an incident record or a documented no-action decision; and a hash chaining the record to its predecessor. The persistent identifier is carried across both the detection pipeline and the system of record. The binding is a data contract between the pipeline and the systems that already generate audit-relevant documentation, not a new dashboard. Because a dropped identifier would otherwise fail silently, a periodic reconciliation control checks that every alert in the detector log has a binding record and that every binding record resolves in the system of record, and reports the orphan rate. Alert-reduction layers upstream must write their suppression events to the same store.

The architecture does not change the detection algorithm and makes no claim to change detection performance.

Table 4. Architecture layers mapped to evidence produced, evaluation criteria and regulatory anchors.

Layer	Function	Evidence produced	Criteria	Regulatory anchors (tier)
L1 Evidence capture	Records provenance automatically in the training pipeline; blocks promotion on missing fields	Dataset source, extraction window, filters, code version, model hash, retraining trigger and model difference	C1, C2, C3	Annex 11 cl. 10 and 21 CFR 11.10(k) for configuration identity (Tier 1); data-integrity guidance (Tier 2)
L2 Change-control gate	Routes any threshold or model-version change through an approval workflow; flags emergency changes	Change request, justification, impact assessment, approval record, emergency-change flag and review record	C3, C4, C5	Annex 11 cl. 10, Annex 15, 21 CFR 211.68(b) (Tier 1); ICH Q10, GAMP 5 (Tier 2)
L3 Qualification checkpoint	Gates promotion from staging to production monitoring of validated infrastructure	Documented qualification review; held-out performance against pre-specified acceptance criteria	C8 (also C4)	Annex 15, Annex 11 cl. 4 and 11 (Tier 1); draft Annex 22 and GAMP AI Guide (Tier 2)
L4 Alert-to-audit binding	Binds each alert or suppressed alert to its downstream audit-trail entry; reconciles identifiers	Append-only, hash-chained binding record; reconciliation report	C6, C7	Annex 11 cl. 7, 9, 17 and 21 CFR 11.10(e) for GMP records (Tier 1); the cross-system binding is Tier 3

