IntuitionLabs

INTUITIONLABS / RESEARCH & PRACTICAL GUIDANCE

Validating AI in GxP: GAMP 5 & Risk-Based Guide

> Build practical AI for pharma and biotech with IntuitionLabs. We help life-science teams turn complex information and workflows into useful software, governed knowledge systems and AI tools.

Published by IntuitionLabs · 2026-02-02 · ai validation · gxp compliance · gamp 5 · machine learning · fda csa · data integrity · computer system validation · eu annex 11 · risk management

Original article: https://intuitionlabs.ai/articles/validating-ai-gxp-gamp5-guide



In this article

1. Executive Summary
2. Introduction
3. Regulatory and Standards Context
4. Key Differences: Traditional Systems vs AI/ML Systems
5. Risk-Based AI/ML Validation Framework
6. Data Integrity and Quality Management
7. Case Studies and Examples
8. Implications, Challenges, and Best Practices
9. Future Directions
10. Conclusion

Executive Summary

The pharmaceutical, biotech, and life sciences industries are on the cusp of a digital transformation driven by **Artificial Intelligence (AI)** and **Machine Learning (ML)**. Generative AI tools alone are projected to add **$60–110 billion annually** to healthcare and pharma productivity (www.mckinsey.com). However, deploying AI/ML models in **GxP** (Good Practice) environments – where patient safety, product quality, and data integrity are paramount – introduces novel challenges. Traditional GxP computer system validation assumes **static, deterministic** software, whereas AI/ML systems are often adaptive, probabilistic "black-box" models that evolve with new data. This fundamental difference demands a tailored, risk-based validation strategy that extends existing frameworks like **ISPE GAMP® 5 (Good Automated Manufacturing Practice)**.

Regulators and industry groups have acknowledged this gap. The second edition of GAMP® 5 (2022) explicitly includes **Appendix D11 (AI/ML)**, prescribing an AI/ML life-cycle covering concept, project, and operational phases (www.scribd.com). A dedicated **ISPE GAMP® Guide: Artificial Intelligence**, published in July 2025, elaborates best practices (ispe.org). Meanwhile, regulatory expectations continue to evolve: EU Annex 11 and FDA 21 CFR Part 11 (which govern electronic record-keeping) remain fully applicable to AI systems, but recent FDA and EMA initiatives emphasize **risk-based Computer Software Assurance (CSA)** and data integrity in AI. For example, the FDA's final CSA guidance (Sept 2025) endorses risk-based validation and explicitly states that its framework **applies to AI tools used in manufacturing or quality systems** (www.hoganlovells.com). The draft EU **GMP Annex 22** (consulted on in 2025 and not yet adopted at the time of writing) and the EU **AI Act** add further expectations, although whether a given pharmaceutical AI use counts as "high-risk" under the AI Act depends on its classification, not on its use in a GMP setting alone.

Against this backdrop, this report offers a **practical, in-depth guide** to validating AI/ML models in GxP environments. We review regulatory contexts and standards, contrast AI-based systems with traditional validated systems, and outline a structured, risk-based validation lifecycle. Key recommendations include treating every AI dataset and output as a part of the regulated record (applying ALCOA+ data integrity), defining clear performance requirements (e.g. accuracy, bias tolerance), and implementing continuous monitoring ("drift detection") after deployment (academic.oup.com). We illustrate best practices with real-world

examples: GSK's "digital twin" for vaccine manufacturing, ML-driven water quality prediction, and AI-based quality management tools have already delivered significant efficiency and quality gains (Table 1). We also analyze emerging tools (GenAI for document drafting) and discuss future directions (federated learning, DevOps with validation).

Our analysis finds that **AI/ML validation is feasible and necessary**. By extending existing GAMP 5 and CSV principles to cover AI artifacts—such as model training data, version-controlled code, and performance metrics—companies can harness AI responsibly. This means risk-proportionate validation: mission-critical AI functions require exhaustive testing and controls, while lower-impact uses can adopt streamlined methods (e.g. unscripted or AI-assisted testing). Crucially, the fundamental goals do not change: systems must be *"fit for intended use,"* preserve patient safety, and ensure complete, accurate data. Achieving this will require close collaboration between IT, Quality Assurance, data science, and subject-matter experts. With **risk-based governance**, robust data pipelines, and thoughtful deployment practices, AI/ML can transform GxP manufacturing without compromising compliance or safety (www.ey.com) (www.mastercontrol.com).

Introduction

The emergence of AI and ML in regulated industries represents a watershed moment. AI-powered algorithms now support tasks ranging from predictive maintenance and process optimization to automated document review and regulatory writing. McKinsey's analysis underscores the scale of the opportunity: **"Generative AI is expected to produce $60 to $110 billion in annual value across the pharmaceutical industry value chain"** (www.mckinsey.com). In practical terms, companies report profound efficiency gains. For example, one AI vendor notes that multivariate models have cut batch deviation rates from 25% to under 10% and raised right-first-time production yield from 70% to over 90% (www.aizon.ai). Such projects not only reduce waste and errors; they bring predictive insights (e.g. anomaly detection) that shift quality assurance from reactive checks to proactive control. In regulated environments, these advances can **accelerate development and reduce shortages**, making medicines safer and more affordable.

However, the **pharmaceutical and medical sectors are among the most highly regulated in the world**. Good Practices—GMP (Manufacturing), GLP (Laboratory), GCP (Clinical) and others—mandate strict controls to ensure product safety, efficacy, and quality. Core regulations like FDA 21 CFR Part 11 (USA) and EU GMP Annex 11 (EU) require that *"computerized systems"* be validated and produce secure, integrity-preserved electronic records. Historically, this meant rigorous **Computerized System Validation (CSV)**: developers and users wrote exhaustive functional specifications and test scripts, performed Installation/Operational/Performance Qualifications, and documented every step in binders. The mindset was static: once validated, a software system was largely frozen, and any change triggered a formal revalidation.

Good Automated Manufacturing Practice (GAMP®) 5 (Second Edition) reframed CSV in the 21st century as a *risk-based lifecycle approach* (www.scribd.com). Instead of 100% testing on all code, effort is scaled to risk: focus on requirements, risk assessment, and critical functions. But even GAMP 5 was conceived before the AI boom. AI/ML systems defy the old paradigm: they *learn* from data, producing models whose inner workings may not be easily understandable or fixed in advance. When new data arrives, models often *adapt*

(exponentially complicating the validation picture). Outputs can be **probabilistic**, not deterministic, complicating traditional pass/fail criteria. In short, **AI systems have characteristics that do not map neatly onto traditional CSV templates** (www.scribd.com) (www.scribd.com).

