

Large language models as the cognitive layer of next generation digital twins in pharmaceutical manufacturing and packaging



Tanvi D. Patel, Meenal M. Rane*

Department of Pharmaceutics, SVKM's Dr. Bhanuben Nanavati College of Pharmacy
(Autonomous), Vile Parle West, Mumbai – 400056, Maharashtra, India.

Email: meenalrane2@gmail.com; meenal.rane@bncp.ac.in

Abstract

Digital Twins (DTs) are a virtual representation of physical processes, equipment, and products that can transform pharmaceutical manufacturing and packaging by enabling sustainable manufacturing (*in-silico* experimentation reducing material waste), real-time monitoring, predictive maintenance, and Quality by Design (QbD) optimization under the Pharma 4.0™ operating model. However, conventional DTs remain largely inaccessible to plant personnel, demanding specialist expertise to interpret outputs from numerical dashboards and physics-based models. Large language models (LLMs), built in the transformer architecture, offer a promising cognitive layer that enables operators to interact with DTs through natural language, supporting root-cause analysis, standard operating procedures (SOP) retrieval, and automated documentation. This perspective synthesizes established DT practice with emerging LLM-DT integration concepts from Industry 4.0 literature, while explicitly distinguishing validated evidence from proposed future directions. Key challenges include hallucination risk, Good Manufacturing Practice (GMP) validation, and regulatory acceptance which are critically examined alongside a roadmap for responsible adoption.

Keywords: Digital Twin; Large Language Model; Pharma 4.0™; Pharmaceutical Manufacturing; Pharmaceutical Packaging; Sustainable Manufacturing; Good Manufacturing Practice; Artificial Intelligence; Process Analytical Technology

1. Introduction

The progressive adoption of the Pharma 4.0™ operational model in the pharmaceutical industry, which connects the principles of Industry 4.0 digitalization with the quality system defined by ICH Q10, has brought about the need for advanced information and decision support systems(1,2). Within the framework of this model, digital twins, which are defined as virtual replicas of real-life processes continuously updated with the help of sensors and information from production processes, have proven to be enabling technologies for continuous process monitoring, in-silico experiments, and predictive quality assurance in the pharmaceutical industry.

The formal introduction of the digital twin concept within the systems engineering framework can be attributed to Grieves and Vickers (2017)(3), while the application of digital twins in the pharmaceutical sector has its roots in the FDA Process Analytical Technology (PAT) regulatory framework established in 2004(4). This framework proposes using real-time information about production processes, which is supported by ICH Q8(R2), Q9, and Q10 guidelines that define the Quality by Design (QbD) concept used specifically by digital twins to improve pharmaceutical development and production processes.

Traditional pharmaceutical DT implementations depend mainly on physically based models or classical ML techniques, where operators interact using numerical dashboards and graphical interfaces, instead of using natural language communication(5). LLMs, which are based on the transformer model introduced by Vaswani et al., have shown some good ideas for reasoning, retrieving information, and creating meaningful texts(6). Using LLMs as a cognitive interface in conjunction with DTs in the pharmaceutical sector provides a valid opportunity to bring DT information to shop personnel, thus speeding up corrective actions, decreasing the frequency of deviations, and fulfilling the requirements for resource-effective Pharma 4.0 production.

2. Digital Twins in Pharmaceutical Manufacturing: A Sustainable Technology enabler

According to manufacturers, pharmaceutical digital twins (DTs) are virtual models of processes that combine sensor and analytical instrument data with real-time data streams to model, predict, and improve operations before taking any action. Using this approach supports sustainable manufacturing because in-silico methods are applied, which helps avoid wasted raw materials and solvents that are associated with standard R&D processes. Examples include real-time monitoring of critical quality attributes (CQAs) by PAT devices and various applications of QbD methods in manufacturing(5,7).

Moreno-Benito et al. (Pfizer, 2022)(7) have presented the largest collection of information about how to utilize the pharmaceutical DT approach. The authors developed digital twins of the continuous direct compression (CDC) line that included a combination of multiple mechanistic and data-driven models. Such digital twins were successfully used for the modeling of CQAs such as tablet hardness, dissolution, and blend uniformity allowing for exploratory design of the process leading to significant reduction in the quantity of necessary raw materials when conducting real experiments. Copot et al. (2026)(8) also presented an additional CDC simulations system.