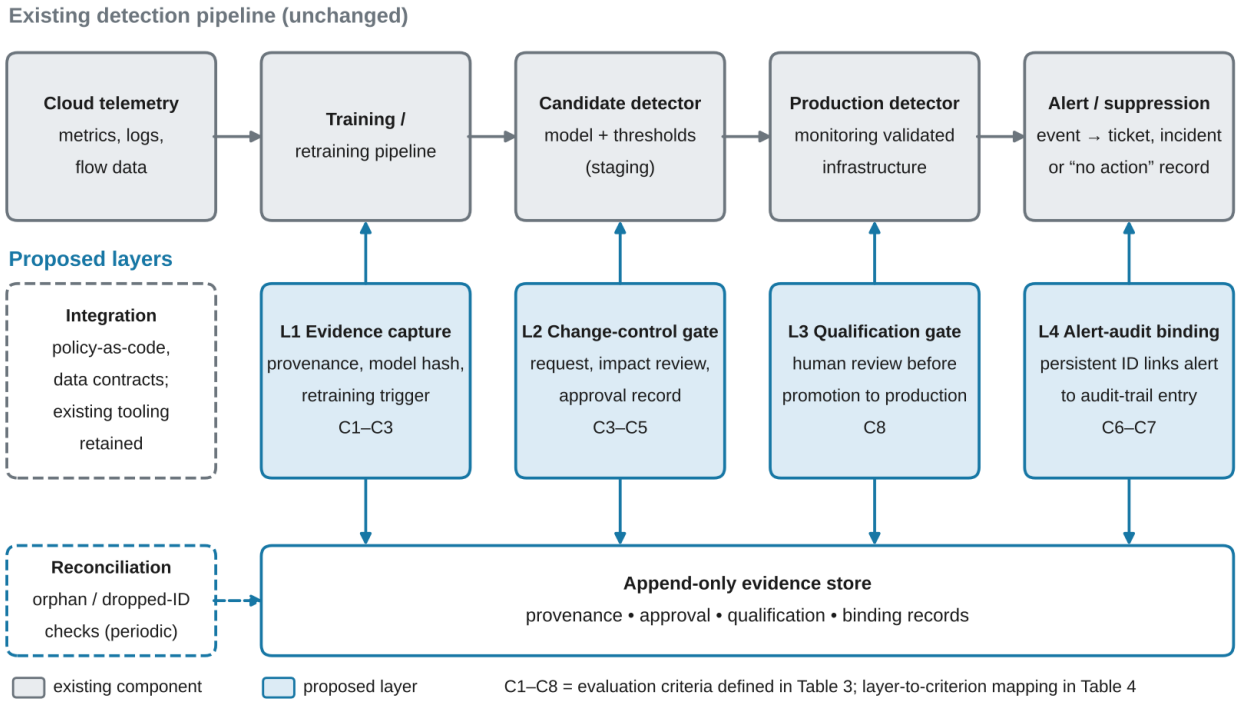

Figure 1. Proposed validation-aware architecture. Grey boxes are the existing detection pipeline; blue boxes are the proposed layers

6. EVALUATION RESULTS

6.1 Criterion-level assessment

Table 5 applies the rubric of Section 4.3 to the two conditions. Baseline levels are assumptions about the reference condition, stated in the table, and are not observations. Illustrative baseline patterns that the author has met in practice, which motivated the reference condition but are not a systematic sample, include training provenance that exists only in scattered notebooks or tickets, threshold changes that travel through software release pipelines built for velocity and not for validated-state approval, and alert records that exist as ephemeral log entries and not as durable links to the audit-trail entries they produced.

Counting criteria at Level 3, the baseline satisfies none of the eight, and its highest level is 2, on C2, where a model registry may exist but deployments are not required to reference it. The proposed architecture reaches Level 3 on five criteria (C1, C2, C3, C6, C7) and Level 2 on three (C4, C5, C8). Three of these scores need explanation. C4 stays at Level 2 because the gate can require an impact assessment and verification evidence, but the quality of the assessment is a human judgement. C5 stays at Level 2 because the emergency path is flagged and its review is tracked, but completing the review is procedural. C8 is scored at the lower of its two parts: the gate enforces that a signed pre-promotion review exists, but periodic evaluation in operation depends on procedures and on drift criteria that remain to be defined. C6 reaches Level 3 only while the reconciliation control is operating; without it, an identifier dropped between systems fails silently, which is arguably worse than an obvious failure because it produces the appearance of compliance without the substance. We treat the architecture's dependence on that control as a finding in its own right (Table 7, F1).

6.2 Traceability scenarios

Table 6 walks the three audit-query scenarios through both conditions. The proposed architecture passes S1 and S3 when every link in the record chain resolves. S2 is only partial in both conditions, for a reason that no record-keeping layer can remove: the architecture can show that an alert was suppressed and under which model and threshold state, but it cannot show that a true anomaly existed that the detector never flagged. This limit follows from the problem and is not a defect of the design, but it bounds what an audit can be told.

6.3 Summary of findings

Three findings follow. First, the gap is chiefly one of linkage and enforcement and not of capability: components such as registries and metadata stores exist, but nothing binds them to change control and to the audit trail. Second, five of the eight criteria can be met by technical enforcement; the other three remain partly procedural and depend on human quality. Third, the dependence of the binding layer on identifier integrity across systems that were not designed to share one is the main implementation risk. We observe, as a hypothesis and not as a finding, that a human qualification gate may also improve the completeness of the records produced

by the other layers, because teams know a reviewer will check them at promotion. This has not been tested here and is included in the validation protocol of Section 7.4.

Table 5. Criterion-level assessment on the four-level rubric (0 absent, 1 ad hoc, 2 defined, 3 enforced). Baseline levels are assumptions about the reference condition; proposed levels are design-level targets assuming correct implementation.

ID	Baseline assumption	Baseline level	Mechanism in the proposed architecture	Proposed level	Residual gap
C1	Metadata store may record dataset references, but fields are not mandatory or tied to promotion	1	Pipeline writes provenance record; promotion blocked on empty fields (L1)	3	Record shows what data were used, not whether they were representative
C2	Registry may version models; deployments need not reference it	2	Deployment admits only registry artefacts identified by hash (L1)	3	Relies on registry integrity and access control
C3	Updates flow through software release pipelines; no validated-state approval record	1	Policy-as-code gate; separation of author and approver (L2)	3	Approver competence; emergency path
C4	CI testing exists; no impact assessment against validated state	1	Gate requires assessment and verification evidence (L2, L3)	2	Quality of the assessment is a human judgement
C5	Hot fixes go through the normal pipeline without flag	0	Flagged expedited path with tracked retrospective review (L2)	2	Completion of the review is procedural
C6	Alerts are ephemeral log entries; manual references in tickets at most	1	Binding record with persistent identifier; reconciliation control (L4)	3 (2 without reconciliation)	Identifier integrity across systems (F1)
C7	Observability platform retains logs; not tamper-evident	1	Append-only, hash-chained store (L4)	3	Retention periods must be aligned with GMP record rules
C8	No detector-specific promotion review or periodic evaluation	0	Enforced pre-promotion review (L3); periodic evaluation procedural	2	Drift criteria and review cadence to be defined