For example, consider an AI used to detect defects in tablets via X-ray images. In a conventional CCD system, developers code explicit rules ("if mark > threshold, flag defect"), and validation testers simply check that rule on set examples. For an ML model, by contrast, the system *learns* what a "mark" or "defect" is from thousands of images. Its performance must be validated statistically (accuracy, false positive/negative rates) on independent data. If the factory introduces a new tablet type, the model may need retraining—and all retraining becomes a validated change. Moreover, with a deep neural net, understanding *why* a particular image was flagged can be very difficult. Regulators expect that even "black box" models are governed: one response is to restrict use to *explainable models* where possible, or to augment them with methods (LIME, SHAP) that shed light on decisions (www.mastercontrol.com). In any case, an implicit requirement emerges: **the validation scope must now include the entire ML lifecycle**—data sourcing, training, and ongoing monitoring (www.scribd.com) (www.scribd.com)—not just the deployed code.

This report provides a **comprehensive guide** to this landscape. We begin by reviewing the regulatory context (existing laws like Part 11/Annex 11, plus new AI-specific regulations) and industry standards (GAMP 5 2nd Ed., forthcoming AI guidelines). We then compare AI/ML systems to traditional GxP systems, identifying key validation differences. Next, we outline a *risk-based lifecycle framework* for AI/ML validation that integrates with standard GxP quality processes: from defining intended use and requirements through data management, model development, testing, and release, to operational monitoring and continuous validation. We emphasize data integrity (applying ALCOA+ to data pipelines) and risk management (identifying AI-specific hazards like bias and drift).

We bolster this analysis with **case studies and examples**. Projects at companies like **GSK, Recordati, and NTT Data** illustrate how AI can be harnessed in a validated way. (For example, GSK's 2020 Siemens/Atos "digital twin" pilot of vaccine manufacturing compared the virtual process with actual process data (www.pharmtech.com). Likewise, an ML water-quality monitoring pilot used historical sensor/lab data to forecast microbial contamination 24 hours in advance.) We also cite industry reports: surveys show that over half of pharma companies plan significant AI investments (especially for document automation) (www.pharmoutsourcing.com), and early adopters (e.g. Moderna) are deploying GPTs by the hundreds to streamline regulated workflows (www.pharmoutsourcing.com).

Finally, we discuss the broader implications. AI/ML hold great promise for product quality and efficiency, but they demand robust validation to ensure "trustworthy AI." We explore current challenges (explainability, data bias, resource needs) and best practices (e.g. AI-for-AI testing, cross-functional expertise). We conclude with forward-looking trends: forthcoming PIC/S and ISO standards, the EU AI Act, and evolving FDA/EMA thinking will further shape the duty to validate AI. In sum, AI/ML validation in GxP is **both essential and achievable**: it builds on existing quality principles (fit-for-purpose, patient safety, ALCOA+ data) while embracing new techniques (risk-based assurance, automated testing, continuous monitoring). With careful implementation, companies can unlock AI's benefits in manufacturing without compromising compliance or safety (www.ey.com) (www.mastercontrol.com).

Regulatory and Standards Context

AI/ML systems in GxP environments must satisfy **all existing regulations** for computerized systems, plus emerging AI-specific guidance. Critically, regulators have made clear that *nothing exempts AI from current rules*: FDA's Part 11 and the EU's Annex 11 requirements for validated, audited electronic records apply just as much to AI models and data streams.

Traditional GxP mandates include:

- **FDA 21 CFR Part 11 (Electronic Records/Signatures)** – U.S. regulation requiring that equivalent electronic records have the integrity and audit trail of handwritten records. This applies whenever ALCOA criteria are needed (i.e. nearly any record related to GxP activities). AI inputs and outputs are "records" in this sense: prompts, data, and results must be traceable, timestamped, and protected with secure access/authentication. In practice, this means treating what the AI model receives and produces as auditable records (www.auriacompliance.com) (www.ey.com).
- **EU GMP Annex 11 (Computerized Systems)** – Similar to Part 11, Annex 11 requires formal validation (IQ/OQ/PQ) of computerized systems to show they *"consistently perform as intended"* (www.ey.com). Annex 11 covers system impacts, backup/disaster recovery, change control, etc. It does **not** explicitly mention "AI," so companies are advised to *treat AI models as part of the computerized system.* For example, any model retraining or code change must go through change control per Annex 11; any model output used in decisions must be captured via audit trail. As EY notes, *"All computerised systems [under Annex 11] undergo formal validation… a structured, documentation-heavy top-down approach"* (www.ey.com). Practitioners therefore apply Annex 11 controls to AI pipelines: performing risk assessments on any ML function, including data integrity controls (audit trails, backups) not just on the hosting environment but on the data pipelines that feed the model.
- **ICH Q9 (Quality Risk Management)** – The ICH Q9 guideline is international risk management guidance underpinning a GMP approach. It mandates that all potential hazards to product quality be identified and controlled. In AI/ML, experts recommend explicitly treating model bias, drift, explainability gaps, and data deficiencies as risk factors under Q9. The guidance does not list AI, but industry writers emphasize that *"new AI risks: model bias, dataset completeness, explainability"* must be folded into the risk assessment plan. In other words, organizations should ask: could a biased model cause harm? Does lack of explainability threaten safety? These become part of the quality risk management (QRM) analysis for an AI project.

In practice, firms map emerging AI factors onto familiar frameworks. Industry commentary makes the same point: Annex 11 and Part 11 still apply, and their controls extend into model training pipelines, cloud platforms, and retraining events (www.scribd.com). For example, a training dataset is treated like a "production batch" that must be documented, protected, and undeleted. Every model version is labeled and archived (often using code version tools like Git). If a model misbehaves or drifts, it can trigger a CAPA under the quality system. Thus, compliance remains anchored in GxP law, only expanded to cover AI's unique objects.

Emerging guidance is rapidly filling the gaps:

