Automation-Assisted Pharmaceutical Annual Product Quality Review: An Integrated Framework for Digital Data Systems, AI-Driven Trend Analysis, And Human Oversight in GMP Compliance

Kulshreshth Sharma[1], Gurmeet Chhabra[2], Lalit Dhakar[3], Ankita Bhadoriya[4], Nimita Manocha[5]
[1,2,3,4,5]*Indore Institute of Pharmacy, Near IIM Indore, Rau–Pithampur Road, Indore, Madhya Pradesh*
doi.org/10.64643/IJIRT209453

***Abstract*—Since 1978, the regulatory authorities have mandated pharmaceutical companies to significantly evaluate the quality of their drug product at least once a year, according to the ICH Q7 guideline and the Good Manufacturing Practice (GMP) guideline of the European Union. The approach to performing this Annual Product Quality Review is largely manual, however, with the data being distributed across laboratory, quality, enterprise and manufacturing. The aim of this review is to gather the literature from the period 2020-2026 and show the necessity of a new model that is not independent that will be able to digitalize APQR. It introduces a hybrid data integration architecture model, a 5-layer hierarchical model for automated analytical functions, a layer for anomaly detection by machine learning algorithms at the top, an AI-aided report generation model based on the validated data obtained from the sources and a 3-tier model for human oversight based on the interpretive complexity of the data and the impact of Good Manufacturing Practice (GMP) on the function. The oversight model conforms to the existing validation models including risk-based Computer Software Assurance and Good Automated Manufacturing Practice, as well as the validation requirements for electronic records in 21 CFR Part 11 and European Union Annex 11. Although there are many published reports and manuals, there is no peer-reviewed literature that assesses the artificial-intelligence tools that are actually being used specifically in the APQR realm of work, especially those using NLP to analyze the story of the deviations and corrective actions. Most critically, aside from the time saving, no study could be found that proved that using automation to aid APQR for the purpose of finding quality trends is effective. The five research priorities that are suggested to address this gap are benchmark datasets, comparative outcome studies, validated evaluation methods, model lifecycle governance and better collaboration between industry and regulators.**

***Index Terms*—Annual Product Quality Review; APQR digitalization; automation-assisted quality review; pharmaceutical artificial intelligence; human oversight boundary; Computer Software Assurance; GAMP 5; data integration architecture**

I. INTRODUCTION

Periodic evaluation of drug product quality is a regulatory requirement, not merely a best practice, for all of the key regulatory jurisdictions. The requirement in the United States is in 21 CFR 211.180(e) which was enacted in 1978 (when the CGMP regulations were issued) [1,3]. The expectation is also captured in the ICH Q7 guideline for active pharmaceutical ingredients (APIs) [2], the guideline has no specific mention of the word annual or APQR, but it is a known regulatory practice, as is the EU guideline [4] which has a similar annual review requirement without explicit terms. The following table (Table 1) is a summary of these frameworks. The terms are not important for this review and the need to test periodically and systematically for continued control of a manufacturing process is essentially the same everywhere.

Table 1. Global Regulatory Frameworks Requiring Periodic Product Quality Review

Framework	Jurisdiction	Key requirement
21 CFR 211.180(e) [1]	United States	Annual evaluation of drug product quality data to determine need for changes in specifications, manufacturing, or control procedures

ICH Q7, Section 2.5 [2]	International (APIs)	Defines "product quality review" as a periodic GMP expectation for active pharmaceutical ingredients
EU GMP / PIC/S PI 009-17 [4]	EU / PIC/S members	Yearly product quality review adopted as established regulatory practice

What has not grown in step with the ongoing regulation over half a century have been the processes for the review. In most of the organizations, APQR has been stuck in most of the existing places including LIMS, QMS, ERP, Deviation/CAPA tracker, Stability data base and standalone spreadsheet, which are the first and most stubborn APQR hurdles and where transcription error and inconsistency can be thrown into the mix. More seriously, when manual trend analysis is added to the huge amount of data over a 1-year time frame, identification of the recurring deviations OOS/OOT patterns and process drift is unscalable. These upstream problems are considered as a downstream effect of the review and approval delays and are not the main problems; a review with disjointed information and manual trending is a structural retro review.

