International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

# From Digitalization to Assurance: Practical Use of Artificial Intelligence in GxP Validation

**Harshitkumar Prajapati**

**Abstract:** The life sciences industry has spent the better part of the last decade digitizing its operations — replacing paper-based systems, automating workflows, and centralizing data across increasingly interconnected platforms. That transition, for the most part, followed a familiar playbook: validate the system, document the controls, and demonstrate to regulators that the technology does what it claims to do. Artificial intelligence does not follow that playbook, and the industry is beginning to feel the friction. AI-enabled systems are now appearing across pharmaceutical and biotechnology operations — from manufacturing analytics and deviation triage to clinical signal detection and quality monitoring. These are not experimental deployments. Organizations are making consequential GxP decisions based on AI outputs, and in many cases, the validation strategies supporting those decisions were built for a different kind of technology entirely. Traditional computer system validation assumed that a system's behavior was fixed at the point of deployment. AI does not make that assumption, and neither does the data it learns from. This paper examines what actually needs to change — and what does not. Regulatory agencies have been consistent: existing quality and risk management principles still apply to AI. What they have been less prescriptive about is how. Drawing on the Computer Software Assurance framework and direct experience implementing AI governance within a GxP-regulated advanced therapy organization, this paper proposes a practical, risk-based approach to AI assurance that focuses on lifecycle control, performance monitoring, and inspection readiness rather than exhaustive algorithmic testing. The intent is not to rewrite validation science, but to extend it honestly into territory it was not originally designed to reach.

***Keywords:*** *artificial intelligence, GxP compliance, computer software assurance, validation, digitalization, quality risk management, data integrity, machine learning*

## 1. Introduction

Something shifted in how life sciences organizations think about artificial intelligence roughly around 2021. Before that point, AI in regulated environments was largely a subject of conference presentations and vendor roadmaps. The technology existed, and interest was genuine, but deployment in GxP-critical processes remained cautious and limited. What changed was a convergence of forces arriving at roughly the same time: large language models became dramatically more capable; cloud-delivered AI platforms became operationally accessible without deep internal infrastructure investment; and competitive pressure to accelerate development timelines intensified—particularly in advanced therapy spaces where organizations are simultaneously scaling manufacturing, clinical operations, and regulatory strategy.

Independent Researcher, USA

The result is that AI adoption in regulated environments is no longer a future consideration. It is a present operational reality, and the compliance functions responsible for governing these environments are, in many organizations, still catching up.

The core challenge is practical rather than philosophical. Computer system validation, as it has been practiced for decades, was designed around deterministic software: systems that, given the same input, produce the same output, behave consistently after deployment, and do not evolve unless a controlled change is introduced. That model of technology behavior is foundational to how validation protocols are written, acceptance criteria are defined, and ongoing compliance is maintained. Artificial intelligence challenges every one of those assumptions. A machine learning model trained on one dataset may produce different outputs when the underlying data distribution shifts. A system designed to learn continuously may improve in some

respects and degrade in others, independent of any formal change event. A model updated through retraining—which is not a code release, does not trigger a traditional change control, and may leave no obvious audit trail—can nonetheless affect every GxP decision it informs.

These are not hypothetical risks. They are operational realities being encountered by validation professionals working with AI-enabled platforms today, and the existing CSV toolkit does not give them adequate tools to address them.

This paper proposes a practical path forward grounded in Computer Software Assurance principles. [4] CSA, as defined by the FDA's 2022 guidance, shifts the focus of assurance from exhaustive documentation and scripted testing toward critical thinking, risk-based prioritization, and evidence of meaningful control over system outcomes. Applied to AI, this means moving focus away from validating what happens inside a model and toward validating the governance, performance monitoring, and lifecycle controls that determine whether model outputs can be trusted in GxP decisions.

The approach presented here reflects not only regulatory guidance but also direct implementation experience across GxP-regulated environments, including deployment of AI-assisted quality monitoring in an advanced therapy manufacturing and clinical operations context. The goal is to give compliance and quality professionals a framework they can actually use — one that is credible to regulators, proportional to risk, and sustainable as AI capabilities continue to evolve.

## 2. The Pace of AI Adoption in Regulated Environments

Understanding why the industry faces a validation gap with AI requires understanding how adoption actually happened — not how it was planned, but how it unfolded in practice.

In most regulated organizations, AI entered through one of several paths. The most common was vendor-embedded AI: a validated SaaS platform that an organization had previously qualified had added AI-enhanced features—intelligent search, anomaly flagging, and predictive recommendations—as part of a general release update. The organization's existing qualification covered the platform but not the new AI capability. A second path was departmental adoption: individual functions introduced cloud-delivered AI tools to improve operational efficiency, initially using them for tasks adjacent to GxP processes and then gradually moving them closer to regulated workflows as confidence grew. A third path was formal strategic initiative: organizations evaluated AI platforms for specific GxP use cases, undertook structured validation projects, and deployed with defined governance from the outset. This was the least common path but consistently produced the clearest compliance outcomes.

The research literature reflects a parallel trajectory. Topol's foundational analysis of AI in high-performance medicine documented the convergence of machine learning capability with clinical and operational decision-making, observing that AI's most transformative impact in life sciences would come not from replacing human expertise but from augmenting it at scale—a pattern that has since proven accurate across pharmaceutical and biotechnology operations. [17] In the manufacturing context specifically, PDA Technical Report No. 84 documented AI and advanced analytics applications already transitioning from pilot to operational deployment, including predictive maintenance of manufacturing equipment, real-time release testing support, anomaly detection in batch records, and intelligent monitoring of critical quality attributes. [12]