In the manufacturing of biopharmaceuticals, digital twins enabled by the application of process analytical technology (PAT) in the production of drugs are used to manage various critical process parameters (CPP). Furthermore, this enables the production of reliable data that can be utilized by machine learning algorithms in the future. Among the implementations of digital twins in industry are examples such as the one of Johnson & Johnson that collaborated with Siemens in coming up with the system. This led it to reduce its production time and use of solvents by around 30% along with overall costs by about 35%. Eli Lilly has been known to also use digital twins to assess scale-up for glucagon-like peptide-1 (GLP-1) receptor agonists Zepbound® and Mounjaro® before embarking on actual manufacturing. The effects of the case studies have not been confirmed.

3. Digital Twins in Pharmaceutical Packaging

The domain of packaging-specific digital twins (DTs) is relatively newly established and lacks the extensive peer-reviewed literature seen in process manufacturing. Secondary packaging DTs make use of data from machine vision cameras, heat-seal monitors, serialization printers, and cartoning systems to create a live virtual model of the packaging line. The only established case is in predictive maintenance, wherein DTs use vibration effects of filling and cartoning machinery along with the seal-bar temperature to detect the build-up of mechanical failures before any significant rejection issues or unforeseen interruptions arise(9). This ability does not only allow DTs to perform maintenance but directly contributes towards lowering materials and packaging waste.

Remote access to optical systems by means of DTs enables virtual commissioning and process optimization when developing new defect detection solutions. Moreover, the issue of serialization, which is now required under the EU Falsified Medicines Directive (FMD 2011/62/EU)(10) and the US Drug Supply Chain Security Act (DSCSA)(11), becomes an

increasingly important application: packaging DTs can recreate the flow of serialization data, check aggregation logic, and detect integration issues before implementation on actual lines.

Cold-chain packaging DTs use temperature-excursion sensor data with existing stability modelling concepts to make instant predictions about shelf-life effects which is the necessary data infrastructure for continuous real-time release testing (RTRt) programmes. This field offers a significant opportunity for using LLM for providing data in situations where errors in packaging, mismatches with serialization or cold-chain accidents require fast inquiries based on the context of the non-expert operators acting in the absence of highly-skilled professionals.

4. Large Language Models as the Cognitive Layer

LLMs based on the transformer architecture(6) are trained using large and diverse sources of information including scientific information, regulatory guidelines, and operational documentation. LLMs possess a good capacity for carrying out multi-step reasoning, retrieving domain-specific knowledge, answering questions, and creating structured documents(12). In the context of its application to pharmaceutical digital twins, the cognitive layer can perform the following four functions: (i) provide a language- based interface for communicating the DT data in real time, (ii) prepare hypotheses of the root cause by checking the sensor data against the batch records and corrective and preventive action databases, (iii) prepare primary version of GMP deviation reports and SOPs, and (iv) provide real-time decision-making support taking into account the applicable regulations (Figure 1).

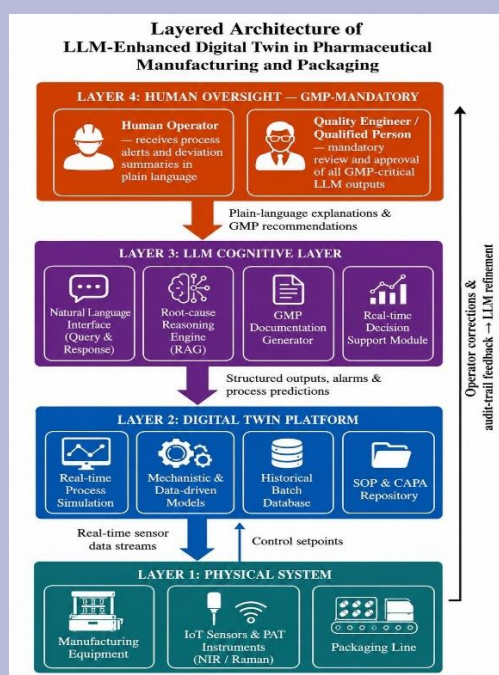


Figure 1. Layered architecture of an LLM-enhanced pharmaceutical Digital Twin (DT).