The remedy for this is digitalisation, which is now being touted by vendors and in industry publications as an alternative to the spreadsheet and data collection automation processes of compilation. Explicitly answering the question of what happens when automated output and human quality judgment meet and overlap is an afterthought of previous accounts where aggregation is solved, not an initial and design question. Automation of compilation is not enough to address problems of interpretation of regulations, root cause investigation or final batch / process judgment, which remain in US and international GMP with a human Quality Unit.

This review finds that the right model for the APQR should not be fully automatic, but be partly data-driven with integrated data systems, automated statistical trending and signal detection based on AI and pattern recognition, but still not cover the volume and pattern-recognition aspects, in addition to regulatory interpretation, investigations and final approval, which remains in the hands of human oversight. The sections that follow follow this architecture end to end as shown in Figure 1 - from data integration, to automated analytics, and finally to AI-assisted reporting and the human oversight boundary, and the regulatory, validation and data-integrity requirements that this architecture must meet, across jurisdictions.

II. FROM MANUAL TO DIGITAL: THE DATA INTEGRATION LAYER

APQR digitalization starts at the data integration layer, and is least consolidated at this layer. Two models are presented in the literature. With an integration-hub (middleware) approach, each system the LIMS, QMS, ERP, MES and stability stays independent, and the data pertinent to the APQR is extracted, mapped and consolidated via API, ETL pipelines or validated middleware.

A centralized-platform model brings together various quality and manufacturing aspects into a single vendor ecosystem, minimizing the interfaces that need to be validated. On the whole, according to industry experts, there is a compromise between consolidated platform and strong data relationship (single sign-on) and best-of-breed, which means there is no dependency but there is a need for strong APIs and data standards. Table 2 shows the comparison of the two models, but does not rank either model in an abstract sense, and the selection of either model will depend on the already widely dispersed system environment of an organization.

Table 2. Integration Architecture Models Compared

Dimension	Integration-hub (middleware) model	Unified-platform model
System ownership	LIMS / QMS / ERP / MES remain independent	Multiple functions consolidated within one vendor ecosystem
Primary advantage	Fits heterogeneous legacy environments; avoids single-vendor dependency	Unified data model, single sign-on, fewer interfaces to validate
Primary risk	Requires robust APIs and common data standards across systems	Vendor lock-in

In most of the companies, most of the systems are legacy systems and a complete system replacement is not possible so that's why most of the companies have integrated each of the systems as an integration-hub.

The architecture of a Pharma 4.0 would include instruments that would interface with LIMS and data would be transferred between instruments and quality systems, MES and ERP, facilitating automation of the data collection from instruments and eliminating manual transcribing of data [6].
In the ISA-95 model, it is assumed that LIMS and MES are on the same level, and provides data interconnection between shop-floor information and enterprise systems [7].

APQR is not about the data making the crossing of this bridge, but about the data making the crossing of this bridge as if it were a GxP electronic record traceable, with an interpretive context (batch, instrument, operator and/or time and date), with a consistent definition, and passed through validated interfaces. Below is a conceptual flow that is depicted in a graphic in this paper, where the Source systems are linked to an Integration layer, which populates a standardised data repository; and a standardised data repository is linked to Analytics, Report generation and human oversight.

The latter is far from satisfactory, as there are several claims about benefits of integration and few to support one such claim is that the batch review time was reduced from seven months to five days only after implementing an integrated LIMS/MES system [8] a good, but once-off claim from the vendor. Lots of good qualitative literature on the reasons for being integrated, and a quantitative evidence base, even if light, and vendor supported.The second design question is about the time, or when the data that is relevant to APQR is transferred from one system to another, either continuously or periodically. It's not a "nightly", "weekly" or "event triggered" continuum, it's the latter of these three and an annual review is not truly in real time.