- **ISPE GAMP® 5 (Second Edition, 2022)** – The world's primary CSV guidance now explicitly addresses AI/ML. The new **Appendix D11 (AI/ML)** introduces an ML-centric system life cycle (concept, project, operation) that parallels traditional GAMP phases (www.scribd.com). It calls out AI tasks: defining performance metrics, training/validation protocol, and continuous monitoring. GAMP 5 still underpins the overall approach, but Appendix D11 signals that data and model must be considered system components. ISPE has since published the **ISPE GAMP® Guide: Artificial Intelligence** (July 2025), which gives in-depth recommendations (e.g. data bias mitigation, supplier expectations, AI design principles) and bridges general GAMP concepts with AI characteristics (www.scribd.com) (www.scribd.com).
- **FDA Computer Software Assurance (CSA) Guidance (2025)** – FDA's latest CSV approach, CSA, shifts from exhaustive testing to risk-based assurance (www.hoganlovells.com). The final CSA guidance (Sept 2025) covers software used as part of medical device production and quality systems (not device software functions themselves); drug manufacturers often apply the same risk-based thinking, although the guidance is written for device makers. It stresses starting with *intended use* and categorizing each feature as "high process risk" or not (www.hoganlovells.com). Importantly, FDA clarified that the CSA framework **can be applied to AI**: *"manufacturers can apply the framework to artificial intelligence tools, again, if used as part of production or quality systems."* (www.hoganlovells.com). Thus under CSA, an AI tool that directly controls a production step or a quality decision (high process risk) would get rigorous testing, whereas a lower-risk supporting tool might need only scenario-based checks. FDA encourages leveraging vendor certifications, audit logs, and test configurators (e.g. automated scripts) over manual proof, which aligns well with AI/ML workflows. CSA explicitly endorses approaches like digital evidence (audit trails, SBOMs, vendor assessments) to support validation results (www.hoganlovells.com) (www.hoganlovells.com).
- **EMA/FDA AI Initiatives** – Though not GxP law, agencies are issuing guidance on "trustworthy AI." FDA's *SaMD Action Plan* and EMA's *Reflection Paper on AI* focus on factors like transparency and human oversight, signaling expectations. For example, the FDA's SaMD (Software as a Medical Device) guidelines recommend Good Machine Learning (GMLP) principles: controlled training data, continuous learning oversight, and output reproducibility. The EU's proposed **GMP Annex 22** (AI), put out for public consultation from July to October 2025, would codify many AI principles; commentary on the draft notes that it covers static models with deterministic outputs in critical GMP applications and excludes dynamic, probabilistic and generative AI models from such uses (www.europeanpharmaceuticalreview.com). It had not been adopted at the time of writing. Meanwhile, the **EU AI Act** (Regulation (EU) 2024/1689, in force since August 2024) reserves its strictest obligations for AI systems classified as "high-risk", for example AI in regulated products such as medical devices (eur-lex.europa.eu). AI used in pharmaceutical manufacturing or pharmacovigilance is not listed as high-risk per se, so classification must be assessed case by case, and the high-risk obligations apply in stages. In practice, regulators expect firms to integrate AI concerns into existing GxP norms: e.g. augment Annex 11 change control to include retraining events, or treat LLM prompt/output logs as regulated records (since they affect decisions) (www.ey.com) (www.mastercontrol.com).

- **International Guidance (PIC/S, ISO, etc.)** – PIC/S (the Pharmaceutical Inspection Co-operation Scheme) took part in the 2025 consultation on the draft Annex 22, signalling that AI expectations are being harmonised across its member inspectorates; in the US, FDA's CSA guidance plays a comparable risk-based role (www.hoganlovells.com). Similarly, ISO has published standards (e.g. ISO/IEC 42001:2023 for AI management systems) that, while non-binding, will shape quality system expectations. The British regulator MHRA and others have also flagged ALCOA+ and AI oversight in inspectorates. The biopharma industry is aligning around a common message: AI/ML models must be treated as critical system components, and validated *in proportion to their impact*.

Summary Table: Regulations and AI/ML in GxP

Regulation / Guidance	Scope	AI/ML-Specific Considerations
21 CFR Part 11 (FDA)	Electronic records & signatures (USA)	AI prompts, training data, and model outputs must be captured in audit trails. Ensure ALCOA+ compliance for all data: timestamps, user IDs, versioning (audit logs).
EU GMP Annex 11	Computerized systems (EU)	Apply Annex 11 controls to AI pipelines: e.g. include model retraining in change control; risk-assess impact of AI on processes; maintain traceable data (training sets) in the Product Quality Review.
ICH Q9 (QRM)	Quality risk management (global)	Include AI-specific hazards (data bias, model drift, lack of explainability) in risk assessments. Ensure training data covers intended population and safety scenarios.
ISPE GAMP 5 (2nd ed.)	Risk-based CSV lifecycle (global)	Contains new Appendix D11 for AI/ML. Recommends an AI life cycle (concept, project, operation) and roles (model owner, SMEs). Encourage continuous monitoring of model performance.
ISPE GAMP Guide: Artificial Intelligence (July 2025)	Best practices for AI-enabled computerized systems in GxP	Consolidates GAMP + AI knowledge. Provides design principles (data quality, bias mitigation), supplier expectations, and a framework for verifying AI-enabled systems.
FDA CSA Guidance (2025)	CSA for production/QMS software (USA)	Endorses risk-based testing (focus on high-risk functions). Explicitly applies to AI tools: recommends starting from intended use and adjusting validation effort by product-safety risk (www.hoganlovells.com) (www.hoganlovells.com).
EU AI Act (Regulation (EU) 2024/1689)	Regulation on AI (EU, phased application)	AI in regulated products such as medical devices (e.g. some clinical decision support) can be "high-risk", requiring documented risk management, data quality and conformity assessment. Manufacturing and pharmacovigilance uses are not high-risk per se and need case-by-case classification.
EU GMP Annex 22 (draft)	AI in GMP manufacturing (EU, PIC/S consulted)	Draft consulted on July to October 2025; covers static, deterministic models in critical GMP applications and excludes dynamic and generative models from them. Not yet adopted at the time of writing.

Table 1: Regulatory framework for AI/ML in GxP. Traditional laws (Annex 11, Part 11) apply fully to AI systems and data; evolving guidance (e.g. GAMP AI, CSA, AI Act) emphasize risk governance, data integrity, and ongoing verification (www.ey.com) (www.hoganlovells.com).

Key Differences: Traditional Systems vs AI/ML Systems

Before detailing validation steps, it is instructive to contrast AI/ML systems with conventional GxP software. Table 2 highlights the principal differences and corresponding validation actions. In traditional systems, requirements are static (functional specs, user stories) and code is deterministic. By contrast, AI/ML introduces:

- **Dynamic behavior:** ML models “learn” from data rather than operate on fixed code. Acceptable output distributions, not exact values, become the validation target. Thus testing revolves around statistical metrics (accuracy, precision, recall, AUC) on reserved validation sets, rather than fixed expected outputs (academic.oup.com).
- **Data-centric risk:** The inputs (training/validation data) become as important as the code. Data must be curated, cleaned, labeled, and integrity-checked (complete, consistent, representative). Every dataset is treated as an original GxP record – subject to ALCOA+ requirements (www.auriacompliance.com). In traditional CSV, validation often assumes good data; with ML, data *quality* is an explicit validation focus.
- **Continuous change:** Traditional validated systems change infrequently. AI models typically degrade over time as operational data diverges from training data (a phenomenon known as *data drift*). This means validated AI/ML systems require ongoing monitoring of performance and triggers for retraining. In practice, AI validation plans incorporate drift detection (e.g. statistical tests on new data distributions) and define when a new training/validation cycle constitutes a “change” that must be re-validated.
- **Explainability:** Conventional software behaves per documented logic. Many ML models (e.g. deep learning) are opaque: their decision logic can be inscrutable. This challenges the traditional “binocular” validation approach (document algorithm and test exhaustively). Industry best practice is to use interpretable models when possible, or to add explanation layers (e.g. LIME/SHAP) that approximate decision rationales (www.mastercontrol.com). In either case, validation must include expert review: e.g., showing SMEs what features drove a classification to ensure it aligns with domain knowledge.
- **Versioning:** In classic CSV, version control applies to code and configuration. For ML, version control must also apply to models and data snapshots. Every time a model is trained or re-trained, that model (and the precise data used) should be labeled, stored, and characterized. This enables traceability: regulators may ask, “Which data and model version produced this alert?” and companies will have documented answers. Git-like versioning of model files (and code notebooks) is increasingly recommended.
- **Documentation:** Beyond the usual URS, DQ. AI projects add new document types. For example, a **Model Design Specification (MDS)** or **Data Integrity Plan** may be created. The MDS defines the AI concept, architecture, and performance targets. The Data Integrity Plan records data sourcing, cleaning, and quality control procedures. Validation artifacts include not only test scripts but also things like data characteristic reports (distribution histograms, outlier analysis), code reviews of training scripts, and details of hyperparameter choices. All of this becomes part of the validation record.