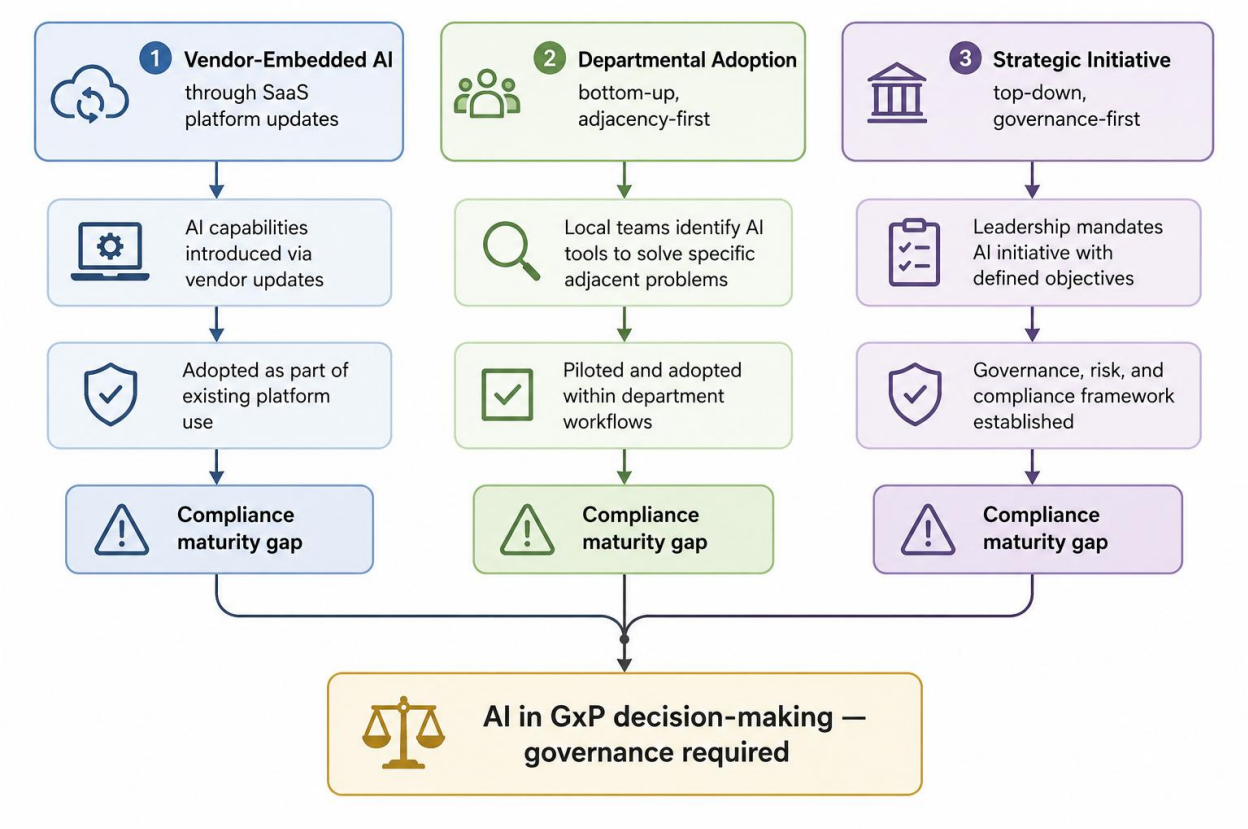


Figure 1: AI Adoption Pathways into Regulated Environments

What these use cases share is that they arrived ahead of a consolidated regulatory framework for governing AI in GxP environments. Organizations were deploying AI while simultaneously determining how it should be governed, what records it needed to generate, and who was accountable for its ongoing performance. The resulting governance gap is not a failure of intent — it is a structural consequence of technology capability outpacing compliance infrastructure.

The use cases now present in regulated environments span a wide range of risk profiles. Some AI applications operate in clearly advisory roles, where system output informs a decision but does not make it. Others are more deeply integrated, with AI outputs directly influencing release determinations, deviation classifications, or supply chain actions. The distinction between these categories is fundamental to any rational assurance strategy and is the basis for the risk framework developed in Section 5.

**Table 1** summarizes representative AI use cases across the GxP risk spectrum, drawing on deployment patterns documented in PDA TR 84[12] and the risk categorization principles established in FDA CSA guidance[4] and ISPE GAMP AI/ML. [6]

**Table 1. AI Use Cases in GxP Environments by Risk Tier**

| Risk Tier | Example Use Cases | Decision Type | Primary Regulatory Concern | Indicative Validation Approach |
|---|---|---|---|---|
| **Low Impact** | Document summarization; SOP drafting support; internal literature search; meeting notes synthesis | Advisory—output informs user but does not enter GxP record or influence GxP decision | Information accuracy; appropriate user awareness of AI limitations | Intended use documentation; user training; basic oversight; minimal documented assurance required |
| **Moderate Impact** | Deviation triage and prioritization; quality data anomaly flagging; audit trail review support; validation document drafting assistance | Determinative-advisory—output directly supports GxP decisions, but mandatory human review is required before any regulated action | Reliability and consistency of output; documented human review workflow; change control for model updates | CSA-based risk-proportional validation; testing of critical functions; defined review procedures; periodic performance monitoring |
| **High Impact** | Batch release decision support; critical quality attribute prediction in manufacturing; clinical safety signal detection; supply chain disposition support | Determinative — output materially drives or constrains a GxP outcome with limited human override | Patient safety; product quality; data integrity; audit trail completeness; human accountability | Risk-based validation; scripted or unscripted testing as appropriate of all GxP-critical functions; continuous performance monitoring; rigorous change control; mandatory qualified personnel sign-off |

*Sources: PDA TR 84,[12] FDA CSA Guidance,[4] ISPE GAMP® AI/ML.[6]*

**3. What Regulators Actually Expect—and What They Do Not**

One of the most persistent misconceptions in discussions of AI validation is that regulators expect something entirely new. They do not. What the FDA, EMA, and other agencies have consistently communicated is that existing regulatory principles, data integrity, system reliability, traceability, and risk-based decision-making apply to AI-enabled systems in precisely the same way they apply to any other technology that influences GxP outcomes. ICH Q9(R1), which governs quality risk management across the pharmaceutical lifecycle, makes no exception for AI. [3] Neither does the FDA's data integrity guidance, nor the electronic records requirements of 21 CFR Part 11, which applies to any computer-generated records used to satisfy regulatory requirements regardless of the underlying technology generating them. [9]

The evolution of the FDA's regulatory thinking on AI is instructive. As early as 2019, the agency published a discussion paper proposing a regulatory framework for AI/ML-based software as a medical device, recognizing that adaptive systems—those designed to improve their own performance over time—posed challenges that existing review and monitoring models were not designed to address. [20] This document acknowledged that the traditional framework's reliance on a "locked" algorithm at the point of market entry was fundamentally incompatible with the value proposition of adaptive AI. The 2021 Action Plan that followed built on those observations, articulating a lifecycle-based approach to AI management that placed ongoing performance monitoring and change governance at the center of regulatory expectations. [1] The continuity between these two documents matters: it shows that FDA's current position on AI governance is not reactive policy but the product of several years of deliberate regulatory development.

EMA has followed a parallel trajectory. The 2023 Reflection Paper on AI in the Medicinal Product Lifecycle addressed the dynamic nature of AI systems directly, emphasizing that organizations must develop strategies to ensure continued reliability, traceability, and transparency as models evolve over time. [2] EMA's 2023 Guideline on Computerized Systems and Electronic Data in

Clinical Trials reinforced the consistent expectation that electronic system governance—including systems with AI capabilities—must meet established standards for data integrity, audit trail completeness, and access control, regardless of the sophistication of the underlying technology. [19] Taken together, these documents establish a regulatory posture that is firm on principles and deliberately flexible on implementation — placing the burden of reasoned, risk-based compliance design squarely on the organization.

The FDA's CSA guidance is valuable here not because it was written specifically for AI, but because its underlying logic maps cleanly onto AI governance challenges. [4] CSA asks organizations to think critically about what actually needs to be controlled, to focus testing and documentation effort on genuinely high-risk functions, and to leverage supplier documentation and operational evidence rather than generating documentation for its own sake. For AI systems, this means accepting that scripted testing alone cannot establish assurance for probabilistic systems, and shifting focus to the quality of governance and monitoring controls instead.

**Table 2** maps the applicable regulatory guidance landscape to AI-specific implications and the inspection focus areas they generate. This mapping draws directly on the source documents cited and is intended to support organizations in aligning their AI governance programs with current regulatory expectations.

