International Journal of Medical and Pharmaceutical Sciences
Open access peer-reviewed Journal
Volume 2, Issue 8, Page 227-241.
https://doi.org/10.5281/zenodo.21806708

Review Article

Data Integrity and ALCOA+ Principles in AI-Driven Regulatory Submissions

Anisha Nalwade*[1], Rupali Waghmode[1], Arti Dagadkhair[1], Sayali Nanware[2], Gauri Patil[3], Rajendra Patil[4]

[1]Department of Regulatory Affairs, Student of Delonix Society's Baramati College of pharmacy, Barhanpur, Maharashtra, India.
[2]Guide, Department of Regulatory Affairs, Asst. Prof. at Delonix Society's Baramati College of pharmacy, Barhanpur, Maharashtra, India.
[3]Head of Department of Regulatory Affairs, Delonix Society's Baramati College of Pharmacy, Barhanpur, Maharashtra, India.
[4]Principal, Delonix Society's Baramati College of Pharmacy, Barhanpur, Maharashtra, India

Artificial Intelligence (AI) is transforming pharmaceutical regulatory affairs by enhancing the efficiency, accuracy, and speed of regulatory submissions through technologies such as Machine Learning (ML), Natural Language Processing (NLP), Generative AI, and Robotic Process Automation (RPA). Despite these advancements, the successful adoption of AI depends on maintaining data integrity, which ensures that regulatory data remain accurate, complete, consistent, reliable, and traceable throughout the data lifecycle. This review discusses the importance of ALCOA+ principles—Attributable, Legible, Contemporaneous, Original, Accurate, Complete, Consistent, Enduring, and Available—as the foundation for preserving data integrity in AI-driven regulatory submissions. It highlights AI applications in data collection, eCTD publishing, regulatory intelligence, document authoring, and lifecycle management, while addressing challenges such as poor data quality, AI hallucinations, metadata integrity, audit trail deficiencies, version control, cybersecurity, and human oversight. The review further summarizes global regulatory expectations and emphasizes the need for risk-based validation, explainable AI (XAI), robust governance frameworks, and Human-in-the-Loop (HITL) approaches to ensure transparency, accountability, and regulatory compliance. Integrating AI with ALCOA+ principles strengthens data integrity, improves submission quality, enhances inspection readiness, and supports reliable regulatory decision-making across the pharmaceutical product lifecycle.

Keywords: Dossier, Electronic submission, Natural Language Processing, Machine Learning , Optical Character Recognition, data governance , Electronic Common Technical Document (ectd) , Explainable Artificial Intelligence (XAI).

INTRODUCTION

The pharmaceutical industry is undergoing a period of rapid transformation driven by advances in genomic science, digital health technologies, informatics, and a deeper understanding of disease biology. Over the past few decades, these developments have enabled the discovery of innovative therapies that significantly improve patient outcomes. [1] The use of novel evidence sources like real-world data and digital biomarkers is enhancing clinical evidence generation and supporting patient-centered drug development. Decentralized clinical trial methodologies using remote monitoring are alleviating patient burdens and improving efficiency. With pharmaceutical R&D becoming more data-

driven, regulatory systems must adapt to manage and assess new scientific evidence, necessitating the modernization of regulatory submissions and review processes to foster innovation and ensure timely access to safe medicines. [2]

1.1 Overview of Regulatory Submissions

Regulatory submissions are vital for the approval of pharmaceutical products, providing essential evidence of quality, safety, and efficacy to regulatory authorities. Utilizing standardized formats like the Common Technical Document (CTD) and Electronic Common Technical Document (eCTD), these submissions facilitate efficient information exchange and multidisciplinary review. They form the basis for marketing authorization applications and ensure compliance and transparency throughout the product lifecycle, ultimately speeding up patient access to safe medicines. [1] Traditionally, pharmaceutical dossiers were cumbersome, paper-based, and prone to errors, prompting the ICH to standardize submissions through the CTD. Despite improvements, challenges in review timelines and data management persisted as drug development grew more complex and globalized. [2]

1.2 Digital Transformation in Regulatory Affairs

The shift from paper Common Technical Document (CTD) to Electronic Common Technical Document (eCTD) in the pharmaceutical industry arises from the complexities of drug development and the need for rapid global submissions. [3] The eCTD enhances traceability, data management, and digital submissions, facilitates international standardization, reduces manual submission errors, and allows for automated validation tests by regulatory bodies. It encompasses lifecycle management tools essential for maintaining compliance during product updates. [4] The evolution of eCTD comprises three stages:

1. Paper CTD Era: Initial submissions were manual, time-consuming, and error-prone. [5,6]
2. Introduction of Electronic Submissions: In the early 2000s, various electronic formats were accepted, improving efficiency but suffering from inconsistencies.
3. Adoption of eCTD Standards: The ICH initiated eCTD standards to standardize electronic submissions, culminating in the current eCTD versions that optimize compliance and regulatory processes. [7]

1.3 Emergence of Artificial Intelligence in Regulatory Operations

Artificial Intelligence (AI) and its subsets, including Machine Learning (ML) and Natural Language Processing (NLP), are transforming the pharmaceutical regulatory environment by automating traditional processes such as document preparation and compliance monitoring [8,9]. AI enhances efficiency, accuracy, and speed in regulatory operations, allowing for improved data extraction, auditing, and quality management. [10] Current applications in regulatory submissions include automated data extraction, document authoring, and quality checks, significantly reducing manual efforts and errors while enabling faster and more consistent submissions. AI-driven decision support systems also provide predictive insights, helping regulatory professionals enhance compliance and decision-making. [11]