This comparison suggests that while the *framework* of V-model validation remains (define scope → design → develop → test → operate), the *content* of each phase shifts in AI/ML projects. Rather than enumerating lines of code, one enumerates model features and performance indicators. Rather than manual scripts, one uses large-scale metric validation. Nevertheless, the underlying goal is unchanged: ensure the system is **"fit for intended use"** and does not endanger patients or product quality.

Table 2: Key Validation Activities – Traditional GxP Systems vs AI/ML Systems

Aspect	Traditional GxP System	AI/ML System
Requirements	Fixed functional & technical requirements (URS, FRS).	Additionally define ML objectives and acceptance criteria: e.g. accuracy targets, allowable bias levels, model explainability needs (academic.oup.com).
Design/Development	Deterministic code configured to spec.	Iterative model development: test multiple algorithms, tune hyperparameters. Each model version (with architecture/hyperpars) is documented (www.scribd.com) (www.scribd.com).
Data Handling	Input data from controlled sources; audit trails on data entry.	Massive training datasets apply. Data must be curated/cleaned and treated as GxP records (ALCOA+). Partition data into training/validation/test sets with strict controls to prevent leakage.
Testing/Validation	Predefined test scripts with known expected outcomes.	Evaluate model on held-out validation set: compute performance metrics (accuracy, sensitivity, specificity, MAE, ROC AUC, etc.). Perform bias checks (splitting test results by subgroup). Continual verification: re-run critical data cases to detect performance drift (academic.oup.com) (nttdata-solutions.com).
Change/Version Control	Formal change control. New code version is validated before release.	Model degradation expected. Maintain versioning: tag each model and its training snapshot. Retraining is treated as controlled "change." Use CI/CD or MLOps pipelines to capture code/data changes; retrospective review of drift triggers retraining along CSA lines.
Operational Monitoring	Periodic reviews (IQ/OQ/PQ); quality reviews at intervals.	Continuous monitoring of live performance: track key metrics (e.g. daily accuracy or anomaly rates). Set thresholds to trigger review or retraining► (e.g. "if accuracy drops below X% for 5 batches, review model"). Tools may regularly score model against sampled ground truth.
Risk Management	Focus on software failure modes (bugs, security, downtime).	Include AI-specific risks: model bias or blind spots, training data gaps, adversarial attack vulnerability. Evaluate scenarios like *if model fails, what harm could occur?* Factor these into QRM (ICH Q9) and link mitigations (e.g. human review in loop).
Documentation	VMP, URS, Design Spec, Test Scripts, etc.	Extend docs: add Model Design Spec, Data Management Plan, Model Qualification Protocol. Capture data provenance, feature engineering notes, and explainability assessments (www.mastercontrol.com).

Table 2: Comparison of validation considerations for traditional GxP systems vs AI/ML-enhanced systems. Reporting requirements now include AI data pipelines and performance metrics, even as risk-based approaches allow more flexible testing for lower-risk models (www.hoganlovells.com) (academic.oup.com).

Risk-Based AI/ML Validation Framework

Validated deployment of AI/ML in GxP environments hinges on a **structured, risk-based framework**. Leading sources (ISPE GAMP, FDA, industry experts) describe a multi-step lifecycle that mirrors traditional CSV but explicitly adds AI tasks (www.scribd.com) (www.hoganlovells.com). A representative AI/ML validation workflow consists of the following phases:

1. **Context of Use & Risk Profiling.** Clearly define exactly *how* the AI model will be used, and the potential impact of errors. Differentiate whether an algorithm is used for safety-critical decisions (e.g. dosing, contamination alerts) versus informational/reporting tasks. Under a risk-based approach such as FDA's CSA, if an AI influences a **high-risk** process step, it needs rigorous assurance; if it merely supports internal analytics, assurance effort can be scaled back (www.hoganlovells.com). Typically, the team develops a risk assessment at project start: list hazards (e.g. "false negative defect detection might allow substandard product") and quantify consequences. This determines the scope: high-risk features get more testing and stricter controls. (This approach directly implements FDA's CSA and ICH Q9 principles: *"intended use drives validation scope"* (www.hoganlovells.com).)

2. **Requirements and Performance Criteria.** Draft an AI-specific User Requirements Specification (URS) and Functional Requirements. These go beyond common features ("classify X"): they include *model-centric* requirements. For instance, one might specify: "The ML model shall classify tablet images with ≥95% accuracy and ≤2% false negative rate, and maintain this performance across all defined batch types." Non-functional specs include data constraints (types, volume, privacy), computational performance (latency), and explainability needs (e.g. ability to inspect top-10 influential features in each decision). Crucially, performance metrics must be pre-defined: accuracy, precision/recall thresholds, RMSE limits, etc. These quantitative targets become the acceptance gates for validation (see GAMP D11 guidance which stresses early definition of metrics (www.scribd.com) (www.scribd.com)).

3. **Data Strategy and Set Construction.** AI validation is only as strong as its data. Teams must inventory all data sources (sensor logs, lab results, EMR data, etc.) and assess suitability: Are there enough cases? Is the data biased? Preprocessing (cleaning, normalization, de-identification) is documented as controlled activities. A **Data Integrity Plan** is developed, treating datasets and labels as regulated "materials" under ALCOA+. For example, any external or third-party dataset must have proof of rights and anonymization. All data transformations (scaling, encoding) are version-controlled to ensure reproducibility (www.auriacompliance.com) (nttdata-solutions.com). The data is then partitioned into non-overlapping sets: training (for model building), validation (for tuning hyperparameters), and test (for final evaluation) (nttdata-solutions.com). This partitioning is critical: the test set must remain completely untouched until final performance evaluation to avoid information leakage. If needed, additional data collection projects may be initiated to fill gaps noted in risk assessment (for example, if edge cases are underrepresented).

4. **Model Development & Engineering.** With data ready, data science or engineering teams iteratively develop candidate models. This typically follows an AI development cycle (design → train → evaluate → redesign) rather than a strict waterfall. Teams experiment with multiple algorithms (random forests, neural nets, SVMs, etc.) and tune hyperparameters (learning rate, tree depth, epochs) to optimize on

the validation set (www.scribd.com) (nttdata-solutions.com). Each training run is logged, capturing versioned code, parameter settings, and performance metrics. Automation tools (AutoML frameworks) can accelerate this, but manual oversight is still needed. Part of this phase is embedding explainability: e.g., if using a deep network, the team might also compute SHAP values or train a simpler surrogate model to help interpret decisions. Throughout, quality assurance monitors data integrity: verifying that no test data has crept into training, and that preprocessing scripts are correct. The result of this stage is one or more candidate model "artifacts," each with a clear performance report.