**Table 2. Regulatory Guidance Landscape for AI in GxP Environments**

| Key Regulatory Principle | AI-Specific Implication | Primary Inspection Focus |
|---|---|---|
| Lifecycle management of adaptive AI/ML systems | Adaptive AI cannot be governed through a locked-algorithm model; requires active, ongoing oversight framework | Evidence of postmarket monitoring strategy and performance governance |
| Predetermined change control plan (PCCP) concept for model modifications | Changes through retraining should be anticipated, categorized, and pre-approved where feasible | Documentation of expected model changes and their evaluation criteria |
| Critical thinking; risk-based testing; leveraging supplier evidence | Validation effort should focus on GxP-critical functions; exhaustive scripted testing is not required nor expected | Rationale for validation scope; evidence that critical risks were identified, assessed, and addressed |
| Electronic records integrity; audit trails; access controls | AI-generated records used to satisfy regulatory requirements must comply with Part 11 controls | Audit trail completeness; access management; record retention for AI outputs |
| AI lifecycle management; model transparency and traceability | Model history, training data provenance, and performance trends must be traceable and documentable | Traceability of model versions; documented lifecycle governance |
| Computerised system governance in clinical operations | AI-enabled systems in clinical settings must meet consistent data integrity and access control standards | Data integrity controls; system validation status; electronic record governance |
| Quality risk management principles | Risk management principles apply to AI without modification; controls must be proportional to GxP impact | Evidence of formal risk assessment informing the validation strategy |
| Computerised system validation, audit trail, access control | AI systems used in GMP operations are computerised systems fully subject to Annex 11 requirements | Validation documentation; audit trail functionality; role-based access management |

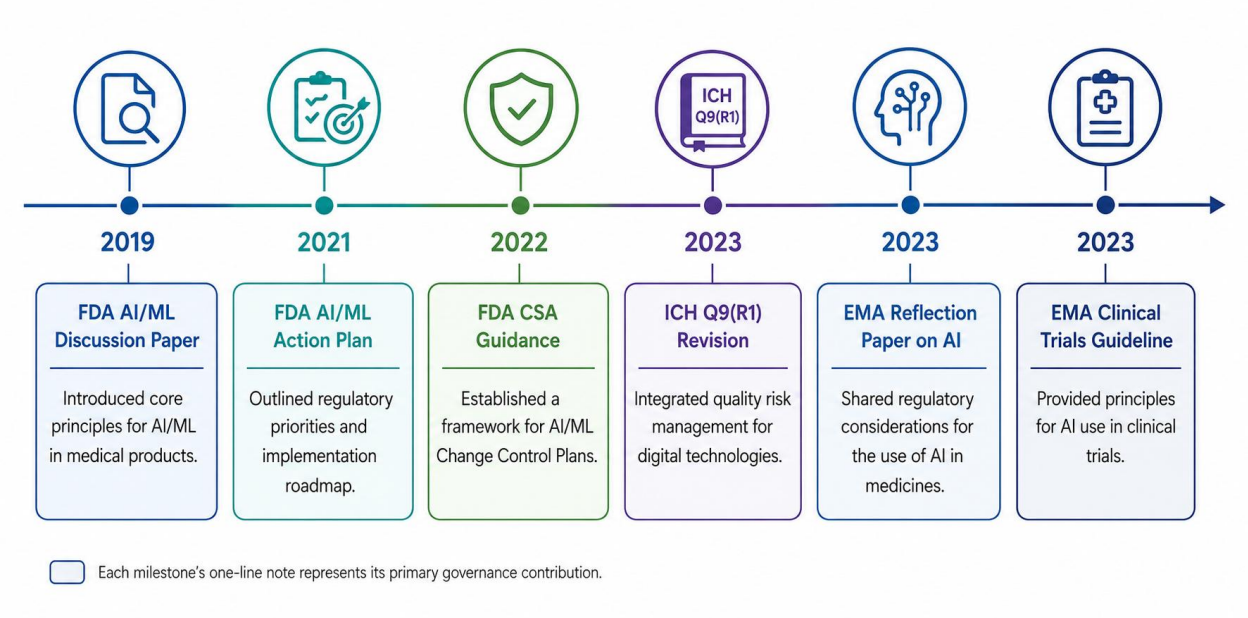


Figure 2: Regulatory Timeline: Evolution of AI Governance Guidance (2019–2023)

One area where regulatory expectations remain genuinely unsettled is continuous learning. Systems designed to update their own models through ongoing data ingestion sit in an unresolved space between software change and operational drift. No major regulatory guidance document has, as of this writing, provided definitive direction on how organizations should manage automated retraining within a validated state. The practical approach most organizations have adopted — establishing threshold-based triggers that convert retraining events into formal change control actions when performance metrics shift beyond predefined bounds — is reasonable and defensible, but it represents emerging industry consensus rather than regulatory mandate.

**4. Why Traditional CSV Falls Short**

The limitations of traditional CSV when applied to AI are worth stating precisely, because the instinct in many compliance organizations is either to apply legacy approaches wholesale or to declare that AI requires something entirely different. Neither response is correct.

Traditional CSV works on a straightforward premise: a system has defined functionality, that functionality is tested against acceptance criteria, and if the system passes, it is considered validated. The validated state is maintained by controlling changes—any modification that could affect system behavior triggers a new round of evaluation. This model is rational, well-understood by regulators, and highly effective for the class of systems it was designed for.

AI disrupts this model at three specific points. The first is acceptance criteria. Scripted test cases for deterministic systems produce pass/fail results against fixed expected outputs. Probabilistic systems do not have fixed outputs—the same input may produce different results depending on the state of the model, the training data at the time of execution, or the operational context. Writing acceptance criteria for this kind of variability requires either accepting a range of outputs (unfamiliar to most validation teams) or limiting tested scope to a narrow set of known-correct scenarios (which provides limited assurance over the full operational range of the system).

The second disruption is the change model. Traditional change control is triggered by modifications to software code, configuration, or infrastructure. AI models can change their effective behavior without any of these events occurring—through retraining on new data, through distributional shifts in incoming data, or through gradual degradation of model performance over time. None of these events would appear in a traditional change management log, and none would automatically trigger a validation impact assessment. Organizations that rely exclusively on existing change management processes to maintain AI systems in a validated state have a structural governance gap.

Third, active compliance. Most validation activities are front-loaded: the bulk of the effort is in

qualifying the system before it is put into use, and continuing compliance involves periodic review and change control. By contrast, AI systems lend themselves poorly to this model because their risk profile is not consistent. A model may perform well in deployment but perform poorly in an alternative operational setting. Long-term compliance for AI, then, requires active oversight rather than a more passive, backward-looking approach. This is not an argument against validation. They are arguments for validating them differently and more cleverly.