Table 6. Traceability scenarios. Pass: answer reconstructable with every link resolvable. Partial: part of the answer recoverable. Fail: not recoverable from records.

Scenario	Records needed	Baseline	Proposed	Condition or limit
S1: Why did this alert fire?	Binding record; provenance record for the model hash; approved threshold state	Partial: log entry exists; model and threshold state may not be recoverable	Pass	All identifiers in the chain resolve
S2: Why did an expected alert not fire?	Suppression record; model hash; threshold state	Fail: no record of suppression or non-events	Partial	Cannot show that a true anomaly existed that was never flagged
S3: Which threshold was in effect at time T, who approved it and on what evidence?	Threshold-state identifier; approval record; impact assessment	Partial: deployment history shows versions; no approval record for the validated state	Pass	Approval records retained with the evidence store

7. DISCUSSION AND LIMITATIONS

7.1 Fit for purpose

The first tension from Section 3 is that general-purpose cloud anomaly detection was never designed against validated-infrastructure requirements. Our analysis does not dispute that design history. It argues that the history is insufficient once the detector's output drives a control-loop action on a validated system. The fix does not require abandoning the detection techniques that the AIOps literature has developed (Li et al., 2020; Notaro et al., 2021); it requires wrapping them in evidence-generating layers that were never part of their design brief. Teams do not need to discard an LSTM-based or autoencoder-based detector that performs well (Hundman et al., 2018; Xu et al., 2018); they need to stop treating its output as compliance-neutral telemetry once the applicability condition is met.

7.2 Retraining against change control

The second tension is the least comfortable. MLOps practice favours continuous retraining in response to concept drift (Kreuzberger et al., 2023; Lu et al., 2019), whereas the change-control gate requires deliberate, reviewed and sometimes slow approval before any change reaches production. A detector retrained weekly against an approval cycle that takes days will spend a meaningful fraction of its life either running a model version that has not cleared qualification or running a stale model that has cleared qualification but fallen behind the drift the pipeline was built to catch.

Draft Annex 22 sharpens the choice. If a detector were classified as a critical GMP application, the draft's exclusion of dynamic, continuously learning models from such applications would turn this from an engineering trade-off into a constraint (European Commission, & Pharmaceutical Inspection Co-operation Scheme, 2025). The pattern consistent with both the draft and the architecture is therefore a locked production model with retraining performed offline in staging. A candidate model is run in shadow mode alongside the locked model on live telemetry to generate verification evidence, and is promoted only through Layers 2 and 3. Where retraining recurs frequently, a predetermined change protocol, modelled on the predetermined change control plan used for medical devices (U.S. Food and Drug Administration, 2024), can pre-approve a bounded class of changes, for example retraining on a defined data window with an unchanged architecture and hyperparameters inside agreed ranges, provided acceptance criteria are met; changes outside the bound receive full qualification review. Drift monitors can raise a retraining request when input distributions move beyond predefined limits, but the request starts the change process and does not deploy a model.

This reframes the tension without dissolving it. The lag between the onset of drift and the promotion of a qualified replacement is bounded below by the approval cycle time, and an adaptive environment can drift meaningfully between one qualification and the next. The architecture does not eliminate that risk; it makes it visible and attributable, and the validation protocol below proposes to measure it as the time a stale or unqualified model is in production.

7.3 Alert volume against audit trails

The third tension resolves more cleanly but not completely. Alert-reduction (Yu & Zhang, 2023) and log-based detection (Landauer et al., 2023) address a real operational problem, analyst fatigue (Tariq et al., 2025) and alert-storm risk (Poghosyan et al., 2021), but reducing alert volume and producing an auditable record of the alerts that remain are separable problems, and a team focused entirely on the first can leave the second unaddressed indefinitely. The binding layer covers only alerts the system generates. If an alert-reduction layer upstream suppresses an anomaly before it reaches the binding layer, that suppression is itself an event that needs a record; otherwise the reduction layer reintroduces the visibility gap the binding layer was built to close (Table 7, F5).