[The physical system layer (manufacturing equipment, IoT sensors, Process Analytical Technology (PAT) instruments, and packaging line) feeds real-time data streams into the Digital Twin Platform, which executes process simulations and maintains historical batch records, standard operating procedure (SOP) repositories, and corrective and preventive action (CAPA) databases. The LLM Cognitive Layer provides a natural-language interface, root-cause reasoning via Retrieval-Augmented Generation (RAG), GMP documentation generation, and real-time decision support. The Human Oversight layer (human operator and qualified person) receives plain-language outputs from the LLM and retains mandatory approval authority for all GMP-critical decisions. A feedback loop enables continuous LLM refinement from operator corrections and audit-trail data. NIR, near-infrared spectroscopy; RAG, Retrieval-Augmented Generation; GMP, Good Manufacturing Practices.]

In 2025, Yang et al. published a study presenting the DPP framework for LLMs in DT modeling (13). The framework consists of three stages of contributions from LLMs to DT workflows — descriptive, predictive, and prescriptive — where descriptive entails the processing of inputs from sensors describing both the state of the plant and the data provided by DT; predictive stage is based on LLM reasoning that assists in providing information on future trajectory of processes; and prescriptive stage involves issuing recommendations on corrective actions relative to the current state of the manufacturing process. Later studies in the pharmaceutical area focused on the research conducted by Salami et al. allowing comparison of the performance of three LLMs including GPT-3.5, GPT-4, and LLaMA offering a well-balanced dataset consisting of data on actual deviations during manufacturing (14). The outcome of the study revealed that LLMs may offer insights regarding the underlying issues prompting companies to use that information indicating at the importance of fine-tuning of such models according to particular area of application as well as employing the expertise of professionals during the implementation process.

The engineering evidence discussed by Gill et al. (2025)(15), presented at IEEE ETFA, gives proof for LLM agents working in a DT environment and launching autonomous fault-management actions with the DT giving the true state of the system which can be used for confirming LLM decisions in real time. Papacharalampopoulos et al. (2026)(16) improved the presented architecture by creating a fault-resistant process and installing a cognitive supervisory control level where an LLM reasoning agent processes DT telemetry data and suggests corrective actions. Thus, the combined results of the described works show the technical approaches to the practical implementation of LLM-DT connectivity. Han and Tao

(2024)(17) describe the manufacturing DT application in a broad context of the application of AI and LLM technologies in the pharmaceutical business.

Table 1. Comparison of conventional AI-driven DT platforms and LLM-enhanced DTs across key operational dimensions.

Feature	Conventional AI-driven DT	LLM-enhanced DT
Operator Interface	Numerical dashboards; graphical process charts	Natural-language Q&A conversational interface
Root-cause Analysis	Engineers manually interpret alarms and multivariate data	LLM synthesizes sensor logs, batch history, SOPs, and CAPA records automatically
SOP Access	Manual document search by operator	LLM retrieves and contextualizes relevant SOP to current DT status
Deviation Report Drafting	Manual GMP writing by qualified person	LLM generates first draft; qualified person reviews and approves
Operator Training Demand	High — requires dashboard and process model expertise	Reduced — conversational interface lowers the specialist skill barrier
Primary Limitation	Cognitive burden on operator; siloed data views	Hallucination risk; probabilistic outputs require rigorous GMP validation

5. Potential Applications across the Pharmaceutical Value Chain

Extrapolating the demonstrated capabilities to the full pharmaceutical value chain yields several technically plausible, though not yet fully validated, applications:

Manufacturing decision support: LLM-generated root-cause hypotheses for process deviations, synthesizing real-time sensor trends with historical batch records and CAPA

databases, could reduce deviation investigation cycle times from days to hours while maintaining full audit-trail traceability.

Packaging troubleshooting: Natural-language queries to diagnose line stoppages, serialization mismatches, or seal integrity failures using DT-linked packaging line data, enabling rapid operator-level triage without specialist engineering intervention.

Quality and deviation investigation: Automated generation of first-draft CAPA documentation and deviation reports from DT data records, subject to mandatory qualified-person review and approval prior to GMP-release.

Regulatory support: LLM-assisted compilation of process validation summaries aligned with ICH Q10 change-control expectations, accelerating the documentation burden associated with post-approval changes (Table 2).

Table 2. Proposed LLM-enhanced DT applications across the pharmaceutical value chain and their current maturity status.