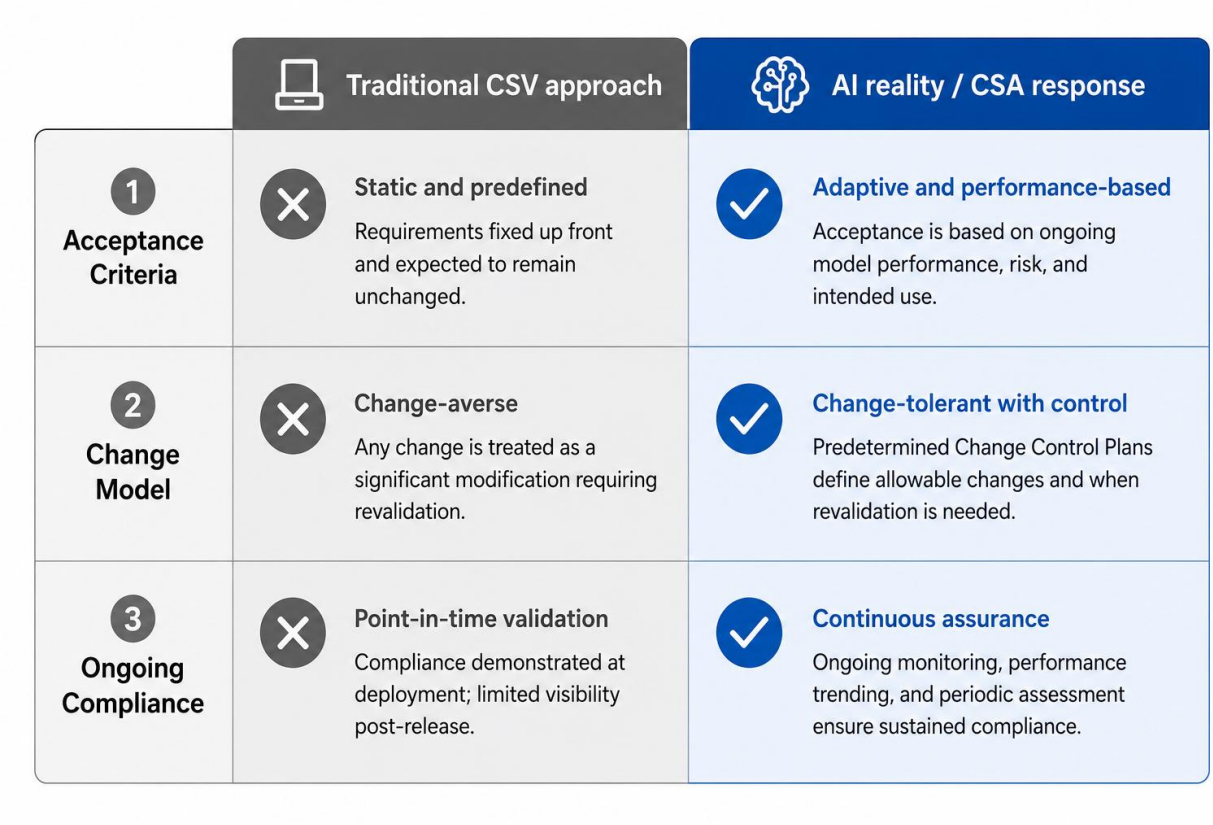


Figure 3: Traditional CSV vs. CSA-Based AI Assurance: Key Differences

**5. A CSA-Based Assurance Framework for AI-Enabled GxP Systems**

The framework proposed here is built on four principles that adapt CSA logic to the specific characteristics of AI-enabled systems. It does not replace CSA — it extends it into territory where traditional validation methods begin to lose their grip.

**Principle 1: Assurance begins with intended use, not system architecture.** The first and most consequential decision in any AI validation project is a precise definition of intended use: what is this system designed to do, in what operational context, and what GxP decisions does its output influence? This is frequently underspecified in practice. An AI system described broadly as "supporting deviation management" might be used simply to search historical records, or it might be providing risk classifications that determine escalation pathways. These represent fundamentally different risk profiles requiring different assurance strategies. Clarity about intended use is the foundation upon which everything else is built.

**Principle 2: Risk classification drives the depth and nature of controls.** Not all AI applications in GxP environments require the same level of assurance. The framework applies a three-tier risk classification: Low Impact, Moderate Impact, and High Impact, calibrated against the degree to which AI outputs influence GxP decisions and the consequences of output error. The pharmaceutical quality system lifecycle governance framework established in ICH Q10 provides the underlying structural logic: quality risk management, document and change management, and performance monitoring are not invented for AI — they are extended from an existing, regulatory-accepted quality system architecture. [11]

**Principle 3: Controls are placed around the model, not inside it.** The objective of AI assurance

is not to validate every possible model output. It is to demonstrate that the model operates within a governed environment—that its inputs are reliable, its outputs are used appropriately, its changes are recognized and evaluated, and its performance is actively monitored. This shifts the validation effort from internal algorithmic testing to governance infrastructure. ISPE GAMP 5 supports this principle directly: risk-based validation focuses on what is critical to product quality and patient safety, not on technical thoroughness for its own sake. [5]

**Principle 4: Assurance is continuous, not event-driven.** The validated state of an AI system is not established once at deployment and maintained passively thereafter. It must be actively sustained through ongoing performance monitoring, periodic review, and timely response to performance signals—an approach that mirrors the continued process verification logic established for manufacturing operations under FDA's process validation framework. [18]

**Table 3** operationalizes these four principles into a control requirements matrix across the three risk tiers, drawing on FDA CSA guidance,[4] ICH Q9(R1),[3] ICH Q10,[11] and ISPE GAMP 5. [5]

**Table 3. CSA-Based AI Validation Framework — Risk Tier × Required Controls Matrix**

| Control Domain | Low Impact | Moderate Impact | High Impact |
|---|---|---|---|
| **Intended Use Documentation** | Brief description of system purpose and user population; AI limitations noted | Formal intended use statement; documented GxP impact assessment; defined decision workflow | Comprehensive intended use with decision authority matrix; explicit human oversight protocol; patient safety impact statement |
| **Testing Requirements** | Unscripted or exploratory testing; user acceptance confirmation sufficient | Risk-based scripted testing of critical functions; supplier documentation leveraged for lower-risk capabilities | Formal test protocols for all GxP-critical functions; independent review of testing results required |
| **Change Control** | Informal review; assess whether revalidation is warranted | Formal change request; documented impact assessment for model or data changes; validation update where required | Full change control with impact assessment required; revalidation for significant model changes; regulatory authority approval pathway where applicable |
| **Performance Monitoring** | Periodic user feedback; no formal performance metrics required | Defined performance metrics; periodic review at scheduled intervals; investigation triggers for threshold breach | Continuous monitoring; real-time or near-real-time performance tracking; automated alert thresholds with defined response procedures |
| **Human Oversight** | User discretion; no formal review requirement | Mandatory human review of AI outputs before any GxP action is taken; documented review trail | Qualified personnel sign-off required for all GxP decisions; defined escalation path; AI recommendation and human decision explicitly separated in records |
| **Supplier Reliance** | Vendor documentation may serve as primary assurance evidence | Vendor documentation supplemented by internal testing of critical functions | Internal validation primary; vendor documentation supporting evidence only |

