



2022

JAPAN

INTERCHANGE

13-14 JUNE | VIRTUAL EVENT

Use of Artificial Intelligence (ML/NLU/NLP/NLG) in Regulatory (Scientific) Documents Authoring

FARHA FEROZE



Meet the Speaker

Farha Feroze

Title: Product Manager

Organization: Symbiance Inc

Farha Feroze is a Product Manager at Symbiance. She has been in the path of Product Management and Business Analysis for a good few years. She is known and appreciated for her sharp acumen and logical mind that easily grasp new business concepts. Her current focus is in providing business solutions that use advanced technologies such as Machine Learning (ML), Natural Language Processing (NLP)& Natural Language Generation (NLG) backed by her experience in Life Sciences Software Applications. She also specializes in designing and documenting business models, rules, workflows, reports and process workflows.



Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*



Agenda

1. Background of CSR and automation
2. Process ad steps
3. Benefits of AI
4. Conclusion



Background of CSR and automation



What are we trying to solve?

Clinical Study Report

- Creating Clinical Study Report (CSR) is highly manual and time consuming.
- The content of CSR is constructed from Protocol, SAP, Safety narratives, in-text tables and medical writer's interpretation of study results.
- No automotive way to handle Manual Errors and multiple iterations

Why automation ?

- Significant manual effort can be saved using AI techniques so MW can focus on the Interpretation and discussion point.
- Effective use and reuse of the template for an organization or therapeutic area or compound.
- The ML prediction algorithm along with NLP/NLG is used to identify/effectively used to read/write the texts from other documents such as Protocol, SAP, In-text Table & etc.,
- The CSR Template will be created based on ICH –E3 guidelines
- The medical writers can focus more on discussion points and interpretation of study results.



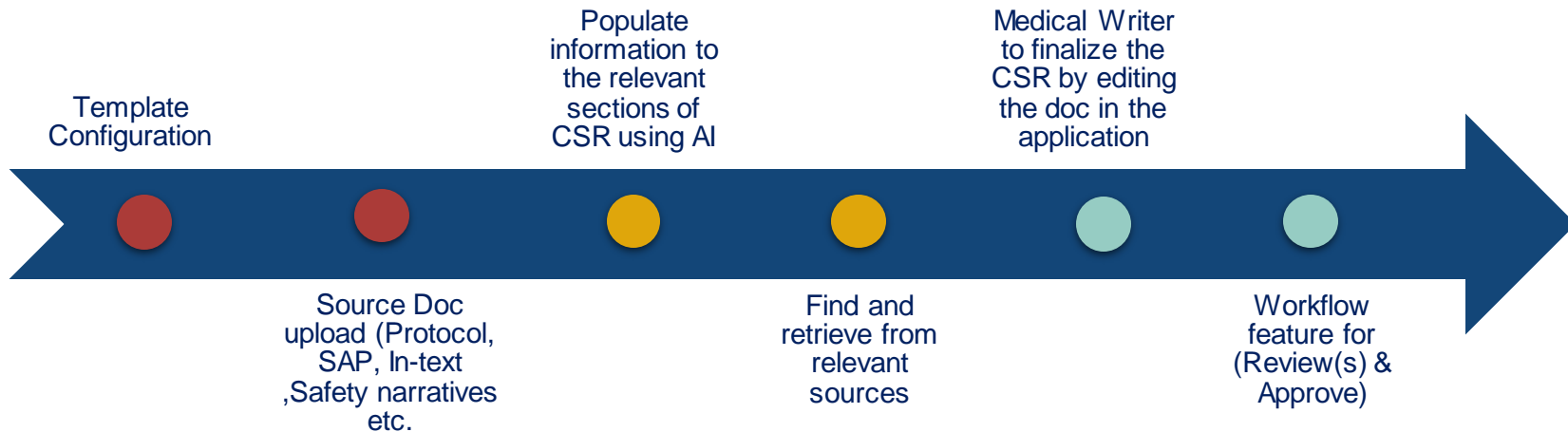
Blank CSR Template

1. TITLE PAGE
2. STUDY SYNOPSIS
3. TABLE OF CONTENTS
4. LIST OF ABBREVIATIONS & DEFINITION OF TERMS
5. ETHICS AND REGULATORY APPROVAL
6. INVESTIGATORS AND STUDY ADMINISTRATIVE STRUCTURE
7. INTRODUCTION
8. STUDY OBJECTIVES
9. INVESTIGATIONAL PLAN
10. STUDY POPULATION
11. RESULTS
12. SAFETY EVALUATION
13. DISCUSSION AND OVERALL CONCLUSIONS
14. TABLES, FIGURES AND GRAPHS
15. REFERENCES
16. APPENDICES