Domain	Proposed LLM-DT Application	Maturity Status
Manufacturing	Root-cause analysis via natural language query; batch deviation hypothesis generation with cross-referencing of real-time sensor trends and CAPA databases	Conceptual / early preprint stage
Packaging	Line-stoppage and serialization troubleshooting; seal-integrity failure diagnosis; cold-chain excursion alerts with shelf-life impact assessment	Conceptual — not yet validated in pharmaceutical packaging
Quality (GMP)	Automated first-draft CAPA documentation and deviation reports, subject to mandatory qualified-person review and approval	Early research — Merck & Co. empirical demonstration (Salami et al., 2025)(14)
Regulatory Affairs	LLM-assisted process validation summaries and ICH Q10 change-control documentation; gap analysis against FDA	No formal FDA/EMA guidance for LLM-

	2025 Draft Guidance (FDA-2024-D-4689)	specific use; emerging regulatory attention
--	---------------------------------------	---

6. Challenges and Regulatory Considerations

LLM-DT integration faces substantive technical, regulatory, and ethical barriers that necessitate systematic resolution before deployment in GMP-regulated pharmaceutical environments.

Hallucination and Factual Accuracy: LLMs are susceptible to generating syntactically plausible but factually incorrect outputs — a critical liability in GMP-regulated environments where decisions must be defensible and auditable. Retrieval-Augmented Generation (RAG) architectures, which ground LLM outputs in verified document retrieval rather than parametric memory, attenuate hallucination risk but do not eliminate it. All LLM-generated outputs must therefore be subject to qualified-person verification before any operational action is taken.

GMP Validation and Electronic Records Compliance: All software components within pharmaceutical manufacturing environments must satisfy applicable validation requirements. FDA 21 CFR Part 11(18) mandates controls for electronic records and signatures; EU GMP Annex 11(19) specifies equivalent requirements for computerized systems under EudraLex. LLM systems, as probabilistic and non-deterministic computational components, present unprecedented challenges for traditional Computer System Validation (CSV) frameworks, which were designed for deterministic, rule-based software.

Data Privacy and Cybersecurity: LLM-DT platforms process proprietary batch records, formulation parameters, and equipment-specific operational data. Deploying external API-based LLM services introduces data confidentiality risks incompatible with standard pharmaceutical data governance protocols. Private on-premises or air-gapped LLM deployments represent likely prerequisites for GMP-compliant operation.

Sustainability Trade-Offs: Quantified lifecycle analyses comparing the compute energy cost of LLM deployment against measurable reductions in solvent consumption, batch failures, and packaging scrap are currently absent from the peer-reviewed literature. This represents a substantive evidence gap that must be addressed before net-sustainability claims for LLM-DT integration can be rigorously validated.

Regulatory Acceptance and Human Oversight: Both the FDA and EMA are actively developing governance frameworks for AI and ML in pharmaceutical manufacturing. The FDA's 2023 Discussion Paper on Artificial Intelligence in Drug Manufacturing (Docket No. FDA-2023-N-0487)(20) and its 2025 Draft Guidance on the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products (Docket No. FDA-2024-D-4689)(21) signal accelerating regulatory attention to AI in pharmaceutical operations. LLM outputs must function exclusively as decision-support tools, with qualified personnel retaining unambiguous approval authority for all GMP-critical actions. A phased implementation and validation pathway structured around these regulatory anchors is illustrated in Figure 2.