*Sources: FDA CSA Guidance,[4] ICH Q9(R1),[3] ICH Q10,[11] ISPE GAMP® 5,[5] ISPE GAMP® AI/ML.[6]*

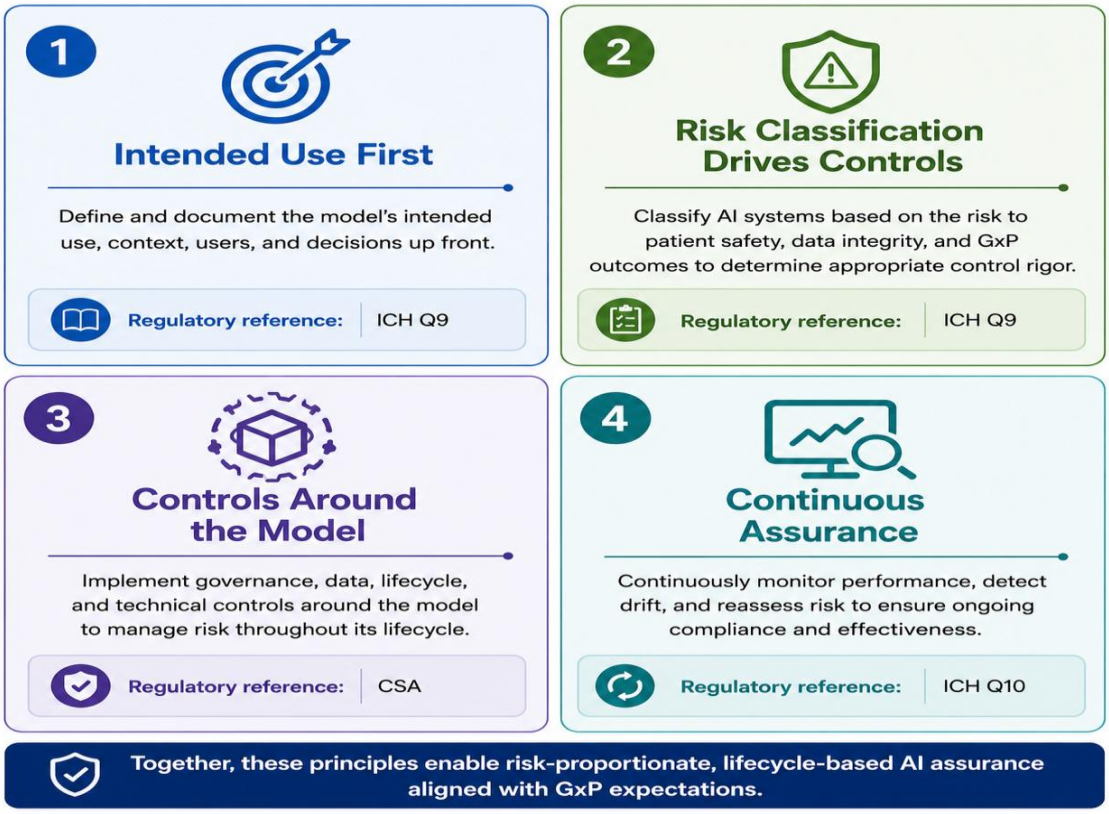


Figure 4: CSA-Based AI Assurance Framework — Four Governing Principles

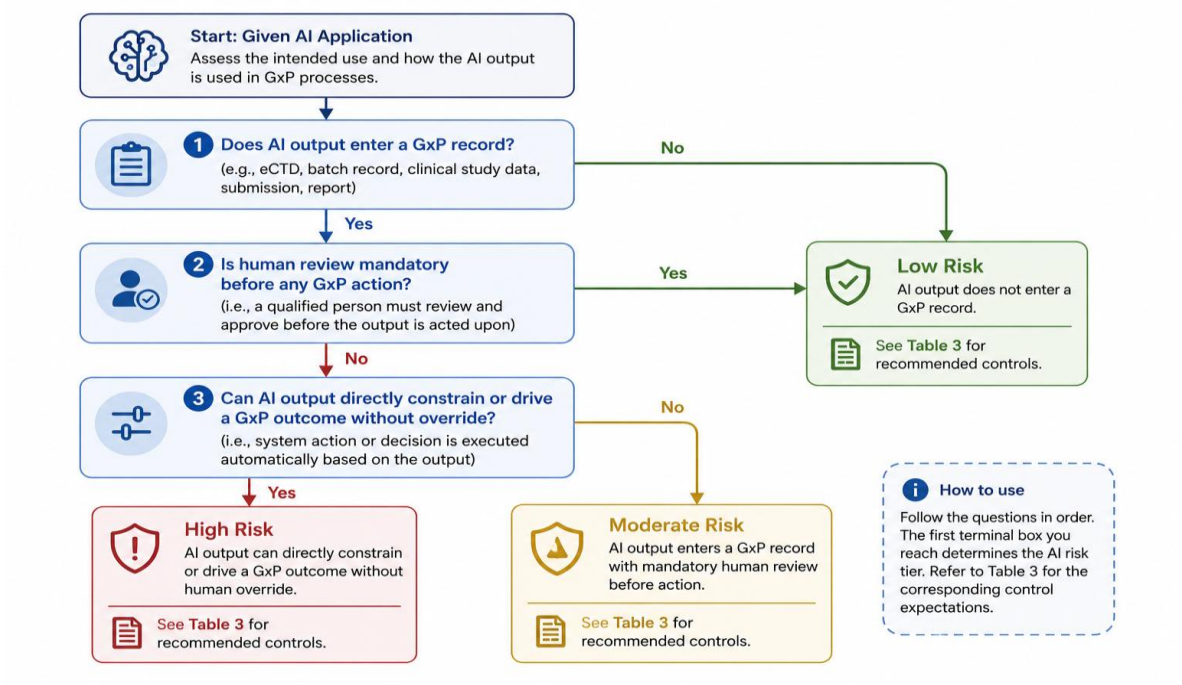


Figure 5: AI Risk Tier Classification Decision Tree

For example, an AI-assisted anomaly detection system for GxP manufacturing might monitor and flag live process data for relevant deviations or anomalies for quality reviewers to take action. In this scenario, the application would be Moderate to High Impact: outputs have an impact on deviation initiation, which is GxP, but final disposition is performed by qualified personnel. The validation measures include procedures to ensure that the system raises known high-risk situations (e.g.,

targeted testing of critical functions), that inputs to the application are sourced and monitored for quality, and that alerts are escalated to quality events. Ongoing assurance may involve periodic checks against plausibility flags, tracking false negative rates against limits and thresholds, and change control processes that allow for retraining in response to shifts in the distributors' data regime. This focused, proportional approach grounded in the operational context may be more promising for regulatory approval than exhaustive testing of edge cases.