The nearer equivalent to real time monitoring in the field of end-to-end testing (E2E) is end-to-end monitoring in real time (RT) as a strategy for Process Analytical Technology (PAT) [9] which is one of the more complex target states moving to near real time monitoring. The position of an organization is on the continuum, it is not a command.

Figure 1. Conceptual flow of an automation-assisted APQR system

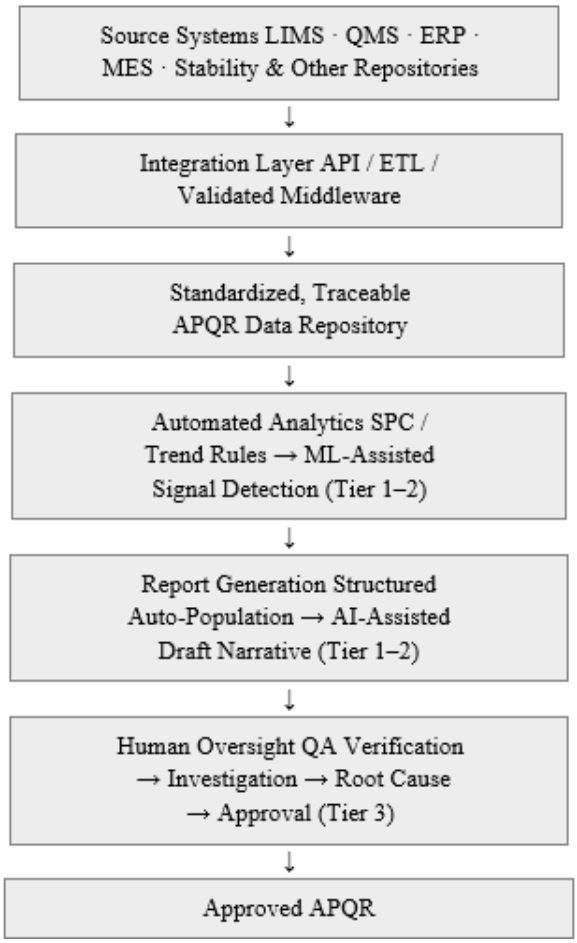


Figure 1. End-to-end conceptual flow from source data to a human-approved APQR, mapped to the Tier 1–3 oversight model developed in Section 5.

III. AUTOMATED ANALYTICS: TRENDING, STATISTICS, AND SIGNAL DETECTION

The Automated Analytics for APQR might be best viewed as a hierarchy as shown in Table 3: Descriptive statistics, SPC/trend rules, automated quality-signal detection, ML-based anomaly detection & prediction, and and always will be human quality interpretation & decision.

The lower layers are not "preliminary" to be avoided as there are more sophisticated methods.

Table 3. The Automated Analytics Hierarchy

Layer	Example techniques	Maturity
Descriptive statistics	Means, ranges, frequency counts	Established
SPC / trend rules	Control charts, Cp/Cpk, Pp/Ppk, spec-limit evaluation	Established, core near-term layer [10,11,12]
Automated quality-signal detection	Rule-based trend/threshold alerts	Established
ML-based anomaly detection / prediction	Random forests, SVMs, autoencoders	Advanced, emerging [13]
Human QA interpretation & decision	Root-cause analysis, regulatory judgment	Always human (Tier 3)

Data obtained from the Continued Process Verification programs are used for the components of the near-term layer, which are routinely trend; calculations of Cpk/Ppk, Capability indices (Cp/Cpk, Pp/Ppk), Specification-limit evaluation, and Trend rules are all explicitly supported by FDA's Process Validation Guidance [11] and are included in the near-term layer. Therefore, SPC is the assumed interest of the automatic; as it can be understood without any explanation and accepted in the pharmaceutical quality system. However, the classic capability indices have some strong distributional assumptions, and may not be suitable for processes that can be monitored more effectively by other techniques such as wavelet-based monitoring or ARIMA decomposition [10].

