

# Challenges of using Artificial Intelligence (Natural Language Processing) tools for regulatory documentation

Harshada B. Pawar<sup>1</sup>, Neelanshu<sup>1</sup>, Gowri M. Bhat<sup>2</sup>, Arun Srikanth<sup>3</sup>, Pradeep M. Muragundi<sup>1\*</sup> 

<sup>1</sup>Department of Pharmaceutical Regulatory Affairs and Management, Manipal College of Pharmaceutical Sciences, Manipal Academy of Higher Education, Manipal, India.

<sup>2</sup>Bioplus Life Sciences Pvt. Ltd., Pharmed Gardens, Bengaluru, India.

<sup>3</sup>Onesource Specialty Pharma Ltd., Bengaluru, India.

## ARTICLE HISTORY

Received on: 30/06/2025

Accepted on: 09/09/2025

Available Online: XX

## Key words:

Artificial Intelligence (AI), Natural Language Processing (NLP), regulatory documentation, compliance challenges, ethical concerns.

## ABSTRACT

Artificial intelligence (AI), specifically Natural Language Processing (NLP), is revolutionizing pharmaceutical regulatory documents by facilitating automated extraction of data, real-time regulatory change monitoring, and standardization of submissions like the Common Technical Document (CTD), enhancing accuracy and efficiency over the conventional manual approach. In spite of these advantages, obstacles exist in dealing with dynamic regulatory needs, intricate legal and scientific terminology, algorithmic uncertainty, and issues relating to data quality, ethics, and compliance, particularly when NLP technologies become more integrated into Software as a Medical Device. For an investigation of these concerns, this narrative review reviewed regulatory materials from authorities such as the U.S. Food and Drug Administration, European Medicines Agency, Medicines and Healthcare Products Regulatory Agency, and international organizations such as the World Health Organization and co-operation and development, complemented by a subject-specific search of Scopus, PubMed, and Google Scholar (2015–2025). English-language peer-reviewed articles, regulatory guidelines, and industry reports relevant to the topic were included and thematically synthesized under opportunities, challenges, ethical issues, and regulatory views. The results indicate that despite the immense potential of AI/NLP solutions to improve regulatory effectiveness and transparency, they pose intricate technical, regulatory, ethical, and operational issues requiring harmonized international standards, stringent validation frameworks, and multistakeholder engagement to deploy safely and effectively in regulatory science.

## 1. INTRODUCTION

Professions involving big text data have been revolutionized with the introduction of Artificial Intelligence (AI) and, more importantly, with the evolution of Natural Language Processing (NLP) tools [1]. NLP, a branch of AI, enables computers to read, analyze, and produce human language, transforming unstructured text into structured, readable data [2].

Among the sectors impacted by NLP, regulatory documentation has been one of the key areas of use. Organizations operating in highly regulated sectors such as pharma, healthcare, finance, and legal compliance have to prepare vast dossiers of technical files, clinical-study reports, manufacturing documents, and compliance statements to present evidence of international standards of quality, safety, and efficacy compliance [3]. Tedious, time-consuming, and error-prone conventional manual preparation and revision of document processes severely hinder efficiency, increase the chances of inconsistencies, and extend regulatory submission times [4]. The use of AI-enabled NLP technologies is on the verge of offering new alternatives by automating the handling of large amounts of regulatory documents, improving accuracy and efficiency. NLP can extract crucial data from complex text,

## \*Corresponding Author

Pradeep M. Muragundi, Department of Pharmaceutical Regulatory Affairs and Management, Manipal College of Pharmaceutical Sciences, Manipal Academy of Higher Education, Manipal, India.  
E-mail: [pradeep.mm@manipal.edu](mailto:pradeep.mm@manipal.edu)

detect regulatory changes in real-time, and enforce consistency in submissions [3].

Although these benefits exist, the use of AI-based NLP systems on regulatory texts is not free of challenges. The constantly evolving regulations need along with multifaceted legal and scientific jargon, are issues that present gigantic challenges to algorithms being able to properly read and understand context and generate compliant output [4].

Concurrently, NLP modules are increasingly incorporated into Software as a Medical Device (SaMD), where they support health care decision-making. Although these technologies promise greater efficiency and precision, they pose questions regarding reliability, possible clinical mistakes, and regulatory monitoring [5,6]. These technologies have raised concerns, for example, among the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA) have recently published formal guidance on Artificial Intelligence/Machine Learning (AI/ML)-based SaMDs and regulatory use of AI [7,8]. The World Health Organization (WHO) has also stressed the need for ethical governance of AI in health [9].

This review discusses the integration of AI and NLP into pharmaceutical regulatory texts, including both the benefits and challenges. It takes a technical, regulatory, ethical, and operational view, and considers the emerging regulatory structures that will determine the future of AI-powered tools in this critical area.

2. METHODOLOGY

This narrative review essentially centered on regulatory agency documents and global guidelines to examine applications of AI and NLP technologies in regulatory science. Major sources were official documents of the U.S. FDA, EMA, Medicines and Healthcare Products Regulatory Agency, WHO, and the Organisation for Economic Co-operation and Development. Professional societies and regulation-oriented industry forums such as Drug Information Association and Regulatory Affairs Professionals Society were also in consideration since they are representative of the developers and consumers of AI-based regulatory software.

To supplement these regulatory resources, a search was also conducted through academic databases such as Scopus, PubMed, and Google Scholar with an emphasis on peer-reviewed scientific literature published in the last 10 years (2015–2025). Key regulatory documents like International Council for Harmonisation (ICH) guidelines were supplemented as needed for historical and context background.

Criteria for inclusion were English-language articles that discussed AI or NLP as it pertained to regulatory science, compliance, or medicine. Exclusion criteria were literature that was not English language, noncredible sources, and publications that were not related to healthcare or regulation. Furthermore, reference lists of pivotal articles were manually reviewed to find other relevant sources.

The chosen literature was synthesised thematically under categories such as opportunities, challenges, ethical issues, and regulatory views to allow for a comprehensive understanding of AI/NLP implementations in regulatory documentation.

3. APPLICATIONS OF AI IN PHARMACEUTICAL REGULATORY DOCUMENTATION

AI technology, i.e., NLP technologies, is transforming the preparation of pharmaceutical regulatory submissions by saving time and increasing the accuracy and consistency of the submissions [3]. Regulatory dossiers are prepared in the Common Technical Document (CTD) format, which is harmonized worldwide by the ICH. The CTD consists of five modules: Module 1 has administrative data; Module 2 has abridgments of technical reports; Module 3 has quality data; Module 4 has nonclinical (safety) data; and Module 5 has clinical data. Module 3 is the most time-consuming and strictly reviewed by the regulatory authorities as it includes in-depth quality-related data of the drug [10]. As summarized in Table 1, AI tools are applied across various CTD sections to streamline documentation.

These computer-generated tools facilitate the transformation of unstructured text and images to neat tables and coherent descriptions, enhancing document clarity and consistency. Automation lowers the level of manual effort, reduces human error, and speeds up dossier preparation, ultimately improving submission quality [11,12].

The use of NLP and AI in regulatory text illustrates how technological advancements can improve efficiency and accuracy in compliance activities. In the case of pharmacovigilance, for example, NLP was utilized to triage individual case safety reports by identifying salient information in unstructured text. Netherlands Pharmacovigilance Centre Lareb created an NLP-driven prediction model that picked up on serious adverse drug reactions more effectively, decreasing the workload manually and enhancing the timeliness of safety reporting [13].

Likewise, AI-based solutions have been utilized to automate the preparation of electronic common technical document (eCTD) submissions. Peer AI showed how stability protocols and batch records in raw Chemistry, Manufacturing, and Controls (CMCs) documentation could be converted into submission-ready drafts through AI-based automation. This reduced errors by a great extent, boosted efficiency, and expedited regulatory compliance processes [14].

Table 1. The table below highlights how AI tools are applied across various sections of the CTD to streamline documentation.

NLP tool	CTD section	Action
NLP text extraction	Module 1: Administrative info	Extracts and organizes key data from large documents
Machine learning	Module 3: Quality data	Assesses and organizes batch records and stability data
NLP compliance check	Module 3: Analytical data	Cross-references submission with global regulatory guidelines
Predictive models	Module 4: Stability	Predicts long-term stability based on current and historical data
NLP formatting tool	Module 5: Effectiveness data	Standardizes documents, ensuring uniformity in style and format across regions

### 3.1. AI enhancements in pivotal module 3 of regulatory submissions

#### Section 3.2.P.3 – Manufacturing process:

AI tools can extract data from batch manufacturing records to generate automatic, accurate, standardized flowcharts and detailed process descriptions, reducing workload and errors [1].

#### Section 3.2.P.4 – Batch analyses:

AI-driven tools assist in gathering, validating, and consolidating batch analysis data in compliance with regulations into tables and graphical presentations, and improve data correctness and approval prospects for dossiers [1].

#### Section 3.2.P.5 – Analytical procedures and validation reports:

AI can generate and automate validation reports for analytical procedures such as drug specifications and impurity profiling to meet regulatory requirements [15].

#### Section 3.2.P.8 – Stability data:

Machine learning (ML) algorithms analyze history and real-time stability data to predict shelf life and optimal storage conditions in an attempt to make it easier to create stability summaries [1].

AI also facilitates regulatory compliance itself by being integrated into agency regulations such as the ICH, U.S. FDA, and EMA, so that tools for validation can identify missing data or compliance discrepancies in real time. It maintains submitted documents up to date according to evolving regulatory needs [2]. For global pharmaceutical corporations, multilingual AI-driven translation models are essential as they can translate intricate technical and legal terminology properly, ensure consistency with regulatory terminology across geographies and languages, and therefore, reduce the risk of error in multijurisdictional submissions. Aside from these benefits, there are still challenges related to the use of AI in pharmaceutical regulatory documents. Maintaining data integrity in terms of ALCOA+++ principles, which include Complete, Consistent, Enduring, and Available to the original Attributable, Legible, Contemporaneous, Original, and Accurate requirements, is required to maintain regulatory trust [16].

Ethical concerns like preventing bias in AI algorithms and understanding AI-based decisions are necessary. In addition, compliance with data privacy laws like the General Data Protection Regulation (GDPR), which regulates data about individuals in the European Union, adds complexity when dealing with sensitive information [17]. Legal concerns like liability and ownership of intellectual property rights of artificially produced content must also be addressed. Finally, there are technical concerns such as AI model validation, explainability, and conformity to the current regulatory environment, which are matters that should be worked out further [18].

It must be noted that though AI application in CTD modules is high-profile, the preparation of dossiers is not the sole application of the software. AI finds application in other aspects of regulatory operations, such as pharmacovigilance, labeling

compliance, and post-marketing surveillance, extending its portfolio in pharmaceutical regulatory affairs [19].

## 4. DATA INTEGRITY AND STANDARDIZATION CHALLENGES IN REGULATORY SUBMISSIONS

Data integrity is a key element in pharmaceutical regulatory materials, particularly when preparing regulatory dossiers such as the Common Technical Document (CTD). Data integrity maintains the accuracy, completeness, consistency, and reliability of the data throughout its life cycle. Good-quality data integrity is the foundation for regulatory submissions to ensure that the drug product complies with strict regulations imposed by regulatory authorities such as the USFDA, EMA, and PMDA. Inconsistency in the data provided can lead to delayed approval and rejection, and consequently, affect market access and patient safety [20].

For maintaining data integrity, the pharmaceutical companies are asked to follow the ALCOA guidelines: Attributable, Legible, Contemporaneous, Original, and Accurate. These were initially put forth by the US FDA and subsequently applied worldwide as the benchmark for data quality and integrity for good documentation practices [21]. The implementation of these guidelines ensures that data is documented and kept in a way that maintains its integrity and acceptability to the regulatory agencies.

### 4.1. ALCOA and its significance in regulatory documentation

**Attributable** requires that all the data be traceable to their origin. The individual recording the data should be identifiable, and data changes should be readable. AI tools like digital audit trail management systems and version control systems aid in ensuring attribution by attributing data entries to identifiable individuals or processes [22].

**Legible** applies to document legibility and written readability. Legible information in regulatory filings prevents misinterpretation. Legibility can be enhanced through NLP algorithms by implementing language standardization, eliminating typographical errors, and ensuring format consistency across dossiers [23].

**Contemporaneous** guarantees that information is documented at the time the information is being created. It is particularly vital during pharmaceutical manufacturing and clinical trials. Laboratory Information Management System that are driven by AI Laboratory Information Management Systems pick up real-time data from processes and equipment automatically, thus achieving contemporaneous status.

**Original** emphasizes that the original record should be maintained, as opposed to duplicated or transcribed data. In computer systems, this is facilitated by blockchain technologies that offer unalterable and stamped records that guarantee data originality and authenticity.

**Accurate** demands to be an actual representation of the facts observed. Inconsistencies or errors can lead to noncompliance. AI software with in-built validation rules and ML models can check inputs against pre-defined parameters to detect inconsistencies and ensure data accuracy before submission. As illustrated in Figure 1, AI tools support ALCOA principles to maintain data integrity.



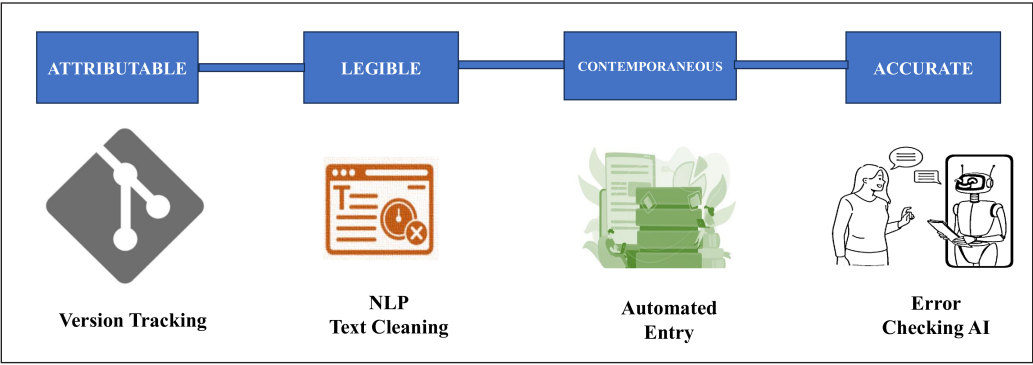


Figure 1. ALCOA and AI for data integrity.

4.2. Addressing challenges of standardization across regulatory bodies

One of the biggest regulatory filing challenges is the nonstandardization of global regulatory agencies. Every agency has its own submission format, organization, and data requirements. For example, stability study reporting should be done on the basis of ICH climatic zones, geographically diverse [24]. Regulatory affairs professionals can use AI tools to simplify this complexity by automating dossier formats, region-synchronizing reports, and cross-referencing agency-specific guidelines.

While there are benefits, the AI systems themselves need to be validated to avoid recasting algorithmic shortcomings or biases that could have an impact on compliance. Therefore, stringent system validation procedures need to be in place, for example, according to good automated manufacturing practice 5 (GAMP 5) guidelines, so that AI tools deployed in regulatory documents are compliant and reliable.

4.3. Ensuring real-time and unique data capture

In clinical trials and pharma manufacturing, data should be captured in real-time to avoid retrospective input, which compromises data quality. AI systems automate this process and generate timestamps, reducing the likelihood of data tampering. With digital tools increasingly becoming a part of our daily lives, there is also the risk of entering duplicate or derived data. “Derived data” refers to second-level data generated by calculations, transformations, or extrapolations rather than written down directly from measurement or observation.

To counter these threats, firms are increasingly adopting blockchain-based digital ledgers that enable secure, verifiable records of all document changes. This technique not only preserves originality but also enables end-to-end tracing of changes, which is especially useful in auditing [25].

To counter these risks, businesses are now using blockchain-based digital ledgers that give secure, verifiable proofs of all document changes. This approach does not just preserve originality but also ensures complete tracing of changes, which proves to be especially useful during audits [25]. Table 2 illustrates how AI tools support each of the ALCOA principles to ensure data integrity in regulatory submissions.

Table 2. ALCOA and AI support.

ALCOA principle	What it means	AI support
Attributable	Data traceability	Version tracking
Legible	Clear to read	NLP text cleaning
Contemporaneous	Real-time logging	Automated entry
Original	Primary data	Blockchain backup
Accurate	True and correct	Error checking AI

5. CROSS-CUTTING CHALLENGES IN APPLYING AI/ NLP TO REGULATORY DOCUMENTATION

5.1. Technical challenges

Technically, one of the most elemental challenges is the guarantee of data quality and standardization. Regulatory documents come in various forms (e.g., scanned PDFs, CTD modules, and region-specific templates), which causes it to be arduous for NLP systems to maintain consistency. Poorly standardized inputs undermine ALCOA principles’ legibility, attribution, and accuracy, and accordingly undermine data integrity in regulatory filings [20,21].

Following this, the algorithmic constraints of AI models add another dimension of challenges. Most AI/NLP tools are “black boxes” in which decision-making paths cannot be easily examined. This lack of transparency lowers the degree of confidence among regulators and slows adoption. The newer field of Explainable AI (XAI) attempts to solve this by offering interpretable results that can enhance regulatory confidence [19,26]. A second issue of relevance is model drift and reproducibility: models trained on one corpus might underperform when regulatory templates or guidelines shift, producing variable results [5,6,12].

Equally critical are the requirements for validation under regulatory quality standards. Because AI/NLP systems have direct access to regulatory submissions, they need to meet GAMP 5 as well as more general GxP rules. Risk-based validation, strong change-control processes, and performance qualifications are necessary to comply [16,22].

As encapsulated in the Technical category in Table 3, the challenges cut across several regulatory areas, from quality-related documentation to clinical submission and pharmacovigilance reporting. Therefore, while AI/NLP have the potential to speed up

data handling and enhance documentation clarity, their safe use is dependent on overcoming continued hurdles of standardization, transparency, validation, and reproducibility.

## 5.2. Regulatory challenges

Regulatory, privacy, and data protection remain a primary concern. The use of patient-level health data triggers the GDPR in the EU [17,27,28] and the Health Insurance Portability and Accountability Act (HIPAA) in the US [29,30]. Both frameworks call for robust measures such as de-identification, lawful processing, and controlled cross-border data transfer. Failure to comply not only risks patient privacy but can also render submissions invalid.

The other important regulatory hurdle is the lack of harmonized international standards for AI products. Though the FDA has published its AI/ML Action Plan and white papers [5,31,32] and the EMA has issued reflection papers on AI applications throughout the product life cycle [6,33], regulatory frameworks for adaptive learning algorithms and post-market maintenance are still in the process of development. Requirements for submission also vary significantly between agencies, leaving industry players uncertain.

Pharmacovigilance regulatory bodies, for instance, mandate adherence to Good Pharmacovigilance Practices. In using AI or NLP for case processing or safety signal detection, such technologies need to be subjected to heavy validation, ongoing monitoring, and by trained staff to ascertain patient safety [20].

Another complexity is caused by nonstandard submission formats by geography. Though the ICH Common Technical Document (CTD) sets the global standard, regional differences occur i.e., variations in eCTD implementation necessities, terminological preferences, and national appendices. These nonuniformities add to the regulatory professional's workload and point to the necessity for harmonization that is long overdue [24,34].

Together, these problems, increasing governance attempts on the one hand and ongoing fragmentation on the other, are captured under the Regulatory domain in Table 3.

## 5.3. Ethical challenges

The use of AI/NLP in regulatory documentation presents several ethical challenges. Perhaps the most urgent problem is biased training data. If training data are deficient, unrepresentative, or skewed towards specific populations, AI-generated output can be biased, resulting in discriminatory or

untrustworthy regulatory judgments [12,35]. A second concern is accountability. When an AI system produces an incorrect result, it is unknown whether the developer, the deploying pharma company, or the relying regulator is at fault [3,10,19]. The accountability gap generates legal and ethical ambiguity within the regulatory sphere. Ethics in patient data is also essential. Secondary use of clinical or health data must have lawful bases under GDPR or HIPAA, with robust protections to avoid re-identification [17,29,30]. In the absence of rigorous regulation, the application of AI in regulatory environments threatens patient rights. Last, intellectual property and ownership issues are emerging. Application of proprietary corpora, models, and algorithms calls into question licensing, derived work ownership, and fair access. International bodies like the World Intellectual Property Organization (WIPO) are already considering these issues in the context of AI innovation [36]. As described in the "Ethical" domain of Table 3, such difficulties stand in counterpoint to the potential for AI to improve patient safety through automated pharmacovigilance, and demonstrate the twin sides of ethical threats and opportunities.

## 5.4. Operational challenges

At an operational level, regulators and firms struggle to integrate AI into established quality frameworks. Ongoing model and dataset updates put traditional change-control processes under pressure, developed for static systems and not for adaptive algorithms [16,34]. Updates will unintentionally violate compliance without effective governance. There are also risks from skills and governance gaps. Regulatory affairs teams tend to be deficient in AI literacy, whereas data scientists tend to be deficient in regulatory capabilities. This disparity highlights the importance of training and cross-disciplinary governance structures to control AI in a responsible manner [16,37]. An additional operational issue is cybersecurity. Regulatory filings harbor extremely confidential intellectual property. AI/NLP systems that process such content should be protected with encryption, access restriction, and ongoing surveillance, in accordance with GxP data integrity requirements [16,21,38]. Finally, audit preparedness and long-term record storage are problematic in the AI setting. Regulators can insist that firms replicate AI-supported decisions even years afterward. This requires permanent storage, tamper-evident logs, and full version-pinning of models and data [16,21,38]. In their absence, it becomes virtually impossible to reconstruct regulatory submissions. Such difficulties, encompassed in the "Operational" domain of Table 3, indicate that although AI holds

**Table 3.** Comparative framework of AI/NLP tools in regulatory applications and their associated risks.

Application domain	Example AI/NLP tools	Primary use in regulation	Associated risks/challenges
Regulatory writing	NLP text extraction, NLP formatting tools	Automating drafting, extracting key information, standardizing submissions	Risk of misinterpretation errors; unexplainability, risk of bias in output content [19,26]
Safety reporting/ Pharmacovigilance	NLP signal detection, ML classifiers	Automated case processing, detection of adverse events, signal handling	False negatives/positives; bias in the training dataset; gaps in accountability if there is misclassification [20,35]
Submission management	Compliance check algorithms, predictive models	Cross-referencing guidelines, predicting stability, handling formats	GxP validation required; model drift with template updates; risk of cybersecurity breaches [16,22,32]
Data integrity & audit	Blockchain, audit trail AI	Ensuring traceability, version control, authenticity of submissions	Auditability issues; integration with existing systems; regulatory acceptance not consistent [16,21,38]

out the promise of efficiency benefits, ongoing compliance and audit preparedness are a formidable obstacle.

## 6. REGIONAL VARIATION IN REGULATORY REQUIREMENTS

One of the primary difficulties for AI deployment in regulatory matters around the world is the presence of diversity in legislative models. Various locations and regions have varying regulations and, consequently, varying constraints. Therefore, region-wise variations as far as regulations are concerned present particular challenges for worldwide businesses that try to incorporate and govern AI.

There are legislation on the one side like those of the European Union, which are not only holistic in nature but also imposes stringent restrictions on the application of AI, especially which are classified as high-risk. Then there are legislative regimes of many Asian nations which emphasize innovation and economic development and, therefore, have a comparatively business-friendly inclination. It can thus be realized why the multinational corporations are having a lot of problems trying to operate in such varied and disparate regulatory environments [39]. The varying regulations result in unnecessary increases in operating costs and convoluted legal interactions since there is no single and global regulation of AI [39].

## 7. ETHICAL AND LEGAL CONSIDERATIONS IN PHARMACEUTICAL REGULATORY DOCUMENTATION

AI is revolutionizing regulatory documentation by improving efficiency, minimizing human errors, and supporting quicker decision-making. However, this integration also poses serious ethical and legal issues. These issues are especially important in the preparation of regulatory filings such as the Common Technical Document (CTD), Drug Master Files (DMFs), Investigational New Drug (IND) applications, and post-marketing surveillance reports, all of which contain sensitive proprietary or patient data [34].

One such high-profile ethical concern is algorithmic bias. If AI systems are trained on nonrepresentative data, notably demographic data in post-marketing safety monitoring, they could generate biased results or exclusion errors that adversely affect particular populations disproportionately [35]. For instance, adverse event predictions could overlook vulnerable groups because they are underrepresented in training data. The regulatory agencies like the U.S. FDA and EMA highlighted the importance of utilizing representative and validated datasets in order to ensure fairness in documentation and regulatory assessments [38]. Data protection and privacy are paramount legal issues. Regulatory filings often include extremely sensitive patient data as well as proprietary drug composition information. To illustrate, Module 3 of the CTD (Quality) or parts of the DMF hold confidential manufacturing and composition information. When AI machines take over the processing and analysis of such data, the likelihood of data breaches or unauthorized access is substantially greater. Regulatory systems such as the GDPR in the European Union [27] and the HIPAA in the United States [29] call for intense control of the use, storage, and exchange of personal information. They call for mechanisms such as encryption, access control,

audit trails, and anonymization to be incorporated into AI-based platforms used in regulatory documentation.

The second major issue is transparency. Most AI systems, especially deep learning models, are “black boxes” with low interpretability. This becomes an issue in high-risk documentation activities like safety signal detection or benefit-risk assessment in Module 5 (Clinical Study Reports). Regulatory bodies are now requiring XAI frameworks to provide transparency in AI-driven decisions [26]. For instance, the FDA AI/ML-Based SaMD Action Plan requires reproducible and transparent models to be used in regulations [16]. Similarly, the EMA’s Reflection Paper on the Use of AI in the Medicinal Product Lifecycle recommends that transparency should be provided in AI-based decision-making, particularly in documentation impacting approvals or labelling [2].

Legal ambiguity over authorship and responsibility in documents created by AI also poses severe threats. For example, when an AI program generates content in the CMC section of a new drug application, it is uncertain whether responsibility lies with the AI developer, the drug sponsor, or both. This issue becomes increasingly complex when addressing patent applications or labeling claims based on AI-generated information. In order to address this, it is suggested that there be regulatory guidelines clearly set to outline ownership and liability in such a situation. Pharmaceutical sponsors must have strong internal verification procedures to ensure the correctness and legalese-defensibility of AI-created content before submitting it to the regulatory agencies. There are global legal bodies such as the WIPO, which already have noted this gap and provided recommendations regarding AI and intellectual property rights, that can significantly contribute towards establishing integrated international standards [36].

Finally, for AI to realize its potential without compromising ethical or legal integrity, it has to be balanced and cooperative. Regulatory agencies, pharmaceutical companies, and developers of AI have to work together to ensure data integrity (ALCOA+ principles), intellectual property protection, and compliance with international data protection regulations. This ensures that AI systems enhance the credibility, transparency, and equity of pharmaceutical regulatory submissions and do not undermine them.

## 8. DISCUSSION

The inclusion of AI, and more specifically NLP, in pharmaceutical regulatory filings provides great benefits in enhancing efficiency, consistency, and accuracy of submissions like the Common Technical Document (CTD). Through automation of processes such as data extraction, formatting, and validation, particularly for Module 3, AI lessens human effort and reduces errors. But applying NLP in this space has some major challenges like traversing regionally specific regulation requirements, understanding highly technical scientific and legal vocabulary, and upholding data integrity on ALCOA+ principles. There are also concerns around ethical issues like algorithmic bias, transparency issues, and ambiguity around accountability for AI-created content, which threatens compliance with the regulator. AI use also necessitates proper compliance with data privacy regulations such as GDPR and HIPAA when dealing with confidential personal and proprietary



information. Although regulatory agencies such as the FDA and EMA have taken first steps towards offering guidelines for AI-driven systems, successful integration needs more concerted effort between industry players, regulatory agencies, and technology innovators, in addition to targeted training for regulatory professionals. Therefore, while its transformative power cannot be denied, the responsible and well-regulated application of AI to this area needs to accommodate innovation with reliability, transparency, and legal strength.

## 9. FUTURE DIRECTIONS AND RECOMMENDATIONS

AI can potentially revolutionize the pharmaceutical regulatory environment by improving efficiency, accuracy, and compliance in submissions. But seamless integration demands strong legal, regulatory, and ethical frameworks to tackle changing challenges around data security, transparency, and interpretability [2].

AI systems should incorporate mechanisms for continuous learning to remain aligned with evolving regulatory expectations set by agencies such as the FDA, EMA, and other global authorities. Real-time updating processes would enable synchrony of regulatory files like the Common Technical Document (CTD) and IND filings, minimizing errors and delays [33].

The FDA AI/ML Action Plan and following discussion documents prioritize transparency, data accuracy, and human review as conditions for ethical implementation [31,39]. Likewise, the EMA Reflection Paper follows a product lifecycle-based, risk-related approach, which highlights that AI tools throughout the product lifecycle must be aligned with the principles of transparency, traceability, and robustness [6,33].

In spite of this achievement, there are still operational and technological issues. NLP methods can introduce interpretation errors, such as in the mapping of pharmacovigilance terms, resulting in imprecision in safety signal detection. Cloud-based applications present additional concerns about localization of data, jurisdictional access, and traceability, which are not always covered in present regulatory frameworks [32].

Another issue is readiness in the workforce. Regulatory specialists tend to be lacking in digital competence and AI literacy, whereas data scientists lack knowledge of regulatory demands. Closing this skills gap requires specialized training and cross-domain governance systems [40]. However, AI-based document verification and automated communication platforms can minimize

administrative delays, colloquially called “red tapism,” and enhance regulatory review efficiency [37].

For the full potential of AI to be realized while the regulatory integrity remains protected, closer cooperation among global agencies, industry players, and technological developers is required. Best practices should include:

- Adoption of ALCOA+ principles (Attributable, Legible, Contemporaneous, Original, Accurate, Complete, Consistent, Enduring, and Available) in AI-based systems [21].
- Compliance with data protection regulations like GDPR [31] and HIPAA [30].
- Instituting ethical review panels to oversee AI usage in regulatory guidelines [36,40].

### 9.1. Opportunities and challenges

The opportunities and challenges of AI/NLP integration in regulatory documentation are synthesized in Table 4.

The integration of findings points to the fact that AI/NLP tools provide disruptive potential while at the same time presenting multifaceted risks. Technically, these technologies enhance submissions and the quality of documentation but are held back by concerns with validation, reproducibility, and obscurity of algorithms [16,19,21,22]. Regulatorially, efforts by the FDA and EMA reflect worldwide momentum toward organized regulation, but the lack of harmonized standards and the vagaries of adaptive AI still present formidable hurdles [5,6,17,27,28,31]. Ethically, AI offers enhanced pharmacovigilance and patient safety [20,35], but issues of dataset bias, accountability loopholes, and intellectual property issues still linger [10,19,36]. Operationally, AI can lower delays and increase efficiency [37], but there are issues in regards to workforce capacity, cybersecurity protection, and long-term auditability [16,38].

Overall, AI/NLP implementation within regulatory science represents both a challenge and an opportunity. The future is in getting a balance between technological advancement and the establishment of harmonized regulations, robust ethical guidance, and capacity development among regulators and industry players [5,6,31,32,36,37].

Furthermore, on the basis of recent global and regulatory guidance, unambiguous recommendations follow for the sustainable implementation of AI in regulatory science. Transparency, data reliability, and human control are highlighted by the U.S. FDA as guiding principles for the use of AI/ML in drug development [7]. The EMA also recommends the lifecycle

**Table 4.** Opportunities versus challenges of AI/NLP in regulatory documentation.

Theme	Opportunities	Challenges
Technical	Quicker processing of data; NLP enhances the clarity and consistency of text [23,37]	Lack of good data standardization; black-box models; GxP validation [16,19,22]
Regulatory	New frameworks (FDA AI action plan, EMA reflection paper) [5,6,31]	Shortage of worldwide harmonization; hazy routes for adaptive AI; compliance of data privacy [17,27,28]
Ethical	Enhanced patient safety through robotized pharmacovigilance [20,35]	Dataset bias; unclear responsibility; intellectual property issues [10,19,36]
Operational	Automation decreases delays and improves efficiency [37]	Skills gap among regulators; cybersecurity threats; auditability issues [16,38]

approach, proportionate to risk, so that AI tools can be traced, made robust, and aligned with changing medicinal product requirements [8]. Adding to these regulatory views, the WHO emphasizes the ethical aspect, calling for governance structures that respect patient safety, equity, and accountability in AI-based healthcare systems [9]. Taken together, these guidelines highlight the need for harmonized global structures that embed technical, regulatory, and ethical protections to achieve the full potential of AI while sustaining public confidence and regulatory integrity.

## 10. CONCLUSION

AI, and more specifically NLP, holds transformative value to revamp regulatory submissions through increased accuracy, consistency, and speed; though, its successful and ethical use is hindered by concerns regarding data integrity, ALCOA+ adherence, transparency of algorithms, and dispersed world regulatory requirements. To progress, regulators must define validation standards for NLP based on current GxP and GAMP 5 standards, whereas industry players must work towards the integration of AI into ICH guidelines to ensure harmonization. Developers and researchers must prioritize explainability, bias reduction, and reproducibility to enhance trust, and all stakeholders must enter into multistakeholder collaborations to ensure ethical regulation, cybersecurity protocols, and workforce training. Together, these steps can unlock the full potential of AI, streamlining regulatory procedures and speeding up access to safe and effective drugs without undermining integrity or compliance.

## 11. AUTHOR CONTRIBUTIONS

All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed to submit to the current journal; gave final approval of the version to be published; and agreed to be accountable for all aspects of the work. All the authors are eligible to be an author as per the International Committee of Medical Journal Editors (ICMJE) requirements/guidelines.

## 12. FINANCIAL SUPPORT

Open access funding was received from Manipal Academy of Higher Education (MAHE), Manipal. No additional funding was received for the preparation of this review article.

## 13. CONFLICTS OF INTEREST

The authors declare no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Two authors are affiliated with industry organizations; their contributions were made in a purely academic capacity.

## 14. ETHICAL APPROVALS

This study does not involve experiments on animals or human subjects.

## 15. DATA AVAILABILITY

All data analyzed in this review are from publicly available regulatory guidelines, peer-reviewed publications, and official databases, and are cited within the manuscript.

## 16. PUBLISHER'S NOTE

All claims expressed in this article are solely those of the authors and do not necessarily represent those of the publisher, the editors and the reviewers. This journal remains neutral with regard to jurisdictional claims in published institutional affiliation.

## REFERENCES

1. Ajmal CS, Yerram S, Abishek V, Nizam VPM, Aglave G, Patnam JD, et al. Innovative approaches in regulatory affairs: leveraging artificial intelligence and machine learning for efficient compliance and decision-making. *AAPS J.* [cited 2025 May 27]. Available from: <https://link.springer.com/article/10.1208/s12248-024-01006-5>
2. European Medicines Agency (EMA). Reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle. 2024. Available from: [https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf)
3. American Pharmaceutical Review. Advancing regulatory compliance with natural language processing. *American Pharmaceutical Review. The Review of American Pharmaceutical Business & Technology* [Internet]. [cited 2025 May 27]. Available from: <https://www.americanpharmaceuticalreview.com/Featured-Articles/607985-Advancing-Regulatory-Compliance-with-Natural-Language-Processing/>
4. Licari D, Comandè G. ITALIAN-LEGAL-BERT models for improving natural language processing tasks in the Italian legal domain. [cited 2025 May 27]. Available from: <https://www.sciencedirect.com/science/article/pii/S0267364923001188>
5. U.S. Food and Drug Administration (FDA). FDA releases artificial intelligence/machine learning action plan. FDA [Internet]. [cited 2025 May 27]. Available from: <https://www.fda.gov/news-events/press-announcements/fda-releases-artificial-intelligencemachine-learning-action-plan>
6. European Medicines Agency (EMA). Artificial intelligence. European Medicines Agency (EMA) [Internet]. [cited 2025 May 27]. Available from: <https://www.ema.europa.eu/en/about-us/how-we-work/data-regulation-big-data-other-sources/artificial-intelligence>
7. U.S. Food and Drug Administration. Artificial intelligence and machine learning in drug development: discussion paper and action plan [Internet]. Silver Spring (MD): FDA; 2023 [cited 2025 Aug 30]. Available from: <https://www.fda.gov/media/167973/download>
8. European Medicines Agency (EMA). Reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle [Internet]. Amsterdam: EMA; 2023 [cited 2025 Aug 30]. Available from: <https://www.ema.europa.eu/system/files/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle-en.pdf>
9. World Health Organization (WHO). Ethics and governance of artificial intelligence for health [Internet]. Geneva: WHO; 2023 [cited 2025 Aug 30]. Available from: <https://www.who.int/publications/i/item/9789240029200>
10. International Council for Harmonisation (ICH). M4 QCommon technical document for the registration of pharmaceuticals for human use – quality. Geneva: ICH; 2002.
11. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;25(1):44–56.
12. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* 2019;17:195. [cited 2025 May 27]. Available from: <https://bmcmmedicine.biomedcentral.com/articles/10.1186/s12916-019-1426-2>
13. Lieber T, Voskuil R, Mol P, Härmark L. Natural language processing for automated triage and prioritization of individual case safety reports. *Front Drug Saf Regul.* 2023 [cited 2025 Sep



- 4]. Available from: <https://www.frontiersin.org/articles/10.3389/fdsfr.2023.1120135/full>
14. Peer AI. Transforming regulatory bottlenecks into strategic accelerants. Peer AI Blog; 2025 [cited 2025 Sep 4]. Available from: <https://getpeer.ai/blog/ai-in-cmc-medical-writing-transforming-regulatory-bottlenecks-into-strategic-accelerants>
15. ICH Harmonised Tripartite Guideline. Guidance for Industry: M4Q – The CTD – Quality. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), 2003.
16. MHRA (UK Medicines and Healthcare Products Regulatory Agency). GXP Data Integrity Guidance and Definitions [Internet]. 2018 [cited 2025 May 27]. Available from: <https://www.gov.uk/government/publications/guidance-on-gxp-data-integrity>
17. General data protection regulation (GDPR). EUR-Lex [Internet]. [cited 2025 May 27]. Available from: [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM%3A310401\\_2](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM%3A310401_2)
18. Chatila R, Havens JC. The IEEE global initiative on ethics of autonomous and intelligent systems. In: Aldinhas Ferreira MI, Silva Sequeira J, Singh Virk G, Tokhi MO, E. Kadar E, editors. Robotics and well-being [Internet]. Cham: Springer International Publishing; 2019 [cited 2025 May 27]. pp. 11–6. (Intelligent Systems, Control and Automation: Science and Engineering; Vol. 95). Available from: [http://link.springer.com/10.1007/978-3-030-12524-0\\_2](http://link.springer.com/10.1007/978-3-030-12524-0_2)
19. Ribeiro MT, Singh S, Guestrin C. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access. 2018;6:52138–60. doi: <https://doi.org/10.1109/ACCESS.2018.2870052>
20. European Medicines Agency. Good pharmacovigilance practices (GVP). European Medicines Agency (EMA) [Internet]. 2025 [cited 2025 May 27]. Available from: <https://www.ema.europa.eu/en/human-regulatory-overview/post-authorisation/pharmacovigilance-post-authorisation/good-pharmacovigilance-practices-gvp>
21. Purdie FP. Data integrity and compliance with CGMP guidance for industry. U.S. Food and Drug Administration. 2017. Available from: <https://www.fda.gov/files/drugs/published/Data-Integrity-and-Compliance-With-Current-Good-Manufacturing-Practice-Guidance-for-Industry.pdf>
22. International Society for Pharmaceutical Engineering. GAMP 5 a risk based approach to a risk-based approach to compliant GxP compliant GxP computerized systems [Internet]. 2008. [cited 2025 May 27]. Available from: [https://www.academia.edu/36720108/GAMP\\_5\\_A\\_Risk\\_Based\\_Approach\\_to\\_A\\_Risk\\_Based\\_Approach\\_to\\_Compliant\\_GxP\\_Compliant\\_GxP\\_Computerized\\_Systems](https://www.academia.edu/36720108/GAMP_5_A_Risk_Based_Approach_to_A_Risk_Based_Approach_to_Compliant_GxP_Compliant_GxP_Computerized_Systems)
23. Wu Y, Liu Z, Wu L, Chen M, Tong W. BERT-based natural language processing of drug labeling documents: a case study for classifying drug-induced liver injury risk. Front Artif Intell. 2021;4:729834. [cited 2025 May 27]. Available from: [https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2021.729834/full?utm\\_source=chatgpt.com](https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2021.729834/full?utm_source=chatgpt.com)
24. International Council for Harmonisation. ICH Q1A (R2) Stability testing of new drug substances and drug products - Scientific guideline. European Medicines Agency (EMA) [Internet]. 2003. [cited 2025 May 27]. Available from: <https://www.ema.europa.eu/en/ich-q1a-r2-stability-testing-new-drug-substances-drug-products-scientific-guideline>
25. Kuo TT, Kim HE, Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications. J Am Med Inform Assoc. 2017;24(6):1211–20. [cited 2025 May 27]. Available from: <https://academic.oup.com/jamia/article/24/6/1211/4108087?login=true>
26. Ribeiro MT, Singh S, Guestrin C. “Why Should I Trust You?” Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. 2016 [cited 2025 May 27]. Available from: <https://dl.acm.org/doi/10.1145/2939672.2939778>
27. European Parliament and Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Available from: <http://data.europa.eu/eli/reg/2016/679/oj>
28. Publications Office of the EU. General data protection regulation (GDPR) [Internet]. 2018 [cited 2025 May 27]. Available from: <https://op.europa.eu/en/publication-detail/-/publication/e899a0de-be12-405c-84ae-34e8a22d6acd/language-en>
29. Edemekong PF, Annamaraju P, Afzal M, Haydel MJ. Health Insurance Portability and Accountability Act (HIPAA) Compliance. StatPearls Publishing; 2024 [cited 2025 May 27]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK500019/>
30. Centers for Disease Control and Prevention (CDC). Health Insurance Portability and Accountability Act of 1996 (HIPAA) [Internet]. 1996 [cited 2025 May 27]. Available from: <https://www.cdc.gov/php/php/resources/health-insurance-portability-and-accountability-act-of-1996-hipaa.html>
31. U.S. FDA. Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan; 2021. Available from: <https://www.fda.gov/media/145022/download>
32. U.S. FDA. Using artificial intelligence and machine learning in the development of drug and biological products. Discussion Paper and Request for Feedback. 2023 Available from: <https://www.fda.gov/media/167973/download>
33. Covington Digital Health. EMA releases reflection paper on AI/ML in the medicinal product lifecycle [Internet]. 2023 [cited 2025 May 27]. Available from: <https://www.covingtondigitalhealth.com/2023/07/ema-releases-reflection-paper-on-ai-ml-in-the-medicinal-product-lifecycle/>
34. International Council for Harmonisation. ICH M4 Common technical document (CTD) for the registration of pharmaceuticals for human use - organisation of CTD - Scientific guideline. European Medicines Agency (EMA) [Internet]. 2003. [cited 2025 May 27]. Available from: <https://www.ema.europa.eu/en/ich-m4-common-technical-document-ctd-registration-pharmaceuticals-human-use-organisation-ctd-scientific-guideline>
35. Obermeyer Z, Powers BW, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447–53. doi: <https://doi.org/10.1126/science.aax2342>
36. World Intellectual Property Organization (WIPO). Revised Issues Paper on Intellectual Property Policy and Artificial Intelligence [Internet]. 2023 [cited 2025 May 27]. Available from: [https://www.wipo.int/meetings/en/doc\\_details.jsp?doc\\_id=499504](https://www.wipo.int/meetings/en/doc_details.jsp?doc_id=499504)
37. Deloitte UK. AI risk and approaches to global regulatory compliance [Internet]. 2025 [cited 2025 May 27]. Available from: <https://www.deloitte.com/uk/en/Industries/technology/perspectives/ai-risk-and-approaches-to-global-regulatory-compliance.html>
38. European Medicines Agency. Guideline on computerized systems and electronic data in clinical trials. 2025. Available from: [https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-computerised-systems-electronic-data-clinical-trials\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-computerised-systems-electronic-data-clinical-trials_en.pdf)
39. NAVEX. The evolving AI regulatory landscape in Asia: what compliance leaders need to know. NAVEX. 2025 [cited 2025 May 27]. Available from: <https://www.navex.com/en-us/blog/article/the-evolving-ai-regulatory-landscape-in-asia-what-compliance-leaders-need-to-know/>
40. World Health Organization (WHO). Ethics and governance of artificial intelligence for health [Internet]. Geneva: WHO; 2023 [cited 2025 May 27]. Available from: <https://www.who.int/publications/i/item/9789240029200>

#### How to cite this article:

Pawar HB, Neelanshu, Bhat GM, Srikanth A, Muragundi PM. Challenges of using Artificial Intelligence (Natural Language Processing) tools for regulatory documentation. J Appl Pharm Sci. 2025. Article in Press. <http://doi.org/10.7324/JAPS.2026.267918>