5. **Model Testing and Selection.** The candidate models are then rigorously evaluated on **held-out test data**. Statistical "scorecards" are produced: accuracy, ROC-AUC, confusion matrices, bias metrics, etc. Teams inspect not only aggregate scores but also failure modes: for classification, reviewing specific false positives/negatives can reveal model blind spots (e.g. the model misidentifies container abnormalities on vials of a certain color). If performance meets the predefined criteria, the best model is selected for release. Qualitative review by domain experts is essential here: an expert might catch systematic errors or unjustified correlations that metrics alone could miss. Integration testing follows: verifying the model's ability to run in the target environment, consume real-time inputs, and hand off outputs to downstream systems. Any code wrappers (for data pre/post-processing) are also tested. This phase effectively serves as the **verification** step: confirming that the model meets all user requirement specifications under controlled conditions (nttdata-solutions.com) (www.mastercontrol.com).

6. **Formal Validation and Release.** Now the selected model undergoes official validation documentation. A Validation Plan is written (covering model development, data, and environment aspects). During the Independent Verification (or IQ/OQ/PQ) phase, test cases are executed: for instance, re-running the final model on the test set to reproduce performance metrics, checking that model outputs match previously logged results. Any stochastic elements (random weight initialization, parallel execution) should not unduly affect outcomes; reproducibility checks ensure that reruns are within acceptable tolerance. Additional stress tests might be applied (e.g. adding noise to inputs to observe robustness). All results are documented. Approval leads to formal release: the model, along with its version, training data snapshot, and validation record, is moved into production status under controlled release status.

7. **Operational Deployment and Monitoring.** After go-live, AI models depart from traditional static deployment. As new real-world data flows in, model performance must be continuously monitored. KPIs (e.g. daily accuracy, false alarm rates, input distribution statistics) are tracked in dashboards. If drift is detected (see below), alerts are generated. Governance states the retraining policy: e.g. if a performance KPI breaches its threshold for N consecutive production runs, a retraining cycle is triggered. This retraining is treated as a controlled change: the model goes through steps 3–6 again with updated data (whereby regressing to earlier lifecycle steps under change control). Periodic audits compare model outputs with ground truth to ensure ongoing validity. This "monitor & manage" phase embodies Quality by Design for AI: the goal is to preserve the spirit of traditional validation (fitness for intended use and data integrity) even as binders give way to dashboards, pipelines, and neural networks (www.auriacompliance.com) (www.mastercontrol.com).

8. **Supporting Quality Processes (CAPA, Change Control, Training).** Throughout the lifecycle, standard GxP quality processes are adapted. The risk management plan is revisited as new hazards appear (e.g., cybersecurity attacks on the model or data breach possibilities). The change control SOP explicitly describes how model retraining, algorithmic adjustments, or data pipeline updates are authorized and documented. Training programs educate data scientists on GxP norms (and QA on AI concepts). CAPA systems may be used to address ML issues (such as mislabeled training data), just as they are used for equipment deviations. Ultimately, an AI project is not exempt from CAPA: each anomaly or incident triggers investigation to determine if there's a root cause in data or model logic.

This life-cycle framework emphasizes **proportionality**: not all AI systems require the same intensity of validation. High-impact uses (e.g. dose determination, direct patient management) mandate thorough analyses and controls (www.hoganlovells.com), whereas lowimpact applications (e.g. internal inventory forecasting) may be handled with streamlined checks. The overarching principle is that “the level of assurance should align with risk” (www.hoganlovells.com). By following the above steps—adapted from GAMP 5 Appendix D11 and CSA guidance—companies create a compliant, transparent pathway for AI deployment in GxP.

Data Integrity and Quality Management

Among all aspects of AI validation, **data integrity** is paramount. Data integrity has long been central to GxP expectations, and for AI/ML it matters even more: a model is only as good as the data it learns from. Regulatory agencies and experts stress that AI pipelines must adhere to the same ALCOA+ principles as any other GxP record (www.auriacompliance.com) (www.ey.com).

Concretely, this means that **every data element** in an AI project is treated as a regulated record: raw sensor readings, patient data, annotated images, training labels, and even intermediary features count. In practice, quality teams must ensure: Attribution (who captured which data?), Legibility (are data in readable formats?), Contemporaneity (are timestamps correct?), Original and Accurate (has the data been altered?), plus the “+” aspects (Complete, Consistent, Enduring, Available) (www.auriacompliance.com) (www.auriacompliance.com). For example, an annotated image used for defect detection (say, an X-ray of a tablet with defect contours drawn by a human) is an original document that must be archived in full. If the model is retrained, the exact snapshot of the training set at that time should be frozen (perhaps hashed and time-stamped) and retained.

Data Quality Control: Initial steps involve rigorous data QC. Data undergo normalization, outlier removal, deduplication, and labeling error checks (www.auriacompliance.com). Bias detection is critical: if older batches dominate the training data, the model may fail on new products. Domain experts work with data scientists to identify and correct such biases. All data transformations (e.g., imputation of missing values, augmentation) are documented in the project records. Pipeline software often notes e.g. “training_data_v1: normalized and scaled to [0,1], removed measurement errors >3σ” with timestamps. If external data is imported (e.g. public datasets), legal clearance and privacy (GDPR/HIPAA) must be confirmed – a task analogous to supplier qualification.

Data Partitioning: As described above, the data is split into training, validation, and test sets (nttdata-solutions.com). This partitioning itself must be auditable: version-controlled splits (e.g. using random seeds) ensure the test set remains isolated. A simple mistake (accidental overlap between training and test) would

invalidate the whole validation. Industry best practice (as reinforced by GAMP D11) is to guard this separation as stringently as manufacturing controls: e.g. store test set on a read-only medium, and give only the verification team access at the release phase. Some firms even conduct blind challenges, where an independent auditor supplies the test data.

Model Input/Output Logging: In a deployed AI system, all live inputs and outputs should be logged in detail. This is akin to saving every paper copy of a form or reagent measurement. For instance, if an AI model processes batch sensor data and outputs a contamination alert, the exact sensor readings, model version, and alert result are saved. Using a centralized logging platform (audit trail database) is recommended. This way, if later a question arises ("why did batch 47 trigger a hold?"), one can reconstruct the decision exactly. Tools like MLflow or custom MLOps platforms often provide model version tracking and run logs that satisfy this need.

Auditing Data Pipelines: Regulatory inspections will scrutinize not just code but data lineage. Inspectors may ask: "How do you know this data wasn't tampered with?" Standard controls – user permissions, checksums, encryption, and periodic review – are needed. For cloud-based services, vendors' compliance documents (e.g. SOC 2, ISO 27001) and data retention policies are reviewed. Pharmaceutical companies often perform vendor audits or rely on validated data connections (APIs) rather than manual file transfers. For example, if a LIMS injects data into a model, an automated link might tag each record with origin metadata. Such measures "lock" the data lineage and become part of the audit trail.