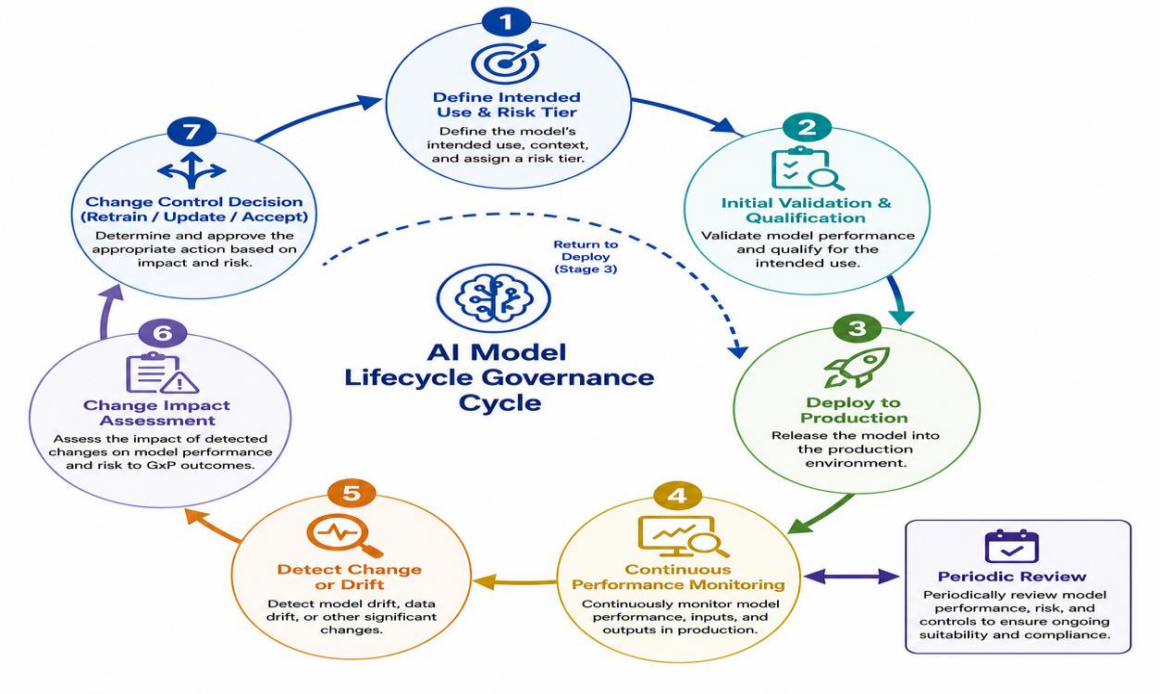


Figure 6: AI Model Lifecycle Governance Cycle

## 6. Practical Controls: Model Change, Data Drift, and Explainability

### 6.1 Model Change and Lifecycle Governance

The most significant governance gap most organizations face with deployed AI is the absence of a model change management framework. Traditional change control processes were not designed to recognize model retraining as a change event, and in many organizations model updates occur without triggering any formal review. This is not a niche concern — it is a structural compliance vulnerability with direct inspection implications.

A functional model change framework should start with a common set of definitions of what constitutes a change to be scrutinized. Broadly, the minimum should consist of retraining events that introduce new training data or modify the training dataset in ways that could affect model behavior, changes to feature selection, model architecture or hyperparameters; changes to the data pipelines feeding the model; and changes to thresholds or business logic for outputs. However, not all these events will require the same level of scrutiny. A risk-based approach, in which the depth of analysis depends on GxP impact and nature of change, is reasonable and consistent with CSA principles. [4]

Integrating AI model change management into existing change control systems is preferable to creating parallel processes. Most validated environments already have change control infrastructure capable of managing complex changes. The key adaptation is in the impact assessment: traditional assessments focus on technical implementation details, while AI change assessments should focus on GxP risk—specifically, whether the change could affect the accuracy, reliability, or interpretability of outputs used in regulated decisions. ISPE GAMP 5 provides a useful risk categorization structure for this purpose, aligning review depth with potential impact on regulated functions. [5]

### 6.2 Data Drift Detection and Performance Monitoring

Data drift describes the phenomenon in which the statistical distribution of data arriving at a deployed AI system changes over time in ways that may cause

model performance to deteriorate without any formal change event having occurred. The risk is particularly acute in pharmaceutical and biotechnology environments where manufacturing processes evolve, patient populations shift, and the operational context in which AI systems function is rarely static.

This risk is not disconnected from the data integrity principles regulated environments already apply. The FDA's data integrity guidance establishes that organizations must ensure the accuracy, completeness, and consistency of records throughout their lifecycle. [13] MHRA's data integrity framework requires that data be attributable, legible, contemporaneous, original, and accurate — principles that apply equally to data inputs feeding an AI model. [10] PIC/S PI 041-1 extends this logic into harmonized international expectations for data management governance in GMP and GDP environments, providing a multinational baseline against which organizations can assess the integrity of AI input data pipelines. [14] ISPE's GAMP Records and Data Integrity Guide further addresses the governance of electronic records in regulated environments, including capture, retention, and traceability requirements that apply to AI-generated outputs used as GxP records. [15]

The concept of continued process verification from FDA's process validation framework is the closest existing regulatory analog to ongoing AI performance monitoring. [18] Just as a validated manufacturing process must be actively monitored to confirm it remains in a state of control, a deployed AI system must be monitored against defined performance criteria to confirm it continues to produce outputs appropriate for their intended GxP use. Relevant metrics—flagging accuracy, confidence score distributions, and false negative rates on known high-risk scenarios—should be defined before deployment, tracked at defined intervals, and reviewed against thresholds that trigger investigation or retraining when breached.

**Table 4** presents a monitoring framework for AI system performance in GxP environments. As noted, the referenced guidance documents establish monitoring as a governance requirement but appropriately defer threshold-setting to organizational risk assessment rather than prescribing universal numerical values. The framework below reflects this correctly.

**Table 4. Monitoring Parameters for AI System Performance in GxP Environments**

| Metric Category | Example Metric | Monitoring Frequency | Threshold-Setting Basis | Escalation Pathway |
|---|---|---|---|---|
| **Predictive Accuracy** | Rate of confirmed true positives among flagged events | Monthly, additionally triggered by significant operational changes | Baseline performance established during validation; acceptable operational risk defined through risk assessment | Investigation; root cause analysis; potential retraining event initiation |
| **False Negative Rate** | Proportion of known high-risk events not identified by the system | Monthly; retrospective review during periodic performance assessment | Risk assessment reflecting patient safety and product quality implications of missed detections | Immediate investigation if safety-critical; potential system suspension pending review |
| **Confidence Score Distribution** | Statistical shift in model confidence outputs relative to validated baseline | Weekly trend analysis or continuous where infrastructure supports | Deviation from distribution baseline established during validation; statistical process control principles | Drift investigation; upstream data quality review; retraining assessment |
| **Input Data Quality** | Completeness, consistency, and integrity of data entering the model per processing cycle | Per processing cycle; integrated with data governance monitoring | Organizational data quality rules; data integrity principles per FDA,[13] MHRA,[10] PIC/S[14] | Data quality CAPA; upstream system review; deviation initiation if GxP record integrity affected |

| | | | | |
|---|---|---|---|---|
| **Model Version Integrity** | Confirmation that the production model version matches the change-control approved version | Per deployment event; confirmed in change control closure | Change control procedure; version management system | Change control escalation; production halt pending reconciliation |
| **Periodic Performance Review** | Holistic assessment of all above metrics against validated baseline performance | Quarterly minimum, or per monitoring plan; additionally after significant operational changes | Validated baseline performance; FDA CSA,[4] FDA Process Validation[18] governance principles | Revalidation determination; CAPA initiation; governance escalation to quality management |