The new ML (machine learning) layer of anomaly detection is a non-alternative and advanced addition to SPC layer. The advantage of using ML as opposed to statistical methods for detection is that the latter are still interpretable and accepted for regulatory purposes, and the former methods can detect multivariate or nonlinear patterns—random forests and SVMs can detect faulty versus normal states and autoencoders can detect deviations based on reconstruction error [13]. The middle ground is becoming more and more hybrid; both ML determination and/or automatic regulatory conclusions the paper does not consider an ML flagged anomaly as a reduction to an automatic deviation or an OOS determination or a regulatory conclusion.

Numerical and text-based analytics are required to recognize recurring-deviation. Lots of information is stored in deviations, CAPA and complaints, where a lot of the information that is driving a recurring root cause on a structured trending is stored. A completely numerical system would treat the three investigations the same that all reported the same failure with the same keyword as three separate failures, while a semantically aware system would cluster them together as being the same failure [14]. In this particular application, there is less level of peer-reviewed evidence in the literature reviewed, with the majority of these being developed by the vendors and/or industry itself and not independently verified.

IV. AI-ASSISTED REPORT GENERATION AND DECISION SUPPORT

Table 4 shows the two different ways of creating the reports. Structured automation includes the data being validated and entered into pre-programmed tables, charts, batch summaries, deviation/OOS/OOT listings, CAPA status, stability results and/or statistical outputs, etc., which are easier to audit.

Table 4. Report Generation Tiers

Tier	Determinism	Governance requirement
Structured auto-population (tables, charts, listings)	Deterministic	Standard CSV / CSA assurance
AI/LLM-assisted narrative drafting (trend summaries, executive summaries)	Probabilistic	Source grounding, hallucination control, mandatory QA verification [15,16,17]

The first step of governance is to start with the ability to create a narrative with AI/LLM if the output is verified, it's a world entirely different from what a generative model can output that may say something that is not supported by the data. Other sources confirmed the findings of supervised, specific-use

LLM as being helpful for extraction, triage, retrieval and documentation, but not for autonomous LLM decision making, and the need for a human resource element that can be held accountable for the use of LLM in high-stakes scenarios. The purpose of the AI generated APQR narrative is not to replace the APQR, but to be used as a document assist until it is QA'd, interpreted, modified and approved.

The number one antidote to hallucination (output not based on controlled and traceable source data presented with false confidence) is that the model is based on controlled and traceable source data not free-ranging thought. The risk of hallucinations is reduced as the outputs are grounded in information retrieved from validated sources (SOPs, investigations, CAPA databases) by searching with the vector database, and at the same time, a decision pathway can be presented to a reviewer allowing a review of the decisions, and not simply relying on a black box this is emerging practice and not established for APQR. The source attribution, reproducibility, version control, audit trails and access control are conditions of application of this tier.

The decision support and leverage from the process variability and process variability recurrence and the frequency of OOS/OOT, recurrence of CAPA, number of complaints, attribute criticality and past APQR findings, are also considered in a composite decision support of the need for QA attention to build on the risk-based prioritization. This is a direct implementation of the risk ranking and filtering tool of ICH Q9 (see reference [18]) which is applicable to portfolios whose comparability can be ensured by using more than one comparison tool. The purpose of an AI supported system is to continually update the portfolio (not a priority decision).

V. THE HUMAN OVERSIGHT BOUNDARY

The Inwards journey towards a new design question is not a yes/no question but instead a 'where' and 'how' question: Where and how should there be a human in APQR? The term human-in-the-loop is too ambiguous because of the results of the human factors research studies—even expert humans will rely more on what is presented by the automated system and will reduce the amount of independent work they will do with the information, known as automation bias [19]. What an APQR system without a man-in-the-loop could do with is a reviewer that takes his/her name to a conclusion that has not been read before.