Demonstrating Trust: A useful analogy is that trustworthy AI requires treating data with the same rigor as any regulated product component (www.auriacompliance.com). In other words, one approaches the AI's dataset like an active pharmaceutical ingredient: with version records, quality testing, and chain-of-custody. The burden is high: companies might spend months reconciling sensor logs with lab results to ensure that a model actually learned the correct cause-effect (as in the water-quality case, where conductivity changes were aligned to microbial counts) (nttdata-solutions.com). But this effort pays off: with clean, traceable data, models can be more reliably validated, and regulatory confidence is bolstered.

Case Studies and Examples

Real-world implementations illuminate both the benefits and the practicalities of AI in GxP settings. Below are selected case studies and industry examples that illustrate key points.

- **GSK's Vaccine Digital Twin.** GlaxoSmithKline partnered with Siemens and Atos to develop a real-time "digital twin" of its vaccine manufacturing process (www.pharmtech.com). This virtual model ingests actual sensor data (temperature, pressure, pH, etc.) and simulates the process flow, with AI-driven optimization in the loop. The 2020 project was a proof-of-concept pilot, which the partners reported as achieved, with GSK aiming to apply the concept at industrial scale. The published account does not describe a GMP validation, but any GMP deployment of such a twin would require its outputs to be **tested against real process data** to verify fidelity (www.pharmtech.com) (www.pharmtech.com). The stated aim is improved process control: as Matt Harrison of GSK Vaccines put it, *"Controlling variability allows us to improve quality and make product 'right first time' every time."* (www.pharmtech.com). This example shows a **hybrid process-model** whose fidelity is checked by comparing virtual outputs with real measurements, blending ML model validation with conventional process validation.

- **Predictive Water Monitoring (NTT Data case).** At a biomanufacturing plant, NTT Data engineers used ML to improve environmental monitoring of purified water (a critical GMP utility). Traditionally, water quality was checked by daily lab samples (e.g. microbial assays); NTTDATA leveraged already-available sensor logs (organic carbon, conductivity, flow etc.) to **predict high microbial counts 24 hours in advance**. They performed a POC: years of historical sensor plus lab results were split into training and test sets, enabling a classifier to be built. Validation involved confirming that model forecasts matched actual lab outcomes. The approach aimed to give earlier contamination alerts with no additional sensors. It highlights key lessons: (a) **High-resolution data is gold** – unexploited sensor logs became a new process control tool; (b) proper data management (timestamp alignment of lab and sensor data) was crucial, effectively merging two "systems of record" in a traceable way (nttdata-solutions.com); (c) the validation plan looked much like a statistical experiment: define hold-out data, and assess model score against real results. In such a setup, the ML model supplements, rather than replaces, traditional sampling with predictive alerts.
- **AI in Quality Management Systems (QMS).** AI tools are increasingly embedded in broader quality functions. A quality-management software company reports clients using AI for **document review, trend analysis, and CAPA management** (quality.eleapsoftware.com) (quality.eleapsoftware.com). For instance, natural language processing (NLP) can instantly extract key data from free-text batch records and flag issues, radically reducing manual review time (quality.eleapsoftware.com). One case saw AI-assisted investigations complete 50–70% faster by accelerating root-cause analysis (quality.eleapsoftware.com). Companies also use predictive AI for QMS: e.g. predictive CAPA systems that score which deviations are likely to recur. The vendor's commentary argues that when AI clarifies traceability (an auditable rationale for each decision), inspectors are more receptive (quality.eleapsoftware.com). These examples show "downstream" validation: while not manufacturing-critical, they still required CSV-like evidence. In practice, firms approached these as GxP systems (e.g. validating AI batch-report reviewers by comparing flagged issues against human review gold-standards).
- **Generative AI for Regulated Document Writing.** Generative models (like GPT-4) are rarely used for production decisions, but are increasingly deployed for document tasks. For example, some companies use LLMs to draft clinical trial protocols or pharmacovigilance narratives. A survey reported that ~12% of firms are piloting AI for automated report generation (www.pharmoutsourcing.com). Early evidence suggests huge labor savings: one study cited in a related IntuitionLabs article reported that an LLM could cut PV report writing time in half. In response, validation practices have emerged: firms track the model version and prompt history for each AI-generated section, and require subject-matter expert sign-off on all outputs. The QA process might include plagiarism checks and factual accuracy tests. Conceptually, this is *validation of AI as a content tool*: audit trails (AI logs) must be kept, and output is always reviewed by a qualified person. While not a traditional manufacturing system, it reflects the expectation that AI handling regulated content be subject to GxP-appropriate controls.
- **Vendor Customers (Aizon case studies).** Start-ups and vendors report practical ROI figures. For example, Aizon (AI for manufacturing) highlights customer outcomes: automated Product Quality Review generation, golden-batch optimization, and deviation reduction (www.aizon.ai). The vendor reports customer outcomes such as batch deviations falling from 25% to under 10% and right-first-time batch success rising from 70% to over 90% (www.aizon.ai). These vendor-reported numbers, not

independently verified, illustrate the scale of impact. They also imply intensive validation: such gains required first building trustworthy models. The blog also paraphrases co-founder Toni Manzano as stressing that strong validation practices are needed to keep AI solutions compliant and trustworthy.

These cases converge on practical themes: **(a) interdisciplinary teams** – data scientists, engineers, and QA personnel worked together; **(b) pilot/proof-of-concept use** – all began with small-scale POCs to verify viability before full GMP deployment; and **(c) leveraging existing data** – none of these projects required curing new novel error modes, but made better use of data already being collected. Importantly, they highlight that AI benefits are often found at interfaces: linking production data with quality outcomes to create proactive controls.

Implications, Challenges, and Best Practices

The case studies and evolving guidance illustrate the **promise and pitfalls of AI in GxP**. On one hand, AI can vastly improve efficiency and quality. The proactive, data-driven insights shift compliance toward prevention. The implications are discussed by analysts: one report suggests generative AI and other advanced tools *"offer dramatic improvements"* for operational processes (www.mastercontrol.com). McKinsey estimates that generative AI could produce $60–110 billion in annual value across the pharmaceutical value chain (www.mckinsey.com). If realized, this means regulators and quality professionals could refocus on oversight rather than menial tasks. ROI can be significant: for example, AI-based document processing has been projected to cut quality labor costs by **20–30%** (quality.eleapsoftware.com).

However, many **challenges** complicate realizing this promise:

- **Explainability and Trust.** AI models – especially deep neural networks – often lack transparent decision processes. Regulators emphasize the need for human oversight and interpretability. In practice, teams must incorporate explainability measures. One approach is to choose simpler models (decision trees, linear models) when possible. If using "black box" models, explanation tools (LIME, SHAP) can highlight which features drove a decision (www.mastercontrol.com). Another tactic is to output confidence scores alongside decisions: any low-confidence prediction is flagged for manual review. Critically, validation documentation should not only state that a model passed its performance tests, but also capture insights on *how* the model reasons. For instance, review notes might record that experts found the model's feature importances reasonable. Without such transparency, auditors may question *why* an AI made a particular call; integrating domain knowledge within the validation evidence is key to building trust.
- **Data Bias and Representativeness.** AI effectiveness hinges on data quality. If training data is biased (e.g. missing capsule color variant), the AI can make systematically unsafe decisions. Therefore, risk assessments must explicitly consider "bias risk" as a hazard (academic.oup.com). Mitigations include assembling diverse training sets that cover all relevant conditions (e.g. all sites, equipment models, variations in raw materials). Statistical bias tests (disparity in errors across subgroups) should be part of acceptance criteria. Ongoing monitoring should include examining model performance by batch, shift, or

other breakthrough factors to catch any selective underperformance. In short, governance must treat the AI's training data as a controlled enrichment project: it often requires clearing ethics/consent, confirming data lineage, and validating labels just as one would any critical component.