*Note: Specific numerical threshold values are not prescribed by the cited regulatory guidance documents and are appropriately determined through organizational risk assessment as part of the validation lifecycle. Threshold values should be documented in the system's monitoring plan and reviewed as part of periodic performance assessments.*

*Sources: FDA CSA Guidance,[4] FDA Process Validation,[18] MHRA Data Integrity Guidance,[10] PIC/S PI 041-1,[14] ISPE GAMP® Records and Data Integrity.[15]*

### 6.3 Explainability and Decision Transparency

Explainability in AI governance discussions tends toward two unproductive extremes: the view that AI cannot be used in regulated decision-making unless its outputs are fully interpretable, and the view that explainability is irrelevant if performance metrics are adequate. The regulatory expectation sits in a more pragmatic space.

What regulators ask of organizations is not a technical explanation of how a model produces its outputs. They ask whether the organization can explain, with demonstrated competence and accountability, how AI outputs are used in the context of GxP decisions, what risks those outputs carry, and what safeguards exist to ensure outputs are not relied upon beyond the system's validated capability. This is an organizational and governance question, not an algorithmic one. EU GMP Annex 11 sets a consistent expectation in this space: computerized systems used in regulated operations must be validated, auditable, and subject to controlled access—requirements that apply to AI-enabled systems without modification.[8]

In practical documentation terms, explainability for GxP purposes means being able to state clearly: this system was designed to produce outputs in the form of Y; those outputs are used by qualified personnel to inform decision Z; and the following controls ensure that Z remains a human judgment rather than an automated outcome. For higher-risk applications, supplementary techniques such as feature importance analysis—which surfaces which input variables most influenced a given output—can provide additional audit and inspection value. The NIST AI Risk Management Framework provides a useful organizational structure for these governance decisions, distinguishing between contexts where interpretability is required for accountability and those where performance monitoring provides sufficient assurance. [7]

The principle of proportionality applies directly: the depth of explainability documentation required scales with the GxP impact of the system. An AI tool used for administrative drafting warrants minimal explainability documentation. An AI system influencing batch disposition decisions requires explicit documentation of how outputs are interpreted, reviewed, and acted upon — and who is accountable for each step.

## 7. Inspection Readiness: What Regulators Actually Assess

Regulatory inspections involving AI-enabled systems are becoming more frequent as adoption rates increase, and the pattern of regulatory focus is becoming clearer. Organizations that align their AI governance programs with these recurring areas of interest consistently perform better than those that approach AI inspections as they would a conventional CSV review.

Intended use is the first and most consistent area of focus. Inspectors will want to see a document that describes, in plain language, what an AI system is used for, what decisions it is used to support, and where it fits into the GxP workflow. Vague descriptions can be misleading. If system functions, decision types, human review requirements, and escalation procedures are specified, it is an indicator of capable, responsible, and accountable human

agency. Ambiguity at this level is one of the most common observations made in the inspection.

Risk-based justification is the second area. Inspectors assess whether validation choices were made deliberately, with documented reasoning, rather than by default or convention. ICH Q9(R1) principles apply here as directly as they do to any quality risk management activity: organizations should be able to articulate the rationale behind their control strategies, not merely describe the activities they performed. [3] The absence of documented rationale—why a certain testing scope was chosen, why specific thresholds were set, and why supplier documentation was judged sufficient—leaves organizations unable to defend their approach when it is questioned.

Ongoing assurance evidence is the third area and one that is increasingly scrutinized. Inspectors have moved beyond asking about initial validation activities to asking what has happened since deployment. Performance monitoring records, periodic review documentation, model change assessments, and deviation records involving AI outputs are all relevant. Proactive monitoring that identified and addressed a performance concern—even a relatively minor one—is typically received positively, as it demonstrates that governance controls function in practice, not only on paper.

Human accountability is the fourth, and arguably the most consistently assessed, area across inspections of AI-enabled systems. Inspectors look for evidence that AI outputs are understood to be tools that support human judgment, not substitutes for it. Clear role definitions, documented escalation procedures, and evidence that qualified personnel review AI-influenced decisions before final action all carry weight. The ICH Q10 quality system framework provides the organizational governance context for these accountability structures: AI assurance should be integrated into the quality system, not managed as a parallel or standalone compliance activity. [11]

An illustrative case from direct operational experience: an AI-assisted system deployed in a GxP manufacturing environment to support real-time process monitoring was assessed during an operational audit. The audit team's primary concern was not the technical performance of the model—performance data was available and adequate. The concern was documentation of how alerts generated by the system were reviewed, documented, and resolved. The absence of a defined workflow for alert disposition — specifically, the mechanism by which a system-generated flag transitioned into a formal quality event — created an observation. Developing an explicit procedure for that transition, which should have been part of the original validation package, resolved the finding. The lesson is clear: AI assurance is as much about governing what happens after a system produces an output as it is about validating the output itself.

## 8. Common Mistakes and How to Avoid Them

Several patterns of failure recur across organizations implementing AI in GxP environments. Understanding them is useful both for organizations building new AI governance programs and for those reviewing existing ones.

**Treating AI as categorically different from other GxP systems.** The instinct to build completely new frameworks for validating AI (separate policies, separate inventory categories, separate governance entities) often creates more problems than it solves. AI systems deployed in existing governance structures do have more strong oversight processes, but rather than trying to create a new system of oversight for AI, existing procedures such as change control, periodic review and risk assessment can be adapted to the particularities of AI systems. WHO good data and record management practices add to this, stressing that data governance frameworks should include all records generated in regulated activities, regardless of when or how they are generated. [16]

**Applying traditional CSV mechanics without modification.** Reflexive application of legacy validation templates to AI systems is equally problematic. Protocol-driven scripted testing that generates extensive documentation but fails to address model drift, retraining governance, or performance monitoring creates a false sense of assurance. Documentation volume is not evidence of control. The ISPE GAMP AI/ML guide addresses this directly, noting that AI governance must account for the dynamic nature of model behavior in ways that static validation approaches cannot. [6]

**Undefined lifecycle ownership.** In most AI systems, there are no owners for the performance, change evaluation, and active compliance of the AI system. This is perhaps the most consistent gap in AI governance analysis. Without ownership, lack of performance monitoring means retraining occurrences are not change events, and compliance posture is no longer protected by controls. Ownership should be established at the system

beginning and recorded in the validation plan , and should be reviewed whenever the organization's responsibilities change.