7.4 Protocol for quantitative validation

The analytical evaluation should be followed by empirical validation. We propose five steps. First, implement the architecture on a representative stack comprising a model registry, a policy engine and an append-only store. Second, run fault-injection tests: drop identifiers in a defined fraction of cross-system calls, attempt gate bypass, and omit provenance fields, then measure how often the reconciliation control, the gate and the schema checks detect each fault. Third, collect metrics: provenance completeness rate, binding integrity rate, orphan rate, time to reconstruct the answers to S1–S3, approval cycle time, and the fraction of time a stale or unqualified model is in production. Fourth, have at least two validation or quality-assurance professionals score Table 5 independently and report inter-rater agreement, for example with Cohen's kappa, which addresses the single-analyst limitation. Fifth, test the hypothesis of Section 6.3 by comparing record completeness with and without the qualification gate, and confirm that the layers leave detection outputs for an unchanged model identical, reporting any added latency.

7.5 Limitations

The evaluation is analytical and practitioner-derived. Levels were assigned by a single analyst; baseline levels are assumptions; and the sources are US and EU texts, so the analysis does not cover other jurisdictions.

Table 7. Failure modes of the proposed controls and qualitative residual risk after mitigation.

ID	Failure mode	Effect on audit readiness	Detection or mitigation	Residual risk
F1	Persistent identifier dropped or mismatched between detector and system of record	Binding fails silently; appearance of compliance without substance	Reconciliation control; referential-integrity checks; canary alerts	Medium
F2	Provenance fields populated incorrectly or pipeline not writing	Records exist but are unreliable	Schema validation at the gate; sampled verification at qualification	Medium
F3	Gate bypass through direct deployment access or misuse of the emergency path	Unapproved change to the validated state	Access control and separation of duties; admission policy; overdue-review alerts	Low to medium

ID	Failure mode	Effect on audit readiness	Detection or mitigation	Residual risk
F4	Qualification review reduced to a formality	Gate satisfied nominally	Reviewer independence; pre-specified acceptance criteria; periodic review of review outcomes	Medium
F5	Upstream alert reduction suppresses events without a record	Visibility gap reintroduced	Require suppression logging to the same store	Medium
F6	Clock skew between systems of record	Time sequence of records unreliable	Synchronised UTC source; record detection and ingestion times	Low
F7	Retraining cadence outpaces the approval cycle	Unqualified or stale model in production	Locked model; predetermined change protocol; drift monitor raises a request	High (structural)
F8	Immature policy-as-code and RBAC foundation	Gate has nothing to attach to	Precondition; out of scope	High where the precondition is absent

The architecture has not been through a regulatory audit, and we make no claim that it would satisfy any authority's inspection criteria; that question is for quality and regulatory professionals. The classification of a detector as inside or outside the validated boundary depends on a risk assessment that this article cannot perform for an organisation. The regulatory landscape is moving, and draft Annex 22 and the revised Annex 11 should be checked for their current status before any application of this work. The architecture also assumes an organisation with mature policy-as-code and RBAC practices; without them the change-control gate has nothing solid to attach to. No public dataset combines cloud telemetry with GxP audit-trail ground truth, so no quantitative claim about detection performance can be supported here. Table 7 summarises failure modes of the proposed controls and the residual risk that remains after mitigation.

8. CONCLUSION

This article started from a gap that, in our experience, is easy to overlook: the anomaly detector watching a validated system may itself sit inside that system's control loop, and treating it as outside is a compliance decision whether or not anyone made it consciously. The contribution is not a new detection algorithm. It is, first, a way of separating what the regulatory sources require from what is guidance and what is proposed; second, an applicability condition and an architecture that move the validated boundary to include the detector where the condition is met; and third, an evaluation framework that states how far the architecture can be trusted to do so.

The analysis indicates that five of eight criteria can be met by technical enforcement, and three remain partly procedural. The four layers are not exotic: they borrow policy-as-code enforcement, RBAC-gated approval, immutable audit logging and staged promotion from infrastructure practice. What changes is the target, a class of software that has escaped that discipline because it looks like monitoring and not like a control-loop component. The tension between retraining and change control remains unresolved, and we think that is the honest note to end on. Locked models, shadow evaluation and predetermined change protocols bound the problem, but adaptive models and change-controlled validated states run on different clocks.