- **Resource Intensity.** Building and validating AI systems is resource-heavy. It demands skilled ML engineers, robust computing resources (perhaps GPUs or cloud services), and new tools. Smaller firms may struggle: indeed, a survey found that 53% of pharma companies admitted lack of internal expertise to implement AI (www.pharmoutsourcing.com). Even for large companies, there is a talent bottleneck; one source remarks that *"AI specialists are already over-subscribed"* (www.pharmoutsourcing.com). Industry response includes user-supplier collaboration: FDA hopes firms will increasingly rely on qualified AI vendors, reviewing their development process instead of doing everything in-house (www.hoganlovells.com). Contract manufacturers and CDMOs may also invest in AI capabilities, offering validated AI modules as services. Regardless, best practice is to involve cross-functional teams early, train QA on AI fundamentals (and ML teams on GxP principles), and document every step to avoid miscommunication.
- **Continuous Lifecycle (Model Drift).** The dynamic nature of ML means validation never truly ends. Models may drift subtly as operating conditions or raw material sources change (academic.oup.com). This implies that a one-time validation is insufficient. Instead, companies should integrate **continuous verification**. Some recommend aligning model "refresher" schedules with routine product reviews; others favor alerts (e.g. monthly evaluations against hold-out benchmarks). Clear retraining governance is needed: decision criteria (e.g. accuracy drop >5% or a batch failure event) must trigger formal change-control cycles. Data pipelines should support seamless data re-ingestion and versioning to make retraining auditable. FDA's CSA guidance also encourages leveraging automated and supplier-generated evidence rather than duplicating it manually (www.hoganlovells.com). While fully automated validation is still emerging, mimicking this vision by embedding test suites into the CI/CD flow for AI models is a goal.

To address these challenges, a set of **best practices** has emerged from regulators and practitioners:

- **Risk-Based Prioritization.** Categorize AI functions by impact. High-impact systems (e.g. those controlling sterile fill processes) receive full CSV-style validation. Lower-impact ones can use leaner controls (e.g. exploratory testing, sanity checks). This proportional approach is explicitly advocated by CSA and GAMP AI guidance (www.hoganlovells.com) (www.hoganlovells.com).
- **Robust Documentation.** Extend documentation templates. For example, Validation Plans should explicitly list AI elements (data selection, model metrics, bias checks). Test Protocols should include ML-specific tests (e.g. re-run model on test set, confirm no performance degradation). Some organizations create new documents like a **Model Performance Qualification Protocol**. Whatever the format, the key is traceability: every AI artifact and decision (features used, hyperparameters chosen, model selection rationale) should have a paper trail (www.mastercontrol.com). Digital records are practical here: using tools that automatically log each training run reduces manual notes.

- **Explainability Controls.** Even if the model is opaque, implement technical controls for interpretability. For example, always store feature importances or intermediate layer outputs. Use standardized model formats (ONNX, PMML) that external auditors can load. Where feasible, deploy an AI "onion" – wrap the core model with simpler checks (e.g. rule-based failure detectors). Provide explanation reports for reviewers, and include them in the dossier.
- **Monitoring Tools.** Invest in real-time monitoring and drift-detection tools as part of the system. Modern MLOps platforms embed telemetry: alert if input feature means shift beyond threshold, or if output distribution changes. These serve like the alarms in a processing plant. Regulatory bodies expect such continuous oversight. Some vendors also offer AI governance platforms that automatically track data lineage and performance, easing the burden on QC teams. Using these tools means following the spirit of 21 CFR 11 audit trails, but at the level of data flows.
- **Cross-Functional Training.** Foster interdisciplinary teams. Data scientists must learn GxP fundamentals (11, 820, GMP) while QA staff need basic AI literacy. Workshops, lunch-and-learns, or joint oversight meetings can build mutual understanding. Industry forums (e.g. ISPE AI SIG) are sprouting to share lessons. The complexity of AI demands that no single expert holds all knowledge: collaborative problem-solving is essential.
- **Culture of Validation.** Finally, companies must cultivate a mindset that *even smarter software must be subject to quality tests*. AI should not be seen as a magic bullet that escapes audit—rather, it is a new kind of equipment. Process engineers have long monitored BIMS and SCADA; likewise, AI/ML should be integrated into the Quality Management System with the same seriousness. This includes embracing emerging ideas like validating AI with AI (using automated prompt generation and checking, as suggested by EY (www.ey.com)) and staying agile to new guidance.

Future Directions

The landscape for AI/ML in GxP is evolving rapidly. From a regulatory standpoint, several developments will shape best practices:

- **Formalization of AI Guidelines.** We expect official guidance on AI in pharma within the next few years. The ISPE GAMP AI Guide (July 2025) was a major milestone, and adoption of EU GMP Annex 22 will be the next. Under the EU AI Act, pharma AI tools that are, or are part of, regulated products such as medical devices (for example some patient-monitoring software) can be high-risk, while manufacturing uses need case-by-case assessment (www.ey.com). Consequently, companies should prepare to produce **Technical Documentation** for AI systems, similar to medical device Tech Docs: clear statements of intended use, design rationale, risk plans, etc. Because PIC/S took part in consulting on the draft Annex 22, its adoption should reinforce global harmonization. International bodies like ICH/PIC/S may also issue reflection papers or trainings on AI data integrity.
- **Technological Trends.** Advances in AI will raise new validation questions. For example, **federated learning** (training models across multiple decentralized sites without sharing raw data) may be employed to build more robust models. This approach complicates validation: it will require ensuring

data integrity across sites and secure aggregation. Another frontier is **generative AI** in core processes (beyond documentation): e.g. using GANs for overlay prediction in tablet coatings. If such use emerges, companies will need strategies to qualify inherently creative models. Moreover, as machine learning pipelines become more automated (AI-driven AutoML, continuous deployment, MLOps), validation may shift towards automated test frameworks. Organizations should watch for new standards (like ISO/IEC 42001, published in December 2023 for AI management systems) that may intersect with GxP.