2. Fundamentals of Data Integrity

2.1 Definition of Data Integrity

According to WHO "Data integrity is the degree to which data are complete, consistent, accurate, trustworthy and reliable and that these characteristics of the data are maintained throughout the data life cycle. The data should be collected and maintained in a secure manner, such that they are attributable, legible, contemporaneously recorded, original or a true copy and accurate. Assuring data integrity requires appropriate quality and risk management systems, including adherence to sound scientific principles and good documentation practices." [13]

2.2 Importance in Pharmaceutical Regulatory Affairs

Data integrity is fundamental in pharmaceutical regulatory affairs, ensuring that data remain complete, consistent, reliable, and traceable throughout the product lifecycle. Regulatory agencies such as the FDA and EMA require compliance with data integrity principles to support product safety, efficacy, and

quality while preventing regulatory actions and delays. With increasing use of electronic systems and AI, adherence to **ALCOA+** principles is essential to maintain data authenticity, regulatory compliance, and patient safety. [12,14–16]

2.3 Data Lifecycle in Regulatory Submissions

The data lifecycle encompasses all stages through which data pass during their existence, including generation, recording, processing, review, approval, archival, retrieval, and eventual destruction. In pharmaceutical regulatory submissions, maintaining data integrity throughout these stages is essential to ensure compliance with regulatory requirements and support reliable decision-making. Data integrity must be preserved throughout the lifecycle to guarantee that data remain accurate, consistent, complete, and available for regulatory review. [16]

- Data generation
- Review
- Approval
- Archival
- Retrieval

3. ALCOA and ALCOA+ Principles

3.1 Evolution of ALCOA Concept

Data integrity in the pharmaceutical industry evolved from paper-based records before the 1980s to electronic systems and LIMS in the 1980s–1990s, improving efficiency but introducing new security challenges. The implementation of **21 CFR Part 11 (1997)** established requirements for electronic records and signatures, while later FDA, WHO, and MHRA guidelines strengthened lifecycle data governance. Since 2020, cloud platforms, electronic batch records, and automation have further improved data traceability, security, and regulatory compliance. [17]

3.2 ALCOA Principles

Data integrity must guarantee that records are: (i) authentic, unchangeable, and transparent—that is, the data cannot be erased;

(ii) traceable or auditable—that is, all data must have audit trails; and

(iii) safe—that is, the data is shielded from corruption and illegal access.

As a result, data must be gathered and stored securely so that it is:

- **Attributable** to the person generating the data
- **Legible** and permanent
- **Contemporaneous** – data is created/recorded when the activity is performed
- **Original** record (or ‘true copy’)
- **Accurate**

These traits are commonly referred to as the ALCOA principles. Data integrity in the pharmaceutical sector is guaranteed by the ALCOA principles. [18]

A. Attributable

The Attributable principle ensures that every data entry or activity can be clearly linked to the individual who performed it. Attribution may be established through initials, a full handwritten signature, a personal seal, the date, and when necessary, the time of the activity. The use of personal seals requires additional risk-management controls, including handwritten dates, restricted access, and secure storage of the seal to prevent unauthorized use. These measures ensure accountability and traceability of all recorded activities.

B. Legible, Traceable, and Permanent

The Legible, Traceable, and Permanent principle ensures that records remain readable, durable, and auditable throughout their lifecycle. Controls include the use of permanent or indelible ink, prohibition of pencil entries and erasures, documentation of corrections through single-line cross-outs with the recorder’s name, date, and reason, and avoidance of correction fluids that obscure original information. Additional controls include the use of controlled bound notebooks with sequentially numbered pages, numbered blank forms, secure archival systems, and

materials that resist fading over time. These practices preserve record integrity and facilitate traceability.

C. Contemporaneous

The Contemporaneous principle requires that data be recorded at the time the activity is performed. Compliance is supported through documented procedures, personnel training, audits, and self-inspections that emphasize real-time recording in authorized documents. Records should include the date and, where applicable, the time of the activity. Well-designed documents and readily available controlled forms promote timely documentation. Furthermore, synchronized and secure time sources should be used to ensure accurate time recording, particularly for time-sensitive activities.

D. Original

The Original principle requires the preservation, review, and retention of original records or verified true copies. Organizations adopting electronic systems must ensure that reviews and approvals are conducted on original records and that any modifications are appropriately documented, justified, and traceable. Reviewer and approver signatures demonstrate accountability, while procedures should define actions for addressing errors or omissions. Secure storage controls include controlled archives, independent archivists, indexed retrieval systems, periodic retrieval testing, and appropriate reading equipment for archived formats such as microfilm. When creating true copies, organizations must verify that the copies accurately preserve the content, meaning, and format of the original records and document this verification. Particular attention should be given to maintaining the integrity of handwritten signatures and ensuring complete traceability throughout the record lifecycle.

E. Accurate

The Accurate principle ensures that data are complete, correct, and reflective of actual observations. Accuracy is maintained through regular validation of analytical methods, routine validation of production processes, and systematic review of Good Documentation Practice (GDP) records. These controls help detect and prevent errors, ensuring that recorded information remains reliable, scientifically valid, and suitable for regulatory and quality-related decision-making. [19]

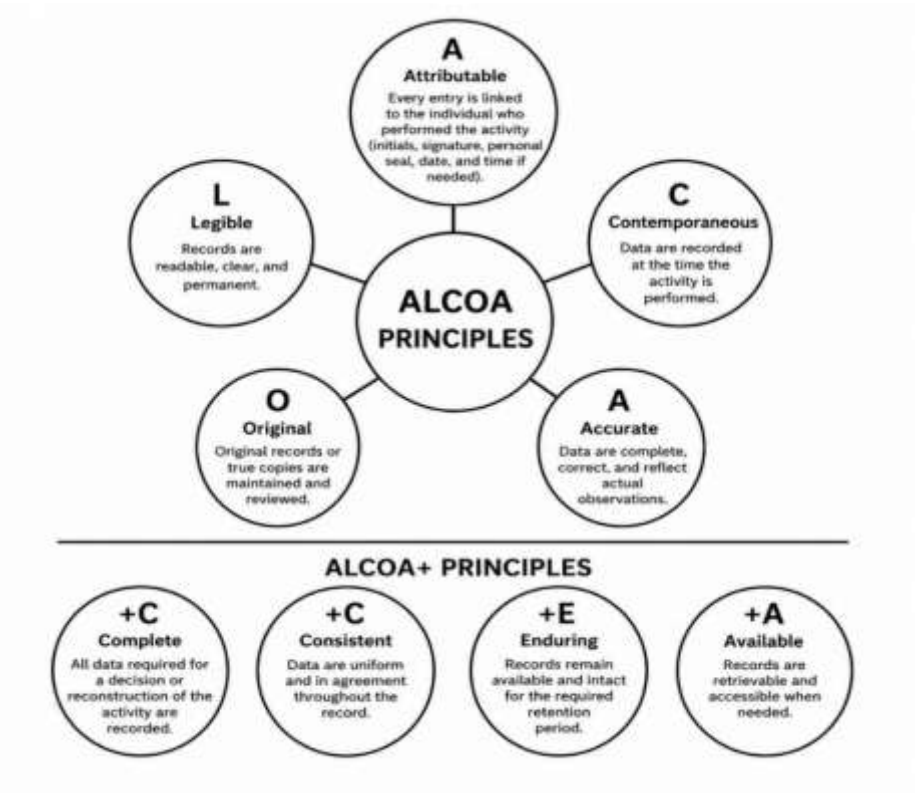


Fig. 1 ALCOA and ALCOA+ Principles.

3.3 ALCOA+ Extensions

- **Complete**

All pertinent information should be included in the data. Analysis and decision-making might be hampered by missing data. Example: To provide a thorough understanding of the procedure and experiment, records should contain all pertinent information. Incomplete or missing data might cause the results to be interpreted incorrectly.

- **Consistent**

Data uniformity across many records and systems is ensured by consistency. Errors and misunderstanding can result from inconsistencies. For instance, standardized protocols must to be adhered to uniformly throughout various studies or operations. Consistency ensures reliability and comparability of data over time

- **Enduring**

Data need to be durable and maintain its integrity throughout time. Archiving, backup, and proper storage are always seen as crucial.

- **Available**

Data must be accessible for validation, inspection, and review. Transparency and compliance depend on accessibility. Both digital and paper records must be readily available for inspections and investigations for the duration of the record. Both paper and electronic materials should have appropriate labelling and/or clear indexing to facilitate retrieval. [20]

4. AI Technologies Used in Regulatory Submissions

4.1 Artificial Intelligence in Regulatory Affairs

AI is transforming pharmaceutical regulatory affairs by automating routine tasks, analysing large datasets, and improving decision-making, accuracy, and efficiency. It enhances regulatory compliance, risk assessment, post-market surveillance, and stakeholder communication, enabling faster product approvals and improved patient safety. [21]

4.2 Machine Learning Applications

Machine learning (ML) is a branch of AI that enables computers to learn from data and make predictions or decisions without explicit programming. Using supervised, unsupervised, and reinforcement learning, ML analyses large datasets to identify patterns, optimize clinical trial design, detect adverse events, support regulatory decision-making, accelerate drug development, and improve patient safety.

4.3 Natural Language Processing (NLP)

Natural Language Processing (NLP) enables the analysis of unstructured text such as clinical trial reports, regulatory documents, and scientific literature. It extracts key information, summarizes lengthy reports, identifies adverse events and safety signals, and monitors social media for real-world drug safety insights, improving data analysis and supporting informed regulatory decision-making. [22]

4.4 Generative AI in Document Authoring

AI-driven document management systems automate the creation, categorization, and submission of regulatory documents using standardized templates. This improves consistency, reduces errors, enhances operational efficiency, minimizes compliance risks, and accelerates regulatory review and approval timelines.

4.5 Robotic Process Automation (RPA)

Although **Robotic Process Automation (RPA)** is not specifically mentioned, the article describes AI-driven process automation for document categorization, template creation, compliance monitoring, and administrative support. AI-powered assistants and chatbots improve operational efficiency, reduce manual effort, and support regulatory compliance.

4.6 AI-Enabled Regulatory Information Management Systems

AI-enabled information management in regulatory affairs automates document management, compliance tracking, real-time monitoring, and advanced data

analytics. These systems standardize submissions, categorize documents, monitor reporting deadlines, detect safety signals, support post-market surveillance, and strengthen data management and regulatory decision-making. [9]

5. AI-Driven Regulatory Submission Process

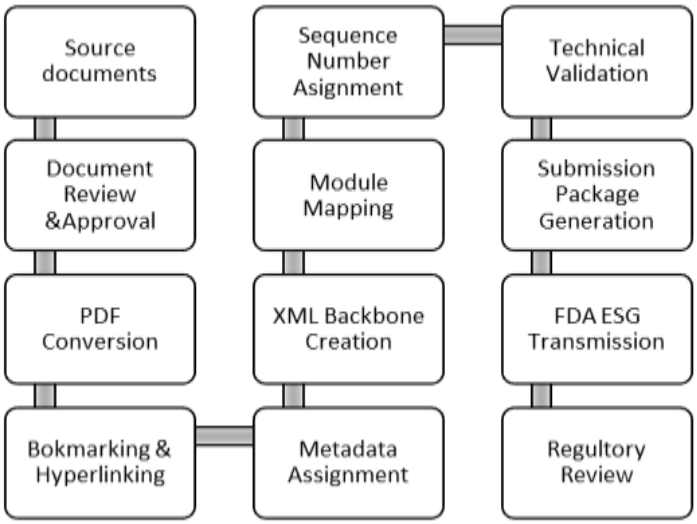
5.1 Data Collection and Compilation

Data collection for AI-driven regulatory submissions involves information from clinical trial protocols, case report forms (CRFs), patient narratives, and medical literature, much of which is unstructured and requires extensive manual processing. [23] AI technologies such as Natural Language Processing (NLP), Machine Learning (ML), and Optical Character Recognition (OCR) automate data extraction and standardization, while Human-in-the-Loop (HITL) frameworks ensure human oversight. Transformer-based NLP extracts key clinical information, OCR converts scanned documents into machine-readable text, and confidence scoring flags uncertain data for human review, improving data quality, efficiency, and regulatory compliance. [24]

5.2 eETD Publishing and Validation

The Electronic Common Technical Document (eCTD) is the internationally accepted format for electronic regulatory submissions to health authorities. The FDA requires eCTD for NDA, ANDA, BLA, commercial IND, and related submissions. eCTD standardizes information into Modules 1–5, supporting efficient review, lifecycle management, archiving, and retrieval. Currently, eCTD Versions 3.2.2 and 4.0 (HL7 RPS-based) are supported for improved interoperability and lifecycle management. [25,26]

eCTD Publishing Workflow [2,27]



eCTD Validation Check

Validation is a critical pre-submission step that ensures technical acceptability and regulatory compliance. The FDA requires submissions to pass validation checks, including XML backbone, checksum integrity, folder structure, file naming, PDF compliance, hyperlinks, metadata, sequence numbering, and submission completeness. Failure to meet these requirements may result in rejection of the submission package. [28] Major eCTD Validation Checks include XML backbone, folder structure, file naming, PDF specifications, hyperlinks and bookmarks, checksum integrity, metadata, sequence numbering, lifecycle operations, and submission completeness. [29]

5.3 Intelligence and Gap Analysis

Regulatory Intelligence (RI) uses AI and Machine Learning to systematically collect and analyze regulatory information, enabling proactive compliance, risk management, and lifecycle decision-making. It supports impact assessment by evaluating regulatory changes, identifying trends, predicting regulatory requirements, and improving submission quality, inspection readiness, and timely regulatory approvals. From a data integrity perspective, AI-

enabled RI upholds ALCOA+ principles—ensuring that regulatory information is Accurate, Contemporaneous, Complete, and Consistent. Automated systems continuously capture updates, while centralized platforms ensure the secure storage and retrieval of data, improving the reliability of regulatory submissions. Ultimately, RI is vital for maintaining compliance and submission quality in increasingly complex global regulatory landscapes. [4,5,11,30,31]

5.4 Submission Tracking and Lifecycle Management

Over the course of a pharmaceutical product's lifespan, regulatory filings continue to change. Continuous management and oversight are necessary for tasks including modifications, renewals, yearly reports, safety updates, and post-approval obligations.AI-enabled lifecycle management systems provide enhanced visibility into these activities and support proactive regulatory oversight.

To support ALCOA+ compliance, lifecycle management systems should maintain:

- Comprehensive audit trails.
- Change history documentation.
- Metadata integrity.
- User accountability.
- Long-term record preservation.

These controls ensure traceability throughout the submission lifecycle and support regulatory inspection readiness.

5.5 AI-Assisted Health Authority Responses

During regulatory review, health authorities often issue requests for information, deficiency letters, and clarification questions that require rapid access to relevant data and coordination among stakeholders. AI supports information retrieval, response drafting, knowledge reuse, and response quality assessment, improving efficiency. However, scientific accuracy, regulatory strategy, final approval, and submission accountability remain the responsibility of qualified regulatory professionals, and all AI-assisted responses must be reviewed before submission.

6. Data Integrity Risks in AI-Driven Regulatory Submissions

Artificial intelligence has improved regulatory operations through enhanced document preparation, data extraction, lifecycle management, and submission quality. However, AI-generated outputs require proper governance, validation, and human oversight to comply with **ALCOA+** principles. Regulatory agencies emphasize that AI-generated data must remain attributable, accurate, complete, contemporaneous, consistent, enduring, available, and traceable to ensure data integrity, regulatory compliance, and reliable decision-making. [32]

6.1 Data Quality Issues

High-quality input data are essential for trustworthy AI systems, as AI models learn from historical datasets. Incomplete, inconsistent, duplicated, outdated, or biased data can produce unreliable outputs, leading to errors in clinical data extraction, adverse event classification, metadata generation, and regulatory dossier preparation. From the perspective of ALCOA+, poor-quality data primarily violate the principles of Accuracy, Completeness, Consistency, and Originality. Since AI cannot distinguish between correct and incorrect source information without appropriate validation, data quality assessment should include automated quality checks, standardized vocabularies, source verification, and routine data governance before AI processing. Regulatory agencies therefore recommend implementing comprehensive data governance frameworks that ensure reliable source data before AI deployment. [32]

ALCOA+ principles affected

- Accurate
- Complete
- Consistent
- Original

6.2 AI Hallucinations and Inaccurate Content Generation

Large Language Models (LLMs) may generate scientifically plausible but factually incorrect information, known as AI hallucinations, including fabricated clinical data, references, statistical values, product specifications, or regulatory interpretations. Such errors can compromise regulatory documents, making independent scientific verification and human review essential. In line with ALCOA+ principles, AI-generated content must be attributable to verified evidence, accurate, and fully traceable to original source records. [33]

ALCOA+ principles affected

- Accurate
- Attributable
- Complete
- Traceable

6.3 Data Traceability Challenges

Regulatory authorities require complete traceability throughout the document lifecycle to verify data origin, processing history, analytical results, and document authenticity. Since AI integrates information from multiple sources, maintaining **provenance records**—including source documents, prompts, model versions, timestamps, reviewer identity, and validation evidence—is essential to meet **ALCOA+** requirements. Electronic provenance systems, unique document identifiers, digital signatures, and audit trails further strengthen traceability in AI-assisted regulatory workflows. [34]

ALCOA+ principles affected

- Traceable
- Attributable
- Original
- Available

6.4 Metadata Integrity Risks

Metadata, including creation date, author identity, electronic signatures, software version, timestamps, modification history, and document status, are an integral part of electronic regulatory records. AI-assisted editing, document conversion, cloud synchronization, and data migration may alter or remove metadata, compromising record authenticity and traceability. Therefore, validated electronic document management systems should preserve metadata throughout the submission lifecycle and prevent unauthorized modification. [32]

ALCOA+ principles affected

- Attributable
- Contemporaneous
- Original
- Enduring
- Available

6.5 Audit Trail Deficiencies

Audit trails provide chronological records of **who** performed an activity, **when** it occurred, **what** changes were made, and **why**, making them essential for compliance with **21 CFR Part 11** and electronic record requirements. AI-assisted regulatory systems should maintain immutable audit logs, secure timestamps, prompt version history, reviewer documentation, model validation records, and automated change tracking to ensure accountability, traceability, and regulatory compliance. [34]

ALCOA+ principles affected

- Attributable
- Contemporaneous
- Enduring
- Available
- Traceable

6.6 Version Control Issues

AI-generated regulatory documents often undergo multiple revisions involving automated content generation, human editing, reviewer comments, and

regulatory updates. Without effective version control, organizations risk using outdated information, conflicting datasets, or obsolete regulatory guidance. Version control failures may produce inconsistencies across different CTD modules, resulting in contradictory product information or duplicated content. Maintaining centralized document repositories with controlled access, electronic signatures, and automated version histories supports compliance with ALCOA+ requirements. [34]

ALCOA+ principles affected

- Original
- Consistent
- Complete
- Available
- Traceable

6.7 Cybersecurity and Data Privacy Concerns

AI-driven regulatory systems process confidential clinical, manufacturing, patient, and regulatory data, making them vulnerable to cybersecurity threats such as unauthorized access, data manipulation, ransomware, and data breaches. To protect data integrity and regulatory compliance, organizations should implement encryption, identity management, multi-factor authentication, role-based access control, continuous security monitoring, and secure cloud governance, ensuring the authenticity, availability, and confidentiality of electronic records. [35]

ALCOA+ principles affected

- Original
- Available
- Accurate
- Enduring

6.8 Human Oversight Limitations

Despite advances in AI, regulatory responsibility remains with qualified professionals. Overreliance on AI and automation bias may lead to acceptance of inaccurate outputs, especially without adequate training or validation. Effective governance requires human-in-the-loop review, multidisciplinary validation, documented approval workflows, periodic model evaluation, and continuous training to ensure regulatory compliance and data integrity. Human oversight remains essential for ensuring that AI-assisted regulatory submissions comply with ALCOA+ principles and regulatory expectations. [33]

ALCOA+ principles affected

- Attributable
- Accurate
- Complete
- Traceable

7. Application of ALCOA+ Principles in AI-Driven Regulatory Submissions

Artificial Intelligence (AI) has transformed pharmaceutical regulatory submissions by improving data processing, document preparation, and decision support. However, AI also introduces challenges related to data integrity, transparency, traceability, and regulatory compliance. [36,37,38] Applying **ALCOA+ principles** ensures that AI-generated regulatory data remain reliable, accurate, and inspection-ready throughout the submission lifecycle. [39] These principles encompass algorithm governance, metadata management, audit trails, validation, and human oversight. [40]

7.1 Attributable Data and User Accountability

All user and AI actions should be traceable through unique user IDs, audit trails, AI output documentation, model versions, timestamps, and reviewer approvals to ensure transparency and accountability. [41,42]

7.2 Legibility and Structured Data Formats

AI-generated regulatory information should remain readable and accessible using standardized templates, eCTD/XML formats, consistent terminology, controlled vocabularies, and preserved metadata. [43,44]

7.3 Contemporaneous Documentation

AI supports real-time documentation through automated timestamps, event logging, workflow records, and continuous audit trails, improving traceability and reducing retrospective data entry. [45]

7.4 Original Source Data Verification

AI-generated outputs must remain linked to original source records through data lineage, validated datasets, and retention of prompts, source documents, reviewer comments, and approval records. [46]

7.5 Accuracy and AI Output Validation

AI-generated information should be validated through expert review, independent verification, model validation, and continuous monitoring to minimize bias, hallucinations, and errors, ensuring accurate regulatory submissions. [47–49]

7.6 Completeness of Submission Packages

AI helps identify missing documents, verify submission requirements, assess dossier completeness, and ensure inclusion of supporting datasets, metadata, audit trails, and validation records. [50–52]

7.7 Consistency Across Submission Modules

AI-based Natural Language Processing (NLP) detects discrepancies, verifies terminology, and ensures consistency across submission modules, improving submission quality. [53,54]

7.8 Enduring Data Retention Strategies

Regulatory records, including AI models, training datasets, validation reports, audit trails, and system configurations, should be securely retained throughout required retention periods. [55,56]

7.9 Availability During Regulatory Inspections

AI improves inspection readiness through intelligent search, automated document classification, knowledge management, and rapid retrieval of complete regulatory records, supporting ALCOA+ compliance. [57]

8. Regulatory Framework Governing AI and Data Integrity

Table 2: Comparison of Global Regulatory Expectations for AI and Data Integrity [58]

Regulatory Authority / Framework	Data Integrity Focus	AI-Relevant Expectations	Key Compliance Requirements
FDA (21 CFR Part 11, Data Integrity Guidance)	Data lifecycle integrity	Transparency, validation of AI-enabled computerized systems	Audit trails, electronic records/electronic signatures (ERES), Part 11 compliance, system validation
MHRA (gxp Data Integrity Guidance)	ALCOA+ compliance	Accountability, governance, human oversight	Data lifecycle controls, audit trails, role-based access, governance
WHO (Good Data and Record Management Practices)	Good data management	Ethical AI use, reliability, transparency	Documentation, oversight, quality management systems
ICH E6(R3)	Clinical data reliability	Risk-based computerized systems, AI-supported clinical processes	Computerized system validation, governance, data integrity throughout clinical trials
ICH Q9(R1)	Quality risk management	AI risk assessment and mitigation	Risk identification, control measures, continual risk review
ICH Q10	Pharmaceutical quality systems	AI lifecycle governance	Continuous improvement, management responsibility, change management
PIC/S (PI 041 Data Integrity Guidance	GMP data integrity	Computerized system controls	Audit trails, metadata protection, system security, data governance

GAMP 5 (2nd Edition)	Computerized system validation	AI lifecycle management, machine learning validation	Risk-based validation, testing, documentation, lifecycle management
NIST AI Risk Management Framework (AI RMF 1.0)	Trustworthy AI	Governance, monitoring, transparency, explainability	Risk management framework, AI monitoring, accountability, human oversight
EU AI Act	AI transparency and accountability	Human oversight, trustworthy AI, risk-based regulation	Technical documentation, transparency, accountability, conformity assessment, human oversight

9. Validation and Governance of AI Systems

Successful adoption of AI in pharmaceutical regulatory submissions requires robust validation and governance to ensure data integrity, regulatory compliance, and patient safety throughout the AI lifecycle. Unlike conventional software, AI requires lifecycle-based validation, following frameworks such as CSV, FDA CSA, 21 CFR Part 11, EU GMP Annex 11, and ISPE GAMP® 5, with assessment of data quality, model performance, bias, explainability, robustness, reproducibility, and continuous monitoring in accordance with ALCOA+ principles.[59–62] Risk-based frameworks, including ICH Q9(R1), GAMP® 5, the NIST AI Risk Management Framework, and the EU AI Act, recommend validation proportional to AI risk, supported by secure audit trails, formal change control, and ongoing monitoring.[61–64,66,67] Human-in-the-Loop (HITL) review remains essential for high-risk applications, while an effective AI governance framework integrates validation, quality management, risk assessment, cybersecurity, ethics, and lifecycle monitoring to ensure reliable, transparent, and compliant AI systems.[61,63,65,66,68]

10. Industry Applications and Case Studies

10.1 AI-Assisted eCTD Compilation

Artificial intelligence is revolutionizing the preparation of regulatory submissions by automating data processing, document assembling, and the creation of structured content. Reusing regulatory data, minimizing mistakes, reducing human transcribing, and expediting the compilation of electronic Common Technical Document (eCTD) submissions are all made possible by the integration of AI with Structured Content and Data Management (SCDM). These solutions facilitate effective administration of regulatory dossiers across the product lifecycle, increase consistency, and improve data quality. [29]

10.2 Automated Regulatory Document Authoring

Generative AI assists in drafting clinical summaries, QOS, investigator brochures, regulatory responses, and labelling documents, improving efficiency and consistency. However, all AI-generated content requires expert scientific review to prevent hallucinations, citation errors, and unsupported conclusions. [69]

10.3 AI-Based Submission Quality Checks

AI improves submission quality by checking document consistency, completeness, technical compliance, terminology, and identifying high-risk areas, while final verification and approval remain the responsibility of qualified regulatory professionals. [70,71]

10.4 Lessons from Regulatory Inspections

Regulatory inspections emphasize data integrity through adequate audit trails, access control, system validation, metadata preservation, and traceability, highlighting the need for strong governance and validation in AI-enabled systems. [72]

11. Future Perspectives

11.1 Explainable AI (XAI)

Explainable AI (XAI) improves transparency, interpretability, and trust by explaining AI decision-making, identifying influencing data, detecting bias, and documenting decision pathways, data lineage,

and confidence scores to support regulatory acceptance and ALCOA+ compliance. [73]

11.2 AI Governance and Regulatory Acceptance

Future AI adoption will depend on robust governance, transparency, validation, accountability, and risk management. Collaboration among regulators, industry, and academia is expected to harmonize AI guidance and expand AI-assisted regulatory review and inspection planning. [74]

11.3 Blockchain for Data Integrity

Blockchain enhances data integrity through immutable records, tamper-resistant audit trails, secure data sharing, and improved traceability, supporting ALCOA+ principles. However, adoption remains limited by scalability, integration challenges, and regulatory uncertainty.[75]

11.4 AI-Powered Smart Submissions

Future AI-powered smart submissions will use structured, machine-readable data, automated validation, regulatory intelligence, and dynamic content generation to improve submission quality, compliance, and regulatory review efficiency. [76,77].

REFERENCES

1. Macdonald JC, Isom DC, Evans DD, Page KJ. (2021)
2. Digital Innovation in Medicinal Product Regulatory Submission, Review, and Approvals to Create a Dynamic Regulatory Ecosystem—Are We Ready for a Revolution?
3. Frontiers in Medicine, 8:660808.
4. Salunkhe P, Nagrale. S. (2026) Electronic Common Technical Document (eCTD): Structure, Submission Process, and Regulatory Updates. International Journal of Drug Regulatory Affairs.
5. Bishnoi M, Sonker A. Emergency use authorization of medicines: History and ethical dilemma. Perspect Clin Res. 2023 Apr-Jun;14(2):49-55. doi: 10.4103/picr.picr_149_22. Epub 2023 Apr 13. PMID: 37325578; PMCID: PMC10267997.
6. Collier R, Singh A, Chopra S. Harnessing AI/ML in drug and biological products discovery and development: The regulatory perspective. Pharmaceuticals (Basel) [Internet]. 2025 Jan [cited 2026 Feb 18];18(1):47. Available from: doi:10.3390/ph18010047.
7. Liu Q, Huang R, Hsieh J, Zhu H, Tiwari M, Liu G, et al. Landscape analysis of the application of artificial intelligence and machine learning in regulatory submissions for drug development from 2016 to 2021. Clin Pharmacol Ther [Internet]. 2023 Apr [cited 2026 Feb 18];113(4):771-774. Available from: doi:10.1002/cpt.2668.
8. Suresh V, Kumar KS. Approval and marketing of medical products in EU, USA, India and Australia. Asian J Med Pharm Sci [Internet]. 2023 [cited 2026 Feb 18];11(1):73 82. Available from: doi: 10.30904/j.ajmps.2023.4619.
9. Sireesha B, Haritha S, Lakshmi Thriveni S, Nikitha P, Vijaya Lakshmi UVS. Regulatory affairs and regulatory requirements for new drug approval. Int J Health Care Biol Sci [Internet]. 2024 [cited 2026 Feb 18];5(4):22-28. Available from: doi:10.46795/ijhcbs. v5i4.643.
10. O.F. Ahmad et al. Barriers and pitfalls for artificial intelligence in gastroenterology: ethical and regulatory issues.
11. Boopathi LK, Raja N. Artificial Intelligence and Machine Learning: Enhancing the Future of Regulatory Affairs. Int. J. Drug Reg. Affairs [Internet]. 2024 Dec15 [cited 2024 Dec 15]; 12(4):53-60. Available from: http://ijdra.com/index.php/journal/article/view/718
12. Patil RS, Kulkarni SB, Gaikwad VL. Artificial intelligence in pharmaceutical regulatory affairs. Drug Discovery Today. 2023;28(9):103700.
13. Ajmal CS, et al. Innovative Approaches in Regulatory Affairs: Leveraging Artificial Intelligence and Machine Learning for Efficient Compliance and Decision-Making. AAPS Journal. 2025;27(1):22.
14. Ajmal CS, et al. Innovative Approaches in Regulatory Affairs: Leveraging Artificial Intelligence and Machine Learning for Efficient Compliance and Decision-Making. AAPS Journal. 2025;27(1):22.

15. World Health Organization. WHO Technical Report Series No. 1033, Annex 4: Guideline on Data Integrity.
16. Madhanraj S, Smith AA. Data Integrity in Pharmaceuticals. Int J Drug Regul Aff. 2025;13(4):26–36.
17. Rattan AK. Data Integrity: History, Issues, and Remediation of Issues. PDA J Pharm Sci Technol. 2018;72(2):105–116.
18. Charoo NA, Khan MA, Rahman Z. Data Integrity Issues in Pharmaceutical Industry: Common Observations, Challenges and Mitigation Strategies. Int J Pharmaceutics. 2023; 631:122503.
19. Himanshu, Himanshu Saini, Km. Reenu. DATA INTEGRITY AND ALCOA+ Principles In The Pharmaceutical Industry. Journal of Advanced Pharmaceutical Sciences and Natural Products. 2026; 1(3).
20. Pharmaceutical Inspection Convention (Ed.), Good Practices for Data Management and Integrity in Regulated GMP/GDP Environments, Pharmaceutical Cooperation Scheme Secretariat, 2018, PI 041-1 (Draft 3).
21. Sihite, E.A.C.P.B. & Sumarna, A.D. Implementation of ALCOA Principles in Paper-Based Documentation Systems (pp. 331–332).
22. Divya Gokulakrishnan, Sowmyalakshmi Venkataraman Department of Pharmaceutical Chemistry, Sri Ramachandra Faculty of Pharmacy, Sri Ramachandra Institute of Higher Education & Research (DU), Porur, Chennai, 600116, India.
23. Shelar S, Bagdane S, Joshi P. Drug regulatory affairs: the potential for machine learning and artificial intelligence. Pharm Reson. 2024;6(2):1-10. Available from: https://pharmacy.dypvp.edu.in.
24. Gude S, Gude YS. The synergy of artificial intelligence and machine learning in revolutionizing pharmaceutical regulatory affairs. Translational and Regulatory Sciences. 2024;6(2):37–45. doi: 10.33611/trs.2024-005.
25. Yogesh S Ahire et al., "Advanced Applications of Artificial Intelligence in Pharmacovigilance: Current Trends and Future Perspectives," Journal of Pharmaceutical Research, 2024.
26. Kamble A. AI-Powered Clinical Document Intelligence: Automating Data Extraction, Standardization, and Review. Journal of Information Systems Engineering and Management. 2026;11(3s):883–903. doi: 10.52783/jisem.v11i3s.14828.
27. International Council for Harmonisation. M4 Organization of the Common Technical Document for the Registration of Pharmaceuticals for Human Use. 2017.
28. U.S. Food and Drug Administration. Providing Regulatory Submissions in Electronic Format—Certain Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications. 2024.
29. Ahluwalia K, Abernathy MJ, Algorri M, et al. The Future of Regulatory Filings: Digitalization. AAPS Open. 2025.
30. U.S. Food and Drug Administration. Providing Regulatory Submissions in Electronic Format: Certain Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications. Guidance for Industry. Silver Spring (MD): FDA; 2024. Available from: FDA eCTD Guidance
31. FDA. Specifications for eCTD Validation Criteria.
32. Nimbalkar S.S, Nagrale S. Pharmaceutical Regulatory Intelligence: Tools, Strategies, and Impact on Product Lifecycle Management. International Journal of Drug Regulatory Affairs. 2026;14(1):61–70.
33. European Medicines Agency. HMA-EMA Joint Big Data Steering Group Workplan 2022–2025. 2023.
34. Rattan AK. Data Integrity: History, Issues, and Remediation of Issues. PDA Journal of Pharmaceutical Science and Technology. 2018;72(2):105–116.
35. Mökander J. Auditing of AI: Legal, Ethical and Technical Approaches. Digital Society. 2023; 2:49.
36. Mora-Cantallops M, Sánchez-Alonso S, García-Barriocanal E, Sicilia MA. Traceability for Trustworthy AI: A Review of Models and Tools. Big Data and Cognitive Computing. 2021;5(2):20.
37. Elsheikh, A.M.A., Rokne, J. & Alhajj, R. Privacy-preserving in foundation models: a systematic review of techniques, threats, and trade-offs. Artif

Intell Rev 59, 145 (2026). https://doi.org/10.1007/s10462-026-11535-4.
38. FDA. Data Integrity and Compliance with Drug CGMP: Questions and Answers. Guidance for Industry. 2018.
39. MHRA. GxP Data Integrity Guidance and Definitions. 2018.
40. WHO. Guidance on Good Data and Record Management Practices. Technical Report Series No. 996; Annex 5. 2016.
41. PIC/S PI 041-1. Good Practices for Data Management and Integrity in Regulated GMP/GDP Environments. 2021.
42. ISPE GAMP 5 Second Edition. 2022.
43. FDA 21 CFR Part 11 Electronic Records; Electronic Signatures.
44. NIST AI Risk Management Framework. 2023.
45. ICH M4 Common Technical Document Guideline.
46. PIC/S Data Integrity Guidance. 2021.
47. ISPE AI Governance Framework. 2024.
48. ICH Q10 Pharmaceutical Quality System.
49. Ji Z, et al. Survey of Hallucination in Natural Language Generation. ACM Comput Surv. 2023;55(12):248.
50. WHO Ethics and Governance of Artificial Intelligence for Health. 2021.
51. ICH M4 CTD Guideline.
52. Ajmal CS, et al. AAPS Journal. 2025;27(1):22.
53. ISPE Data Governance Framework. 2023.
54. EMA eCTD Lifecycle Management Guidance. 2024.
55. Nature Reviews Drug Discovery. AI in Regulatory Science. 2024.
56. ISO 15489 Records Management Standard.
57. ISPE AI Lifecycle Management Guidance. 2024.
58. ICH E6(R3) Good Clinical Practice. 2025.
59. Balta I, Lemon J, Popescu CA, McCleery D, Iancu T, Pet I, Stef L, Douglas A, Corcionivoschi N. Food safety–the transition to artificial intelligence (AI) modus operandi. Trends in Food Science & Technology. 2025 Sep 4:105278.
60. U.S. Food and Drug Administration. Computer Software Assurance for Production and Quality System Software. Silver Spring (MD): FDA; 2022.
61. U.S. Food and Drug Administration. 21 CFR Part 11: Electronic Records; Electronic Signatures. Silver Spring (MD): FDA; 1997.
62. ISPE. GAMP® 5: A Risk-Based Approach to Compliant GxP Computerized Systems. 2nd ed. Tampa (FL): International Society for Pharmaceutical Engineering; 2022.
63. Pharmaceutical Inspection Co-operation Scheme (PIC/S). PI 041-1: Good Practices for Data Management and Integrity in Regulated GMP/GDP Environments. Geneva: PIC/S; 2021.
64. International Council for Harmonisation (ICH). ICH Q9(R1): Quality Risk Management. Geneva: ICH; 2023.
65. International Council for Harmonisation (ICH). ICH Q10: Pharmaceutical Quality System. Geneva: ICH; 2008.
66. Myllyaho L, Virtanen S, et al. Validation methods for artificial intelligence systems: A systematic literature review. J Syst Softw. 2021; 181:111050.
67. National Institute of Standards and Technology (NIST). Artificial Intelligence Risk Management Framework (AI RMF 1.0). Gaithersburg (MD): NIST; 2023.
68. European Medicines Agency. Reflection Paper on the Use of Artificial Intelligence in the Medicinal Product Lifecycle. Amsterdam: EMA; 2024.
69. OECD. OECD Principles on Artificial Intelligence. Paris: Organisation for Economic Co-operation and Development; 2019.
70. Nature Reviews Drug Discovery. AI in Regulatory Science. 2024.
71. DIA Digital Regulatory Transformation Survey. 2024.
72. EMA Artificial Intelligence Reflection Paper. 2024.
73. FDA, EMA, MHRA, PIC/S inspection reports and data integrity observations.
74. ISPE GAMP 5 Second Edition. 2022.
75. DAMA International. Data Management Body of Knowledge (DMBOK2). 2017.
76. PIC/S PI 041-1 Data Integrity Guidance. 2021.
77. Ahluwalia K, Abernathy MJ, Algorri M, Bork M, et al. The Future of Regulatory Filings: Digitalization. AAPS Open. 2025; 11:9.
78. Ajmal CS, et al. Innovative Approaches in Regulatory Affairs: Leveraging Artificial Intelligence and Machine Learning for Efficient Compliance and Decision-Making. AAPS Journal. 2025;27(1):22.

79. Khan MAA, Hamid S, Khan SA, Sarfraz M, Babar ZUD. Stakeholders' views on pharmacovigilance system and policy coordination in Pakistan. Front Pharmacol [Internet]. 2022 [cited 2026 Feb 18]; 13:891954. Available from: doi:10.3389/fphar.2022.891954.

Cite: Anisha Nalwade*, Rupali Waghmode, Arti Dagadkhair, Sayali Nanware, Gauri Patil, Rajendra Patil, Data Integrity and ALCOA+ Principles in AI-Driven Regulatory Submissions, Int. J. Med. Pharm. Sci., 2026, 2 (8), 227-241. https://doi.org/10.5281/zenodo.21806708