The study is analytical, and its levels are design-level targets. Empirical validation under the protocol of Section 7.4 is the next step. Whether a particular binding record satisfies a particular regulator's expectations is a matter for quality and regulatory professionals. The architecture is offered as an engineering proposal whose regulatory anchors are identified so that such review can take place before a detector goes into production, not after an inspector asks.

References:

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys, 41*(3), Article 15. DOI: 10.1145/1541880.1541882

European Commission. (2011). *EudraLex Volume 4: Good manufacturing practice guidelines, Annex 11: Computerised systems*. https://health.ec.europa.eu/medicinal-products/eudralex/eudralex-volume-4_en

European Commission. (2015). *EudraLex Volume 4: Good manufacturing practice guidelines, Annex 15: Qualification and validation*. https://health.ec.europa.eu/medicinal-products/eudralex/eudralex-volume-4_en

European Commission, & Pharmaceutical Inspection Co-operation Scheme. (2025). *Draft EudraLex Volume 4, Annex 22: Artificial intelligence (with draft revised Annex 11 and Chapter 4)*. Stakeholder consultation, July 7–October 7, 2025. https://health.ec.europa.eu/consultations/stakeholders-consultation-eudralex-volume-4-good-manufacturing-practice-guidelines-chapter-4-annex_en

Hundman, K., Constantinou, V., Laporte, C., Colwell, I., & Soderstrom, T. (2018). Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding. In *Proceedings of the 24th ACM SIGKDD International*

Conference on Knowledge Discovery & Data Mining (KDD '18) (pp. 387–395). ACM. https://arxiv.org/abs/1802.04431

International Council for Harmonisation. (2008). *ICH Q10: Pharmaceutical quality system*. https://www.ich.org/page/quality-guidelines

International Council for Harmonisation. (2023). *ICH Q9(R1): Quality risk management*. https://www.ich.org/page/quality-guidelines

International Society for Pharmaceutical Engineering. (2022). *GAMP 5: A risk-based approach to compliant GxP computerized systems* (2nd ed.). https://ispe.org/publications/guidance-documents/gamp-5-guide-2nd-edition

International Society for Pharmaceutical Engineering. (2025). *ISPE GAMP guide: Artificial intelligence*. https://ispe.org/index.php/publications/guidance-documents/gamp-guide-artificial-intelligence

Islam, M. S., Pourmajidi, W., Zhang, L., Steinbacher, J., Erwin, T., & Miranskyy, A. (2021). Anomaly detection in a large-scale cloud platform. In *Proceedings of the 2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)* (pp. 150–159). IEEE. DOI: 10.1109/ICSE-SEIP52600.2021.00024

Kaur, D., Uslu, S., Rittichier, K. J., & Durresi, A. (2023). Trustworthy artificial intelligence: A review. *ACM Computing Surveys, 55*(2), Article 39. DOI: 10.1145/3491209

Kreuzberger, D., Kühl, N., & Hirschl, S. (2023). Machine learning operations (MLOps): Overview, definition, and architecture. *IEEE Access, 11*, 31866–31879. https://ieeexplore.ieee.org/document/10081336

Landauer, M., Onder, S., Skopik, F., & Wurzenberger, M. (2023). Deep learning for anomaly detection in log data: A survey. *Machine Learning with Applications, 12*, Article 100470. https://www.sciencedirect.com/science/article/pii/S2666827023000233

Li, Y., Jiang, Z. M., Li, H., Hassan, A. E., He, C., Huang, R., Zeng, Z., Wang, M., & Chen, P. (2020). Predicting node failures in an ultra-large-scale cloud computing platform: An AIOps solution. *ACM Transactions on Software Engineering and Methodology, 29*(2), Article 13. DOI: 10.1145/3385187

Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., & Zhang, G. (2019). Learning under concept drift: A review. *IEEE Transactions on Knowledge and Data Engineering, 31*(12), 2346–2363. https://arxiv.org/abs/2004.05785

Muralidharan, V., Adewale, B. A., Huang, C. J., Nta, M. T., Ademiju, P. O., Pathmarajah, P., Hang, M. K., Adesanya, O., Abdullateef, R. O., Babatunde, A. O., Ajibade, A., Onyeka, S., Cai, Z. R., Daneshjou, R., & Olatunji, T. (2024). A scoping review of reporting gaps in FDA-approved AI medical devices. *npj Digital Medicine, 7*, Article 273. DOI: 10.1038/s41746-024-01270-x