- **Regulatory Innovation.** Regulators themselves are piloting new approaches. Regulators are also discussing how deployed AI models should be monitored over their lifecycle, which may make evidence of ongoing performance monitoring an inspection expectation. The concept of *Computer Software Assurance (CSA)* will likely become normalized: instead of retrospective binder checks, inspectors may expect evidence of ongoing risk management (e.g. demonstrating that high-risk features have robust alarms in place). In the EU, anticipations around Annex 22 indicate that pharmaceutical GMP will soon formally address AI. Companies should stay engaged in public consultations and pilot projects to shape these rules.
- **Integration with DevOps and Automation.** Mature AI validation will blend with modern software practices. One can envision a "GxP DevOps" pipeline where continuous integration tests include both functional and model performance tests. Automated tools might periodically retrain models on new data and re-evaluate against test benchmarks as part of regular product reviews. AI validation may increasingly leverage test generation by AI itself: as noted by EY, *"AI can validate AI,"* automatically generating thousands of test prompts and scoring responses (www.ey.com). Embracing these tools could transform CSV into a mostly automated process.
- **Cultural Shift.** The move towards AI will also drive organizational change. QA departments may need data scientists embedded in them. Regulatory inspectors will need training on AI concepts (some already have started curricula on AI/ML bias, etc.). Vendors will publish more compliant-by-design AI solutions (with built-in logging and audit features). Over time, AI literacy will become as fundamental in QA personnel as knowledge of humidity controls or dissolution testing is today.

In summary, the future of AI in GxP is **increasingly codified and integrated**. Companies that proactively build AI-savvy quality systems will gain a competitive edge. As one expert observes, generative AI and ML "offer dramatic improvements" in operations (www.mastercontrol.com) – but only for those who couple them with **rigorous governance**. With the combined momentum of regulation and industry best practice, AI systems in pharma will converge on a new standard of "AI-aware validation," which preserves the core quality mission while enabling innovation.

Conclusion

Integrating AI/ML into GxP-regulated manufacturing is both a challenge and an opportunity. Our analysis finds that **validating AI/ML is feasible if approached methodically**. The core GxP requirements have not changed: manufacturers must ensure systems are fit for intended use, protect patient safety, and maintain

complete data integrity. What has changed are the *means*. AI introduces adaptive algorithms, big data pipelines, and “black-box” elements. To address this, companies should **extend** existing validation frameworks with AI-specific controls.

This entails treating AI training, data, and outputs as regulated entities: for instance, ALCOA+ governs training datasets just as it does lab measurements (www.auriacompliance.com). It means building a risk-based validation lifecycle (aligned with GAMP 5/AIC, CSA, and QRM) that explicitly covers model metrics, data strategies, and monitoring. It also requires clear documentation: from the outset, write URS/URS that include model performance targets, and maintain rigorous audit trails for all AI operations (www.mastercontrol.com).

Our case studies illustrate that this works in practice. Real examples—from vaccine digital twin pilots to water monitoring proofs of concept—suggest that innovative AI solutions can be validated under GMP as long as their development and operation are controlled like any critical system. They yield tangible benefits: improved first-pass quality, faster deviation resolution, and efficiency gains often measured in dozens of percentage points (www.pharmtech.com) (www.aizon.ai). Moreover, industry reports indicate regulators and leaders are moving forward with clear guidance (e.g. GAMP AI Guide, Annex 22, FDA CSA) that practitioners can align with today.

The key insight is that **AI does not void compliance; it transforms it**. Through disciplined risk management, robust data governance, and ongoing oversight, organizations can harness AI/ML effectively. As one industry author puts it, when generative AI and other tools “offer dramatic improvements” in operations (www.mastercontrol.com), it is incumbent on companies to ensure these tools are governed and validated. The path is laid out by evolving standards and the experiences of early adopters. By following the principles and practices outlined here—backed by regulatory precedent and case evidence—companies can drive innovation while upholding the highest quality and safety standards.

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://www.mckinsey.com/industries/life-sciences/our-insights/generative-ai-in-the-pharmaceutical-industry-moving-from-hype-to-reality
2. https://www.scribd.com/document/850790152/PE-SeptOct23-CompleteIssue-v2-LR-1
3. https://ispe.org/publications/guidance-documents/gamp-guide-artificial-intelligence
4. https://www.hoganlovells.com/en/publications/fda-finalizes-computer-software-assurance-guidance-for-production-and-quality-system-software
5. https://academic.oup.com/bjr/article/96/1150/20220878/7499000
6. https://www.ey.com/en_ch/insights/life-sciences/gxp-and-ai-tools-compliance-validation-and-trust-in-pharma
7. https://www.mastercontrol.com/gxp-lifeline/gxp-compliance-approaches-data-infrastructure-ai/
8. https://www.aizon.ai/blog/gxp-ai-in-action-real-successes-from-pharmas-hidden-gem
9. https://www.pharmtech.com/view/gsk-pilots-digital-twin-for-vaccine-manufacturing
10. https://www.pharmoutsourcing.com/Featured-Articles/614774-Can-GenAI-Address-the-Soaring-Costs-of-Pharma-Medical-Writing/

11. https://www.auriacompliance.com/gmp-blog/data-integrity-and-ai-integration-key-considerations-for-compliance-in-gmp-pharmaceutical-manufacturing
12. https://www.europeanpharmaceuticalreview.com/what-annex-22-spells-for-ai-in-gmp-manufacturing/2135686.article
13. https://eur-lex.europa.eu/eli/reg/2024/1689/oj
14. https://nttdata-solutions.com/es/blog/machine-learning-for-pharmaceuticals-blog-series-part-2/
15. https://quality.eleapsoftware.com/ai-in-the-pharmaceutical-industry-how-artificial-intelligence-is-transforming-quality-management-systems/

THE PUBLISHER

About IntuitionLabs

IntuitionLabs is an AI consulting, custom software development and data engineering firm serving pharmaceutical, biotechnology, medical-device and other life-science organizations. We work with clinical, regulatory, medical-affairs, commercial, quality and IT teams to connect technology decisions with the work people need to accomplish.

AI consulting and adoption

Our AI enablement services cover readiness assessments, use-case selection, governance and policies, team workshops, adoption measurement and ongoing advisory support. We help organizations structure the information layer behind AI: source material, context, permissions and maintained knowledge that make generated answers useful and reviewable. Private LLM inference and hosted AI options support teams evaluating how to operate AI with appropriate control over their data and infrastructure.

Software, data and life-science workflows

IntuitionLabs develops custom software for pharma and biotech, integrates enterprise systems, and builds data engineering and business intelligence solutions. Areas of focus include AI agents, regulatory research, medical writing, medical affairs, CMC information, competitive intelligence and clinical-document workflows. Our eTMF intelligence work includes cross-system reconciliation and inspection-readiness support.

Enterprise platforms and regulated delivery

We provide Veeva services, application support, managed services, integrations and custom applications, alongside enterprise content work involving platforms such as Egnyte. For regulated workflows, our services include GxP enablement, computer-system validation and software development addressing 21 CFR Part 11 requirements. The applicable controls, validation responsibilities and acceptance criteria are defined for each engagement.

Work with IntuitionLabs

Explore AI enablement, pharma and biotech software development, data engineering and BI, and Veeva services. Contact IntuitionLabs to discuss your workflow, information sources and implementation needs.

IntuitionLabs publishes educational research to help life-science teams make informed technology decisions. Coverage of a product or organization does not imply a client relationship, endorsement or partnership.

Website: https://intuitionlabs.ai

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.