It is also evident that a 'automation versus humans' approach is not appropriate and that the three tier model (table 5) that scales the essential human involvement by the interpretive complexity, ambiguity and impact of GMP of a function is more appropriate. Tier 1 (deterministic automation) will not need to re-calculate the output of the routine line by line, but will need to handle exceptions, and understand the performance of the system.

Tier 2 (AI-assisted decision support) is slightly more abstract, and needs to be verified and validated with the person's product, where there can be automation bias. Tier 3 (human accountable decisions) is not delegateable: decisions, always with qualified human, can not be delegated, regardless of Tier 2 output was taken into consideration or not.

Table 5. Three-Tier Human Oversight Model

Tier	Function / examples	Human role
Tier 1 Deterministic Automation	Rule-based, validated processing: data extraction, calculations, table population, predefined SPC charts, report formatting	Routine output may not require item-by-item manual recalculation, but exceptions and system performance remain governed
Tier 2 AI-Assisted Decision Support	Interpretation, detection, prediction, or prioritization: anomaly detection, NLP clustering, risk scoring, trend summaries, AI-generated narratives	Mandatory QA verification and contextual interpretation
Tier 3 Human-Accountable Decisions	Investigation conclusions, root-cause determination, CAPA adequacy, assessment of significant trends, final APQR conclusions and approval	Qualified human retains decision authority and accountability

While the GMP of the responsible organisation is maintained unchanged, regarding the AI/ML algorithm that it relies on, and in general, in the context in which Human Oversight should be maintained, that is, the more it gets data-driven, the more it gets away from the rules and therefore the more it becomes autonomous [21]. Tier 2: If a reviewer signs off on an AI-generated conclusion, and doesn't read it, then it is not to detract from his/her review responsibility. It is not so much the domain as the mechanism of oversight, which is not well-defined in AI governance research other than in the pharma world; it is not only "approval" but "active oversight" and is able to overrule.

The three scenarios are: 1) An AI-powered trend engine that won't pick up on an important trend; 2) A risk model that won't identify an actual risk; 3) The generative model that will create a fluent and plausible story, which is rubber-stamped by the reviewer because AI-generated text also sounds good. The term "special" is not included in APQR because they are examples of a pattern of automated-decision support and in this case, specific regulated consequences.

VI. REGULATORY, VALIDATION, AND GOVERNANCE CONSIDERATIONS

Firstly, an Automation-assisted APQR system is a collection of computer-based systems that are used in GMP, and has many of the same attributes as CSV: intended use, risk assessment, specification, verification, traceability, change control, access control, audit trails, electronic record integrity and maintaining the validated state. The risk-based structure from which this was extracted from was provided in GAMP 5.

The most explicit recent guidance of the logic has been FDA's Computer Software Assurance (CSA) guidance from September 2025 that recommends a risk-based approach which focuses on documenting specific elements of the software lifecycle that present a risk to the software, such as the process, and maps software to intended use and software process-risk classification [22, 23] and uses vendor assessments and digital evidence as needed. This review is not meant to replace, but is meant to clearly express a principle that should be considered throughout the document: the level of assurance expected of the effort to implement GMP should reflect the impact of the implemented system upon the product, process control and/or QA when compared to the manual system that is to be replaced. This review should not be seen as a substitute for the EU GMP embedding of the same principle which took place when the computerised system was adopted rather than the manual system in 2011, without a reduction in the quality of the product, the control of the product or QA.

EU GMP Annex 11 record integrity and completeness of audit trail, and access control and unreputable linkage of signatures and 21 CFR Part 11 cover the dimension aspect of electronic-record and electronic signature.

Table 6. Validation and Assurance Requirements by Tier

Tier	Validation / assurance requirement
Tier 1 Deterministic Automation	Conventional CSV or risk-based CSA assurance proportional to process-risk classification [22,23]
Tier 2 AI-Assisted Decision Support	Tier 1 requirements plus model performance evaluation, explainability, data provenance, failure-mode characterization, and drift monitoring [24]
Tier 3 Human-Accountable Decisions	No software validation substitutes for qualified-person judgment; decisions are informed by validated Tier 1 and assessed Tier 2 outputs

AI/ML model lifecycle management is another governance category to CSV, as its is not a code change but after the deployment of the AI/ML model, the statistical relationship between the input and output could change.

Relevant training data should be documented and versioned before the training data is deployed, training data drift should be monitored before reusing the training data and/or impact assessment should be conducted when revalidating the training data (if applicable) before the data drift is too much or there are major changes in the training data as defined by the trigger conditions.

This shouldn't be dismissed as preparatory to deployment, as it is, in fact, an assessment of the bias, characterization of false positive and false negative outcomes, explainability and training governance.

VII. DISCUSSION AND FUTURE OUTLOOK

There is a reasonable grasp of data integration but often it's still heterogeneous in nature. Evidence of automated analytics in the context of SPC is well-established and in the APQR narrative text, the evidence is promising, but still nascent.Light evidence for automated analysis of APQR narrative text; A well-established core for SPC, and an advanced body for ML, more recent. In principle, the use of AI in the drafting of reports is good, but there is no published evidence with specific focus on APQR writing. There are no studies that compare the effectiveness of the three-tiered system of oversight with the one system of human in the loop (human in the middle) on the results of APQR.

But this uncertainty is not something this paper is looking to clear up as the regulatory environment that it depends on is still constantly evolving, with Annex 11 mid-revision and only CSA was finalised a year ago. Any review written within this time frame is essentially about a target architecture and not a proven one.

There are a few 'directions' that should be listed without exaggeration, for developments in the near future. The progression of the integration-maturity continuum towards the far right is called continuous quality intelligence. It is possible for report generation and decision support to go in a more focused direction as a future evolution of the questions of oversight, becoming multimodal and agentic coordinating retrieval and analysis from multiple quality systems. The use of digital twins and predictive quality models for predictive maintenance and process optimization has been demonstrated to be successful, but data integration, model accuracy and regulatory complexity are challenges [26]. Continued Process Verification will further drive the trend in Section 3 towards closer alignment. These developments signal a future state scenario of first - retrospective APQR, then - automated APQR periodically, then - an updated quality data environment, then - quality intelligence predictive, and finally - pro-active quality management, all guided by a human. However, it's a concept different from the related but separately regulated real-time release testing (RTRT) that uses real-time process data to release the batch as opposed to APQR automation's endpoint.

The field should have a plan for specific evidence, and not further research as represented in Table 7.

Table 7. Research Agenda Priorities

Priority	Focus
1	APQR-specific benchmark datasets (representative structured data paired with appropriately governed unstructured data)
2	Comparative outcome studies across manual, rule-based, SPC-based, and AI-assisted APQR workflows using measurable outcomes (preparation time, error rate, signal-detection sensitivity, reviewer workload)
3	Validated evaluation methods for NLP/LLM summaries, anomaly detection, and risk-prioritization systems specific to the APQR context
4	Lifecycle governance frameworks for model drift, change control, retraining, explainability, and human oversight developed specifically for GxP environments
5	Stronger industry–academia–technology-provider–regulator collaboration to develop interoperable standards and evidence-based regulatory expectations

Most important, however, the field has to have proof that using AI to assist with APQR is quicker, and that the added benefit of using AI means that they are able to find meaningful trends that wouldn't have been identified without the help of AI, with less false-negatives, less false-positives and no automation bias: This is a hard, but important, and still unanswered question.

VIII. CONCLUSION

The literature reviewed shows that the automation of an APQR is most well developed when the majority of the tractable: deterministic Tier 1 functions exist. The data extraction and building on this is structured in SPC charts and auto-population of the report is based on a precedent of 50 years of experience, starting with

the first codification of ICH Q7 guideline 21 CFR 211.180(e) in 1978 and publishing of ICH Q7 in 2000. One example of this is the integration of a LIMS/MES system that shaved 7 months from review time to 5 days; that's an outer bound and not necessarily an average number of how much value might be gained from one organization.

However, one of the biggest uncertainties is the increasing disparity between the message being delivered about the capabilities of automation-supported APQR and the proven benefits. There is an abundance of literature to support the efficiency claims, but little in the peer reviewed literature that supports any capability that will increase the likelihood of detecting meaningful quality trends. This disconnect is further compounded when Tier 2 functions are developed, as automation bias leads to faster and more confident looking results, which are most at risk for independent assessment.

The closure should be tackled on a case-by-case basis for each stakeholder. Technology providers supplying the technology to use a model should make sure that the outputs of the model are built on retrieved and valid source data and that the technology is accompanied by monitoring of drift and revalidation triggers in the model at the time of use in production. It is important that researchers concentrate on the outcome studies and benchmark data sets identified in this review that are missing, rather than keep generating new ‘isolated case studies. The principles behind the risk-based approach that is already in use for the CSA and in the revision of Annex 11 should be explicitly applied to the requirements for the validation of AI systems, so that organisations don't have to fill in the details by inference.

The comments made here seem well-intentioned and well-spoken, but it is not criticism or rejection. The direction of automation-assisted APQR, based on the three-tiered oversight model proposed here, is coherent and regulator-friendly with a reasonably established deterministic layer and a promising yet thin advanced layer, evidenced. The answer is that the architecture of defensible automation-assisted APQR is mostly in place it's not fantastic, but it's not terrible, either it's just that it lacks the confidence to claim it is producing higher quality oversight, not speedy reporting.

REFERENCES

[1] Code of Federal Regulations, “21 CFR Part 211.180—General requirements.” U.S. Food and Drug Administration. [Online]. Available: https://www.ecfr.gov/current/title-21/chapter-I/subchapter-C/part-211/subpart-J/section-211.180

[2] International Council for Harmonisation, Q7 Good Manufacturing Practice Guidance for Active Pharmaceutical Ingredients. Geneva, Switzerland: ICH, 2000. [Online]. Available: https://database.ich.org/sites/default/files/Q7%20Guideline.pdf

[3] Pharmaceutical Technology, “Product Annual/Quality Review: US-EU Comparative Analysis and Interpretations.” [Online]. Available: https://www.pharmtech.com/view/product-annualquality-review-us-eu-comparative-analysis-and-interpretations

[4] ValGenesis, “Optimizing Your Annual Product Quality Reviews (APQRs).” [Online]. Available: https://www.valgenesis.com/blog/optimizing-your-annual-product-quality-reviews-apqrs

[5] IntuitionLabs, “GxP Compliance Software: eQMS & Document Management.” [Online]. Available: https://intuitionlabs.ai/articles/gxp-compliance-software-eqms-comparison

[6] Technology Networks, “Pharma 4.0: Digital Integration, LIMS, and AI in the Lab.” [Online]. Available: https://www.technologynetworks.com/applied-sciences/articles/pharma-40-digital-integration-lims-and-ai-in-the-lab-413557

[7] Infosys, “Why MES-LIMS Integration Is a Game-Changer.” [Online]. Available: https://www.infosys.com/engineering-services/insights/documents/unlocking-hidden-value.pdf

[8] IntuitionLabs, “Quality 4.0 in Pharma: A 2026 ROI & Economic Analysis.” [Online]. Available: https://intuitionlabs.ai/articles/quality-4-0-pharma-roi-analysis

[9] “Challenges and Opportunities of Implementing Data Fusion in Process Analytical Technology—A Review,” PubMed Central. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9369811/

[10] "A Review of Artificial Intelligence Impacting Statistical Process Monitoring and Future Directions," arXiv:2503.01858. [Online]. Available: https://arxiv.org/pdf/2503.01858

[11] International Society for Pharmaceutical Engineering (ISPE), "The Role of Process Capability in Monitoring Product Quality." [Online]. Available: https://ispe.org/sites/default/files/concept-papers/role-process-capability-monitoring-quality.pdf

[12] Sigma-Aldrich, "Continued Process Verification to Improve Product Quality, Strengthen Process Control, and Meet Regulatory Authority Expectations." [Online]. Available: https://www.sigmaaldrich.com/US/en/technical-documents/technical-article/pharmaceutical-and-biopharmaceutical-manufacturing/monoclonal-antibody-manufacturing/continued-process-verification

[13] "Applying Machine Learning Models for Real-Time Process Monitoring and Anomaly Detection in Pharma Manufacturing," GSC Biological and Pharmaceutical Sciences, vol. 27, no. 1, pp. 315–341, 2024. [Online]. Available: https://gsconlinepress.com/journals/gscbps/sites/default/files/GSCBPS-2024-0153.pdf

[14] Entefy, "A Multi-Model AI Framework for a More Robust Deviation Management in Pharma Manufacturing." [Online]. Available: https://www.entefy.com/blog/a-multi-model-ai-framework-for-a-more-robust-deviation-management-in-pharma-manufacturing/

[15] "Large Language Models in Adverse Drug Reaction Detection and Pharmacovigilance: A Systematic Review of Current Applications, Challenges, and Future Directions," PubMed Central. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC13465311/

[16] ISPE Pharmaceutical Engineering, "Transforming Pharmaceutical Quality with GAMP AI Guardrails." [Online]. Available: https://ispe.org/pharmaceutical-engineering/july-august-2026/transforming-pharmaceutical-quality-gampr-ai-guardrails

[17] "The Imperative for Regulatory Oversight of Large Language Models (or Generative AI) in Healthcare," npj Digital Medicine, 2023. [Online]. Available: https://www.nature.com/articles/s41746-023-00873-0

[18] European Medicines Agency, ICH Guideline Q9 on Quality Risk Management, EMA/CHMP/ICH/24235/2006. [Online]. Available: https://www.ema.europa.eu/en/documents/scientific-guideline/international-conference-harmonisation-technical-requirements-registration-pharmaceuticals-human-use-ich-guideline-q9-quality-risk-management-step-5-first-version_en.pdf

[19] "The Flaws of Policies Requiring Human Oversight of Government Algorithms," arXiv:2109.05067. [Online]. Available: https://arxiv.org/pdf/2109.05067

[20] "Check the Box! How to Deal with Automation Bias in AI-Based Personnel Selection," PubMed Central. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10113449/

[21] "Harnessing the AI/ML in Drug and Biological Products Discovery and Development: The Regulatory Perspective," PubMed Central. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11769376/

[22] U.S. Food and Drug Administration, Computer Software Assurance for Production and Quality System Software, final guidance, Sep. 2025. [Online]. Available: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/computer-software-assurance-production-and-quality-system-software

[23] Lexology, "FDA Finalizes Computer Software Assurance Guidance for Production and Quality System Software." [Online]. Available: https://www.lexology.com/library/detail.aspx?g=f885ead2-5969-4749-a339-6ebeda313551

[24] IntuitionLabs, "ISPE GAMP AI Guide: Validation Framework for GxP Systems." [Online]. Available: https://intuitionlabs.ai/articles/ispe-gamp-ai-validation-guide-gxp

[25] valkit.ai, "EU Annex 11 vs 21 CFR Part 11 Explained." [Online]. Available: https://valkit.ai/blog/annex-11-eu-gmp

[26] "Transformative Roles of Digital Twins from Drug Discovery to Continuous Manufacturing: Pharmaceutical and Biopharmaceutical Perspectives," PubMed Central. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12516570/