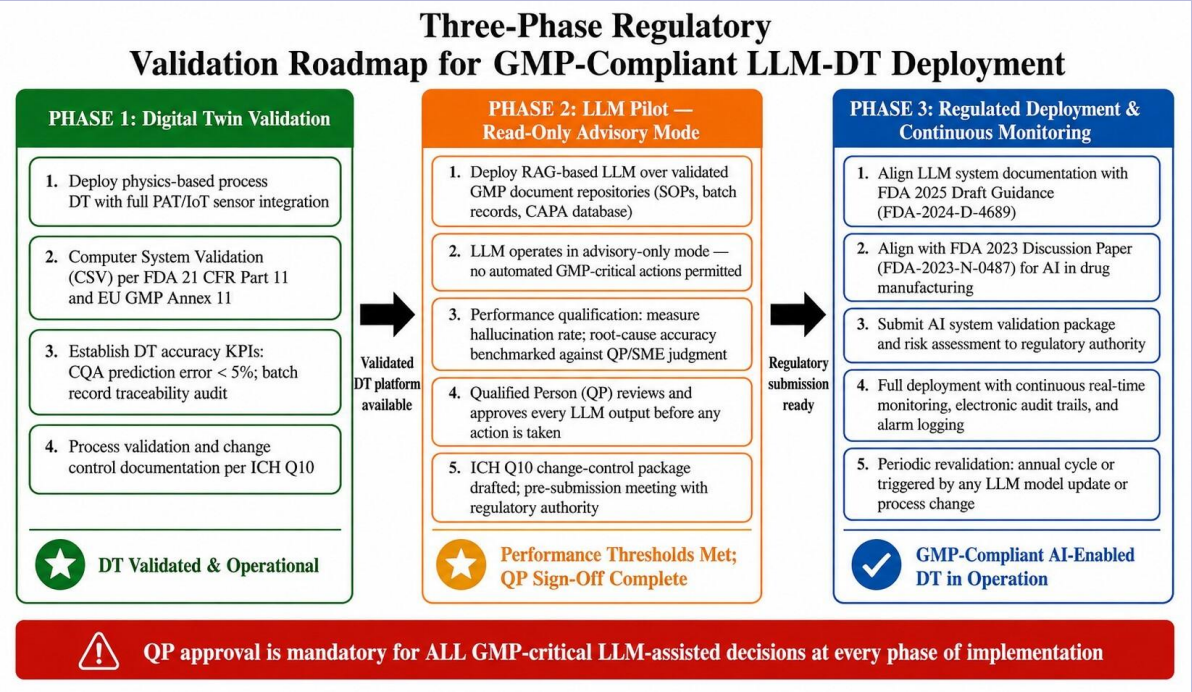


Figure 2. Three-phase regulatory validation roadmap for GMP-compliant LLM-enhanced Digital Twin (DT) deployment in pharmaceutical manufacturing.

[Phase 1 establishes the validated DT foundation through Computer System Validation (CSV) aligned with FDA 21 CFR Part 11 and EU GMP Annex 11. Phase 2 integrates the LLM in read-only advisory mode over a Retrieval-Augmented Generation (RAG) architecture, with performance qualification benchmarked against qualified person (QP) judgment and ICH Q10 change-control documentation submitted to the regulatory authority. Phase 3 delivers GMP-compliant full deployment with continuous real-time monitoring, electronic audit trails, and periodic revalidation, aligned with the FDA 2023 Discussion Paper on AI in Drug Manufacturing (Docket No. FDA-2023-N-0487) and FDA 2025 Draft Guidance (Docket No.

FDA-2024-D-4689). QP approval for all GMP-critical LLM-assisted decisions is mandatory at every phase. CSV, Computer System Validation; RAG, Retrieval-Augmented Generation; QP, qualified person; ICH, International Council for Harmonization; GMP, Good Manufacturing Practices.]

7. Future Outlook

The establishment of validated LLM models for the pharmaceutical industry, which includes the use of RAG across GMP doc archives, regulatory documents, and manufacturing history records, is the next step in the advancement of LLM-DT. In terms of sustainability, LLM-DT allows for recognizing quality issues and recommending corrective measures to assure resource-efficient Pharma 4.0 processes. However, implementation of this idea requires simultaneous developments in the fields of explainability frameworks, regulatory science, and energy-efficient LLM architectures, which are still under study now.

8. Conclusion

The use of digital twins shows visible results in pharmaceutical production and packaging through continual, live tracking of production processes, prediction of machinery failures and detecting problems at every level of CQA, while simultaneously bringing evident benefits in terms of sustainability including reduced material and solvent consumption as well as packaging waste. The introduction of LLMs as a basis for natural-language reasoning in DT platforms is a new trend which is based on certain scientific evidence and the outcomes of the preliminary study of pharmaceutical failures and successes in error rectifying methods in various industries. There are several issues that need to be addressed, particularly in respect of hallucination control, validation of GMP compliance in results of probabilistic software and evaluation of overall sustainability impact.

9. References

1. International Council for Harmonisation. ICH Q10 Pharmaceutical Quality System. Geneva; 2008.
2. Pharma 4.0 Operating Model | Industry 4.0 | ISPE | International Society for Pharmaceutical Engineering [Internet]. [cited 2026 Jul 21]. Available from: <https://ispe.org/initiatives/pharma-4.0>
3. Grieves M, Vickers J. Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. In: Transdisciplinary Perspectives on Complex Systems. Cham: Springer International Publishing; 2017. p. 85–113. doi:10.1007/978-3-319-38756-7_4
4. Fda. Guidance for Industry PAT - A Framework for Innovative Pharmaceutical Development, manufacturing, and Quality Assurance [Internet]. 2004 [cited 2026 Sep 2]. Available from: <http://www.fda.gov/cvm/guidance/published.html>
5. Tao F, Zhang H, Liu A, Nee AYC. Digital Twin in Industry: State-of-the-Art. IEEE Trans Industr Inform. 2019 Apr;15(4):2405–15. doi:10.1109/TII.2018.2873186
6. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention Is All You Need. arXiv preprint arXiv:1706.03762. 2017. doi:10.48550/arXiv.1706.03762
7. Moreno-Benito M, Lee KT, Kaydanov D, Verrier HM, Blackwood DO, Doshi P. Digital twin of a continuous direct compression line for drug product and process design using a hybrid flowsheet modelling approach. Int J Pharm. 2022 Nov;628:122336. doi:10.1016/j.ijpharm.2022.122336

8. Copot D, Ayyaz B, Yumuk E. Realistic process simulator for control strategy evaluation in continuous direct compaction tablet manufacturing. *Control Eng Pract*. 2026 Apr;169:106712. doi:10.1016/j.conengprac.2025.106712
9. Tsallis C, Papageorgas P, Munteanu RA, Dellagi S. Large Language and Foundation Models for Machinery Health Monitoring: A Systematic Review. *Applied Sciences*. 2026 Mar 5;16(5):2493. doi:10.3390/app16052493
10. Directive 2011/62/EU amending Directive 2001/83/EC on the Community code relating to medicinal products for human use, as regards the prevention of the entry into the legal supply chain of falsified medicinal products [Internet]. 2011 [cited 2026 Sep 2]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32011L0062>
11. Drug Supply Chain Security Act (DSCSA) | FDA [Internet]. 2013 [cited 2026 Sep 2]. Available from: <https://www.fda.gov/drugs/drug-supply-chain-integrity/drug-supply-chain-security-act-dscsa>
12. Gopaluni RB, Kotamraju V, Bhushan S. Large Language Models in Process Systems Engineering: Opportunities, Architectures, and Industrial Deployment Challenges. *arXiv preprint arXiv:260611589*. 2026.
13. Yang L, Luo S, Cheng X, Yu L. Leveraging Large Language Models for Enhanced Digital Twin Modeling: Trends, Methods, and Challenges [Internet]. 2025 Mar 4 [cited 2026 Jul 21]. Available from: <https://arxiv.org/pdf/2503.02167>
14. Salami H, Smith-Goettler B, Yadav V. How can language models assist with pharmaceuticals manufacturing deviations and investigations? *Int J Pharm*. 2025 Jan;669:125100. doi:10.1016/j.ijpharm.2024.125100
15. Gill MS, Vyas J, Markaj A, Gehlhoff F, Mercangöz M. Leveraging LLM Agents and Digital Twins for Fault Handling in Process Plants. *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA*. 2025 May 4. doi:10.1109/ETFA65518.2025.11205597
16. Papacharalampopoulos A, Karagianni OM, Stavropoulos P. Cognitive Supervisory Control with LLM Reasoning Agent for Fault-Tolerant Process Systems: A Digital Twin Perspective. *Processes*. 2026 Jul 15;14(14):2298. doi:10.3390/pr14142298
17. Han Y, Tao J. Revolutionizing Pharma: Unveiling the AI and LLM Trends in the Pharmaceutical Industry [Internet]. 2024 Jan 5 [cited 2026 Jul 21]. Available from: <https://arxiv.org/pdf/2401.10273>
18. 21 CFR Part 11: Electronic Records; Electronic Signatures [Internet]. 1997 [cited 2026 Sep 2]. Available from: <https://www.ecfr.gov/current/title-21/chapter-I/subchapter-A/part-11>
19. EudraLex Volume 4: Guidelines for Good Manufacturing Practice for Medicinal Products for Human and Veterinary Use, Annex 11: Computerised Systems [Internet]. 2011 [cited 2026 Sep 2]. Available from: https://health.ec.europa.eu/medicinal-products/eudralex/eudralex-volume-4_en
20. Discussion Paper: Artificial Intelligence in Drug Manufacturing [Internet]. [cited 2026 Sep 2]. Available from: <https://www.regulations.gov/docket/FDA-2023-N-0487>
21. Considerations for the Use of Artificial Intelligence To Support Regulatory Decision-Making for Drug and Biological Products | FDA [Internet]. 2025 [cited 2026 Sep 2]. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-artificial-intelligence-support-regulatory-decision-making-drug-and-biological>