**Over-reliance on vendor documentation.** It is reasonable to leverage supplier-provided validation documentation for AI components, consistent with CSA principles. [4] It is not reasonable to treat vendor documentation as a substitute for organizational governance. Vendor SOC reports, model cards, and performance attestations provide useful supporting evidence, but they do not address how an organization deploys, governs, and monitors the system within its own regulated environment. The electronic records requirements of 21 CFR Part 11 apply to the organization, not the vendor an AI system generating records used to satisfy regulatory requirements must be governed accordingly, regardless of who built or hosts the model. [9]

**Insufficient data governance for AI inputs.** AI systems are only as reliable as the data they consume. Organizations that validate model outputs without governing the quality, integrity, and provenance of input data have addressed the symptom without the cause. Governance expectations apply to the full data lifecycle creation, processing, transfer, storage, and retention with equal force to the data pipelines feeding AI systems. [15],[10] This extends to the fundamental principle that all records in regulated operations, regardless of format or origin, must be attributable, legible, contemporaneous, original, and accurate. [16]

## Conclusion

The entry of artificial intelligence into regulated life sciences operations is not a trend to be managed at some future point—it is a transition already underway, and the industry's compliance infrastructure is actively calibrating to it. The gap between AI deployment and AI governance is real, but it is not unbridgeable. The path forward does not require new science. It requires disciplined application of existing quality principles to technology that does not behave the way those principles originally assumed.

Computer Software Assurance provides the most credible conceptual foundation for that work.[4] Its emphasis on critical thinking over mechanical compliance, on proportional controls over exhaustive documentation, and on ongoing assurance over front-loaded qualification aligns more naturally with AI's operational characteristics than traditional CSV ever could. The practical controls discussed in this paper model change governance, defined performance monitoring, explainability proportional to decision risk, and clear human accountability structures are not novel inventions. They are principled extensions of quality risk management frameworks that regulated environments have applied for decades, anchored in ICH Q9(R1), ICH Q10, and the GAMP architecture. [3],[11],[5]

What changes with AI is not the underlying logic of compliance. What changes is the nature of the risk being managed. AI systems can change without being changed, degrade without failing, and influence decisions in ways that are difficult to reconstruct after the fact. Governance frameworks that account for these characteristics—through active monitoring as codified in the FDA's process validation principles[18]; through data integrity governance reinforced across MHRA, PIC/S, and WHO expectations[10],[14],[16]; and through lifecycle accountability structures embedded in the quality system—are not constraints on AI. They are the conditions under which AI can be used responsibly in environments where patient safety and product quality depend on sustained, demonstrable control.

For organizations navigating the next generation of regulated AI deployments, the practical priority is not to build more elaborate validation packages. It is to build governance infrastructure that can be explained, defended, and sustained—inspection after inspection, model update after model update, as regulatory expectations around AI continue to evolve alongside the technology itself. [6]

## References


1. U.S. Food and Drug Administration. *Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device Action Plan.* U.S. Department of Health and Human Services: Silver Spring, MD, 2021. https://www.fda.gov/media/145022/download
2. European Medicines Agency. *Reflection Paper on the Use of Artificial Intelligence in the Medicinal Product Lifecycle.* EMA/676047/2023. Amsterdam, The Netherlands, 2023. https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf
3. International Council for Harmonisation. *ICH Q9(R1): Quality Risk Management.* ICH Harmonised Guideline. Geneva, Switzerland, 2023.

https://database.ich.org/sites/default/files/Q9%20Guideline.pdf
4. U.S. Food and Drug Administration. *Computer Software Assurance for Production and Quality System Software.* Guidance for Industry. U.S. Department of Health and Human Services: Silver Spring, MD, 2022. https://www.fda.gov/media/188844/download
5. International Society for Pharmaceutical Engineering. *GAMP® 5: A Risk-Based Approach to Compliant GxP Computerized Systems, Second Edition.* ISPE: Tampa, FL, 2022.
6. International Society for Pharmaceutical Engineering. *GAMP® Good Practice Guide: Artificial Intelligence and Machine Learning (AI/ML).* ISPE: Tampa, FL, 2023.
7. National Institute of Standards and Technology. *Artificial Intelligence Risk Management Framework (AI RMF 1.0).* NIST AI 100-1. U.S. Department of Commerce: Gaithersburg, MD, 2023. https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf
8. European Commission. *EU GMP Annex 11: Computerised Systems.* EudraLex — The Rules Governing Medicinal Products in the European Union, Volume 4. Brussels, Belgium, 2011. https://health.ec.europa.eu/system/files/2016-11/annex11_01-2011_en_0.pdf
9. U.S. Food and Drug Administration. *21 CFR Part 11: Electronic Records; Electronic Signatures — Scope and Application.* Guidance for Industry. U.S. Department of Health and Human Services: Silver Spring, MD, 2003. https://www.fda.gov/media/75414/download
10. Medicines and Healthcare products Regulatory Agency. *GxP Data Integrity Guidance and Definitions.* MHRA: London, United Kingdom, 2018. https://assets.publishing.service.gov.uk/media/5aa2b9ede5274a3e391e37f3/MHRA_GxP_data_integrity_guide_March_edited_Final.pdf
11. International Council for Harmonisation. *ICH Q10: Pharmaceutical Quality System.* ICH Harmonised Guideline. Geneva, Switzerland, 2008. https://database.ich.org/sites/default/files/Q10%20Guideline.pdf
12. Parenteral Drug Association. *Technical Report No. 84: Artificial Intelligence and Advanced Analytics in Pharmaceutical Manufacturing.* PDA: Bethesda, MD, 2022.
13. U.S. Food and Drug Administration. *Data Integrity and Compliance With Drug CGMP: Questions and Answers.* Guidance for Industry. U.S. Department of Health and Human Services: Silver Spring, MD, 2018. https://www.fda.gov/media/119267/download
14. Pharmaceutical Inspection Co-operation Scheme. *PI 041-1: Good Practices for Data Management and Integrity in Regulated GMP/GDP Environments.* PIC/S: Geneva, Switzerland, 2021. https://picscheme.org/docview/4234
15. International Society for Pharmaceutical Engineering. *GAMP® Guide: Records and Data Integrity in Regulated GxP Environments.* ISPE: Tampa, FL, 2017.
16. World Health Organization. *Guidance on Good Data and Record Management Practices.* WHO Technical Report Series, No. 996, Annex 5. WHO: Geneva, Switzerland, 2016. https://www.gmp-compliance.org/files/guidemgr/WHO_TRS_996_annex05.pdf
17. Topol, E.J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **2019**, *25*, 44–56.
18. U.S. Food and Drug Administration. *Guidance for Industry: Process Validation — General Principles and Practices.* U.S. Department of Health and Human Services: Silver Spring, MD, 2011. https://www.fda.gov/files/drugs/published/Process-Validation--General-Principles-and-Practices.pdf
19. European Medicines Agency. *Guideline on Computerised Systems and Electronic Data in Clinical Trials.* EMA/226170/2021. Amsterdam, The Netherlands, 2023. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-computerised-systems-and-electronic-data-clinical-trials_en.pdf
20. U.S. Food and Drug Administration. *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device.* Discussion Paper and Request for Feedback. U.S. Department of Health and Human Services: Silver Spring, MD, 2019. https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf