CitiusTech

HEALTHCARE LIMITLESS

Generative AI paradigm shift in regulatory documentation

Table of Contents

Introduction

Regulatory and compliance documentation is one of the most critical components of the clinical development process, impacting every aspect of it. Its purpose is to provide evidence that clinical trials are conducted ethically and procedurally while ensuring the safety of the patients. However, developing and maintaining a large set of regulatory documents in adherence to ICH-GCP guidance, including GDRP (Good Documentation and Record Practices) and ALCOA principles (Attributable, Legible, Contemporaneous, Original, Accurate) is a complex and time-consuming task when dealing with a global patient population and multi-country conduct.

Managing the documentation's completeness, accuracy, and quality is a complex and labor-intensive function for the pharmaceutical industry. The ever-involving nature of clinical trials, such as complex protocol designs (i.e., adaptive trials, platform trials, decentralized trials), digital technology, and increased use of RWE, adds complexity in adhering to documentation compliance. As a result, pharma companies continue to spend more on regulatory affairs while an increasing number of observations on documentation gaps are being made.

From the below graph, it is evident that regulatory documentation management is still a current challenge in clinical operations.

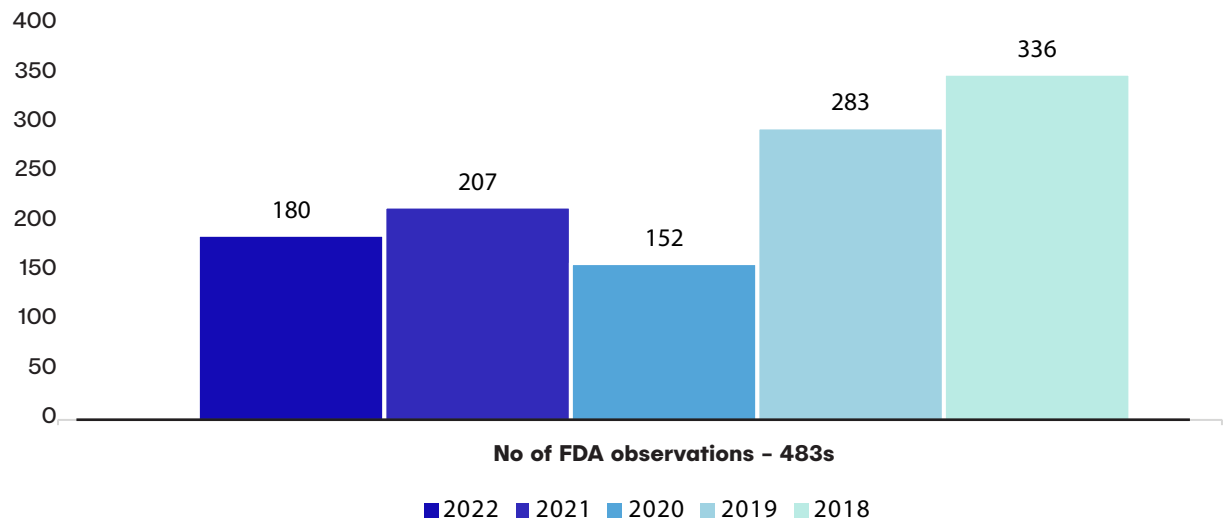

Figure 1 - FDA observations (Form 483s) from the last 5 years issued for regulatory documentation gaps

Current automation landscape in regulatory documentation

Regulatory document automation started with the emergence of NLP and NLG capabilities of AI/ML. Leveraging a template-driven NLP/NLG combined with procedural rules-based engines has streamlined the processes and has provided auto generation capabilities. However, the current approach has limited reusability across problem domains and requires redevelopment of the solution for each trial and subprocess level. Here are some examples of available automation capabilities for regulatory documents.

- **Quality checks and GDP deviations:**

Advanced AI can perform automated quality checks on documents, identifying inconsistencies, errors, and protocol deviations as per regulatory standards.

- **Document classification and organizing for audit:**

Machine Learning (ML) algorithms can categorize and classify documents, making it easier to organize and manage regulatory information—filing of TMF documents and identifying gaps like missing consent form post protocol amendment. ML can identify similar documents, help with version control, and identify outdated or duplicated content.

- **Semantic rule-based information extraction for automated compliance checking:**

Natural Language Processing (NLP) can review and extract relevant critical information from unstructured regulatory documents, such as drug names and dosages. NLP-powered search engines can provide accurate and contextually relevant results when querying large document repositories such as TMF.

- **Content generation and regulatory documents curation:**

Natural Language Generation (NLG) can generate automated reports summarizing standardized sections of regulatory documents, ensuring accuracy and consistency and curating documents such as protocol, ICF, etc., with the utmost precision.

Limitations of current automation technologies

It's important to note that while these available technologies present significant potential to enhance regulatory documentation, they also bring about certain challenges, such as:

- **Rule-based AI and lack of comprehension:**

Rule-based AI systems operate on a pre-defined set of rules and logic. However, when it comes to complex scenarios that deviate from these rules, such systems may struggle to adapt and learn from data. This is because they lack comprehension of the interrelation between documents or concepts and may fail to understand the proper context. E.g., Any change in Protocol w.r.t patient lab test also impacts the informed consent form.

- **Domain specificity:**

Clinical trial documents can differ as per Therapeutic area and Indication. This means an AI model trained on general data may produce inaccurate results. Additionally, it can be challenging to maintain and update the model to meet specific requirements that vary from trial to trial.

- **Quality of training data:**

All automation technologies are heavily dependent on the quality of training data; if the model is trained on biased data, it may lead to judgemental errors.

- **Creation of new content:**

It is challenging to create entirely new content that is beyond the scope of existing data.

- **Data privacy and security:**

Insecure models may pose a risk of data breaches.

- **Adaptability:**

There needs to be more adaptable automation to deal with the ever-evolving complex demands of regulators and the pharma industry.

Unleashing the potential of generative AI in regulatory documentation

Traditional AI operates based on predefined rules and historical data patterns. In contrast, the generative AI approach goes beyond this limitation and creates entirely new content resembling human cognition.

With the emergence of generative AI technologies and Large Language Models (LLMs), cognition as a service can be provided at scale, making it possible to automate regulatory documentation generation, management, and maintenance. This automation can help improve clinical trial operations significantly.

Generative AI can leapfrog the generation of regulatory documents with its ability to process vast amounts of unstructured data and advanced learning capabilities. Moreover, its cognitive and contextual comprehension abilities can be further improved over time, releasing faster and more accurate results. Additionally, potential inconsistencies can be flagged, and errors are tracked, leading to more complete and accurate documents in clinical operations.

Furthermore, generative AI can enable personalized content creation, including protocol creation, multi-lingual informed consent forms, monitoring reports as per the study, therapeutic area, and indication.

Market size

The market size of the global generative AI in Life Sciences was estimated at **USD 160.23 million in 2022,** and it is expected to hit around **USD 1,059,38 million by 2032,** growing at a **CAGR of 20.78%** during the forecast period from 2023 to 2032. The graph below indicates that generative AI has exponential growth and economic potential in various sectors of Life Sciences in the future.

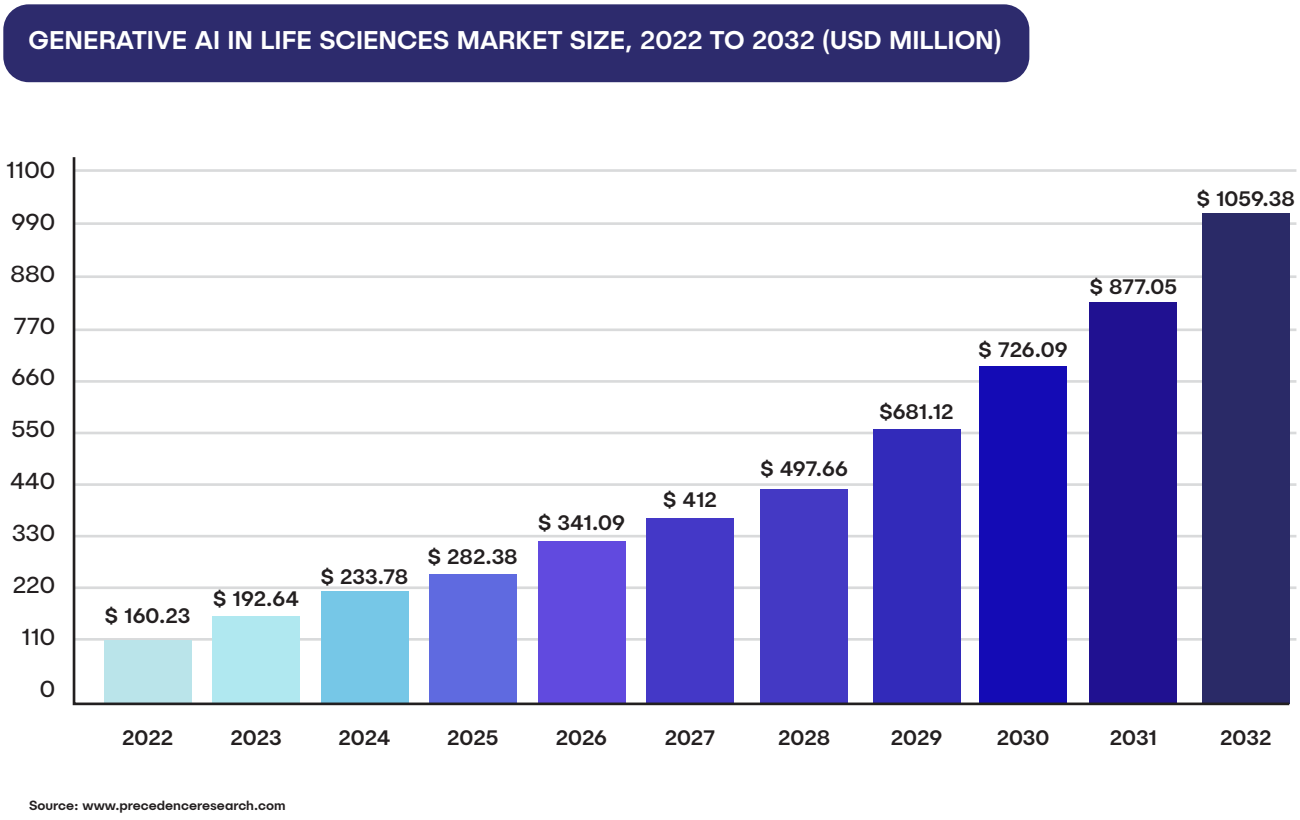

Figure 2 - Generative AI in Life Sciences market

High-level generative AI implementation process – An overview

One of the breakthroughs is the generative AI capability to leverage different learning approaches, such as Supervised, Un, or Semi-supervised training models, which can perform multiple tasks. Unsupervised learning is applied to tasks such as document clustering, text summarization, anomaly detection, semantic analysis, etc., whereas supervised learning helps in document classification, named entity recognition, language translation, etc.

On high-level, below are the steps that show how generative AI implementation can be done:

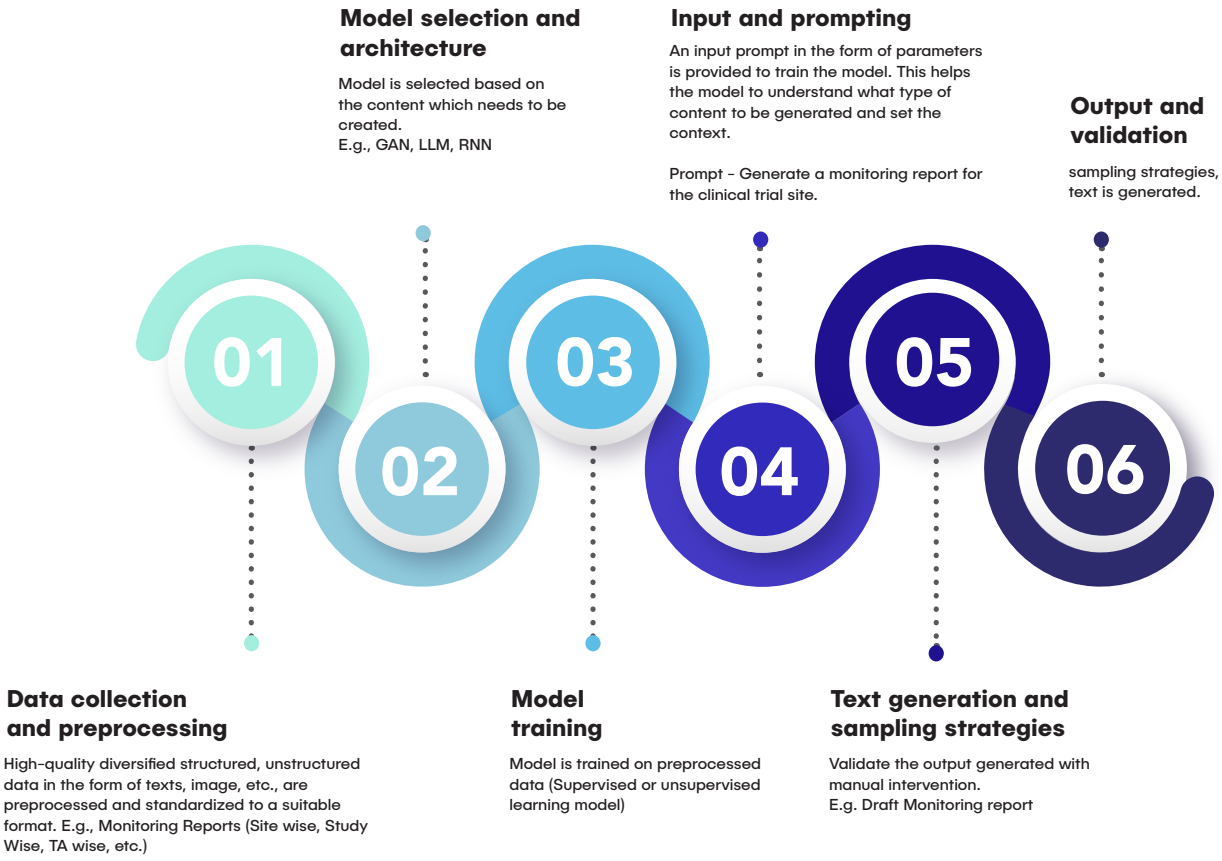

Figure 3 - Gen AI implementation steps

Heat map

The above process can be implemented for content creation and updating of various regulatory documentation. Now let us understand which are the regulatory documents where generative AI can be leveraged and augmented. The image below shows the same.

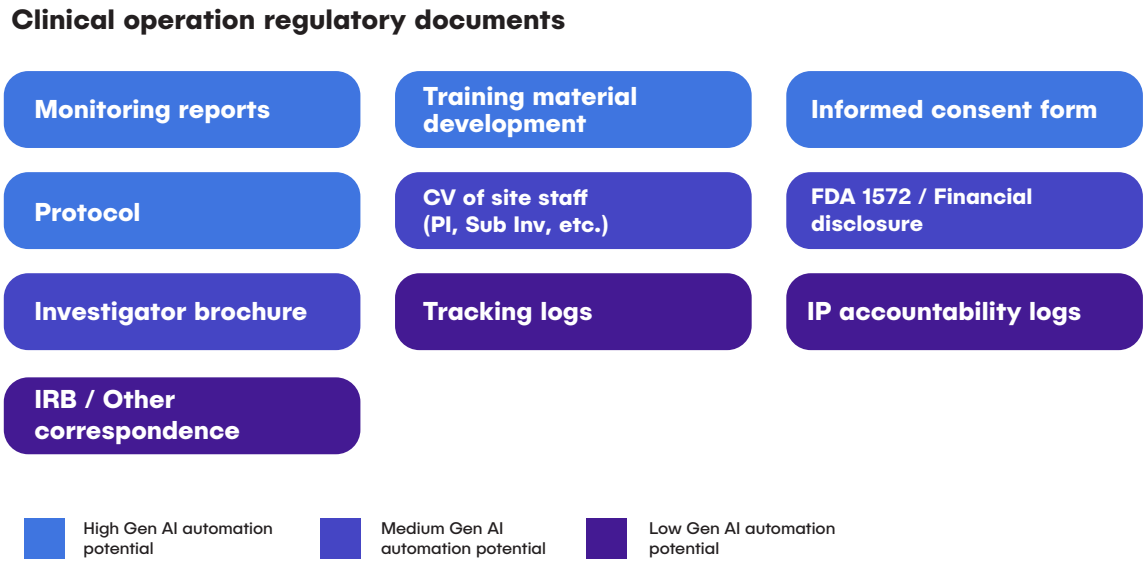

Figure 4 - Leveraging Gen AI for clinical operations regulatory documents

Transforming regulatory document management: Potential generative AI use cases

In accordance with the ICH E6 Good Clinical Practice guidelines, the following list of regulatory documents may potentially utilize generative AI in clinical operations.

Monitoring reports

Market need

- Average phase III trial monitors spend up to 18 hours on this task alone.
- With many top 20 pharma averaging over 40,000 yearly monitoring trip reports, 25%–30% of total clinical trial costs are attributed to site monitoring.
- Many CRAs work across 10+ systems, often siloed, adding effort and complexity to an already challenging job.

Solution

- Based on data available in various clinical systems such as CTMS, IVRS, CDM, TMF, etc. Gen AI can create and pre-fill the monitoring report before the monitor visits the site.

Business value

- Reduce monitor time in preparing the reports and focus on on-site monitoring.

Multi-language informed consent form creation

Market need

- Generation of participant's understandable clinical trial details as per regulatory performa.

Solution

- Create multilanguage ICF creation as per therapeutic area and indication-based protocol.

Business value

- Better patient compliance as per industry standards.
- Patient dropouts are high due to lack of information and confidence.

Protocol authoring

Market need

- Faster development of clinical protocol - Currently, it takes at least six weeks to write a simple early-phase protocol. It can take up to 18 months to develop a prospective clinical protocol.

Solution

- Protocol authoring can be accelerated and optimized by using Gen AI.
- It can scan a sizeable amount of scientific literature, previous protocol histories, databases, and reports and provide endpoints, sample size, treatment arms, statistical considerations, analysis, and procedures that can be followed.

Business value

- Getting the foundation right and reducing protocol amendments and violations, leading to less administrative burden and rework.

Creating training material for PI and site staff

Market need

- Speedy and accurate training material creation with limited resources and ensuring ICH-GCP compliance.

Solution

- Generative AI can create TA and indication-based protocol training material for all study and site personnel staff.

Business value

- Creating effective training content with minimal manual intervention.

Standardized CV creation of PI and site staff

Market need

- CV of Site staff to be formatted in standardized sponsor format.

Solution

- CV in a study is always created in a standard format to keep consistency across the sites. Gen AI can scan CV documents in various formats and create the CV in the standard formats. CV is one of the most critical regulatory documents

Business value

- Compliant ready documents.

FDA 1572 and financial disclosure form

Market need

- Adherence to regulatory standards.

Solution

- Gen AI can auto-create FDA 1572 and financial disclosure (3454 and 3455) forms based on the CV received and CTMS data.

Business value

- Reduce repetitive tasks and error-free documents.

Investigator Brochure (IB) creation

Market need

- Clear, accurate document generation in less time.

Solution

- Gen AI can create certain sections in IB by training the model with historical IB content specific to TA, Indication, and Investigational Product.

Business value

- Reduce deviation.

Short & long-term benefits

Implementing Gen AI in regulatory documentation holds immense potential and delivers greater value.

Short term

- Cycle reduction time in creating content of regulatory documents.
- Removal of redundant and repetitive tasks required in content generation.
- Increase in process efficiency for the creation and maintenance of documents.
- Accurate, Consistent, quality-driven document generation.

Long term

- Real-time audit and inspection ready documents.
- Compliant driven process across document generation, review, management, and maintenance.
- Cost effective.
- Optimize resource management in the document lifecycle.

Challenges

While generative AI has numerous benefits, the author here wants the audience to be cognizance of the limitations such as:

1. Data quality, lineage, provenance:

High quality driven diversified data is required for training; ensure data integrity before training the model. The biased data model cannot give the output accurately, leading to incorrect content creation.

2. Scalability:

The model should be able to process a high volume of data with consistent accuracy and performance.

3. Integration:

Ensure the generative AI model can be leveraged in existing enterprise systems.

4. Data privacy and security:

During training in the LLM model, one must ensure data privacy regulations and security that can lead to data breaches.

5. Transparency:

AI models, often termed as "Black Boxes," need to ensure transparency in the entire process of training the model.

Key takeaways

- Generative AI holds massive potential for transforming the creation and application of technology in the regulatory landscape.
- Simplify the regulatory documentation.
- Game changer in content creation with enhanced accuracy and productivity.

Conclusion

The journey of generative AI is inevitable and is knocking on every door of opportunity. It offers tremendous promise for the Life Science industry. Faster understanding, acceptance, and implementation of generative AI in documentation will reshape the future of clinical trials and be the go-to solution for regulatory documentation.

Authors

Chhaya Vyas,

Sr. Healthcare Consultant, CitiusTech

Ravinder Singh,

SVP, Head of Healthcare Consulting, CitiusTech

About CitiusTech

Our vision is to inspire new possibilities for the health ecosystem with technology and human ingenuity. At CitiusTech, we constantly strive to solve the industry's greatest challenges with technology, creativity, and agility. Together with the world's leading Healthcare and Lifesciences organizations and our partners, we aim to accelerate the transition to a human-first, sustainable, and digital healthcare ecosystem.

To learn more about this contact us at

GenAI@citiustech.com

Visit our website

www.citiustech.com/re-imagen-ai

© Copyright 2023 CitiusTech. All Rights Reserved.

CitiusTech