Myllyaho, L., Raatikainen, M., Männistö, T., Mikkonen, T., & Nurminen, J. K. (2021). Systematic literature review of validation methods for AI systems. *Journal of Systems and Software, 181*, Article 111050. https://www.sciencedirect.com/science/article/pii/S0164121221001473

Nassif, A. B., Talib, M. A., Nasir, Q., & Dakalbab, F. M. (2021). Machine learning for anomaly detection: A systematic review. *IEEE Access, 9*, 78658–78700. https://ieeexplore.ieee.org/document/9439459

Notaro, P., Cardoso, J., & Gerndt, M. (2021). A survey of AIOps methods for failure management. *ACM Transactions on Intelligent Systems and Technology, 12*(6), Article 90. DOI: 10.1145/3483424

Pharmaceutical Inspection Co-operation Scheme. (2007). *PI 011-3: Good practices for computerised systems in regulated "GxP" environments*. https://picscheme.org/en/publications

Poghosyan, A., Harutyunyan, A., Grigoryan, N., & Kushmerick, N. (2021). Incident management for explainable and automated root cause analysis in cloud data centers. *Journal of Universal Computer Science, 27*(11), 1152–1173. DOI: 10.3897/jucs.76608

Sahiner, B., Chen, W., Samala, R. K., & Petrick, N. (2023). Data drift in medical machine learning: Implications and potential remedies. *British Journal of Radiology, 96*(1150), Article 20220878. DOI: 10.1259/bjr.20220878

Schmidl, S., Wenig, P., & Papenbrock, T. (2022). Anomaly detection in time series: A comprehensive evaluation. *Proceedings of the VLDB Endowment, 15*(9), 1779–1797. DOI: 10.14778/3538598.3538602

Tariq, S., Baruwal Chhetri, M., Nepal, S., & Paris, C. (2025). Alert fatigue in security operations centres: Research challenges and opportunities. *ACM Computing Surveys, 57*(9). DOI: 10.1145/3723158

Testi, M., Ballabio, M., Frontoni, E., Iannello, G., Moccia, S., Soda, P., & Vessio, G. (2022). MLOps: A taxonomy and a methodology. *IEEE Access, 10*, 63606–63618. DOI: 10.1109/ACCESS.2022.3181730

U.S. Food and Drug Administration. (2018, December). *Data integrity and compliance with drug CGMP: Questions and answers*. Guidance for Industry. https://www.fda.gov/media/119267/download

U.S. Food and Drug Administration. (2024, December). *Marketing submission recommendations for a predetermined change control plan for artificial intelligence-enabled device software functions: Final guidance*. https://www.fda.gov/regulatory-information/search-fda-guidance-documents

U.S. Food and Drug Administration. (n.d.a). *21 CFR Part 11—Electronic records; electronic signatures*. Electronic Code of Federal Regulations. https://www.ecfr.gov/current/title-21/chapter-I/subchapter-A/part-11

U.S. Food and Drug Administration. (n.d.b). *21 CFR Part 211—Current good manufacturing practice for finished pharmaceuticals, § 211.68*. Electronic Code of Federal Regulations. https://www.ecfr.gov/current/title-21/chapter-I/subchapter-C/part-211

Wall, D. P., Liu-Mayo, S., Salomon, C., Shannon, J., & Taraman, S. (2023). Optimizing a de novo artificial intelligence-based medical device under a predetermined change control plan: Improved ability to detect or rule out pediatric autism. *Intelligence-Based Medicine, 8*, Article 100102. https://www.sciencedirect.com/science/article/pii/S2666521223000169

Xu, H., Chen, W., Zhao, N., Li, Z., Bu, J., Li, Z., Liu, Y., Zhao, Y., Pei, D., Feng, Y., Chen, J., Wang, Z., & Qiao, H. (2018). Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications. In *Proceedings of the 2018 World Wide Web Conference (WWW '18)* (pp. 187–196). https://arxiv.org/abs/1802.03903

Yu, M., & Zhang, X. (2023). AlertInsight: Mining multiple correlation for alert reduction. *Computer Systems Science and Engineering, 46*(2), 2447–2469. DOI: 10.32604/csse.2023.037506

Rajashekar Reddy Gujjula
Independent Researcher, India
rajashekargujjular@gmail.com
ORCID: 0009-0000-8482-9219