Process and Steps

Steps





Use of ML/NLP/ NLG

Prediction Accuracy :

- The text from PDF document such as Protocol/SAP are extracted using ML/NLP and the texts are understood by the system engine.

Identifying Individual Sections:

- ML model is used to predict the best matching content from various source documents(Protocol, SAP, In-text etc.,) for all the sections in a CSR.
- Named Entity Recognition (NER) which is a subprocess of NLP is used to identify the drug names, dosages, duration of drug, sponsor name and protocol number

Title Page & Synopsis:

- Customized ML Algorithm used in title page and study synopsis.

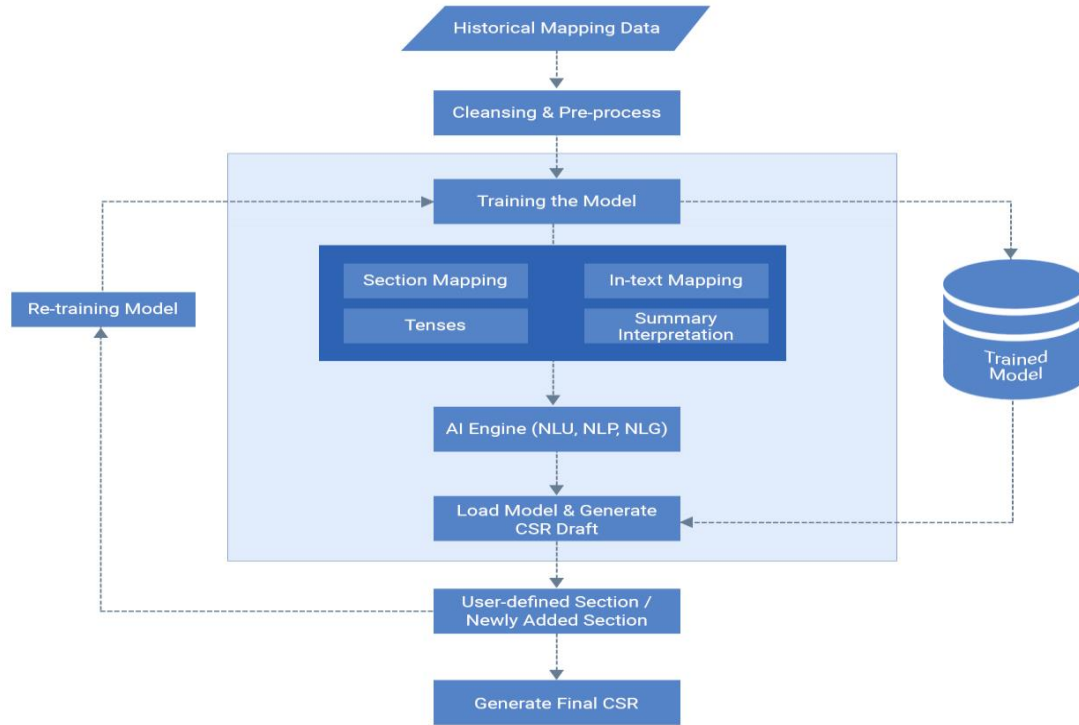
Deep Learning Model usage:

- The section 10 (study population), section 11 (results) and section 12(safety evaluation) is more of In-text tables that utilizes Deep Learning Model to find the best matching In-text tables.

Handling of Un-structured data

- Extracting data from source documents (PDF/Word, formatting and special symbols, subscript & superscripts and diagrams)
- Summarizing large paragraphs using NLU & NLG for CSR synopsis
- Reading the data from the unstructured TLFs & table details extraction of data and interpretation in simple English
- Tense conversion
- Section mapping challenges
- Optimization for performance in ML models

How AI is applied??



Application Edit Screen

The screenshot displays the ZYLiq application interface for editing a document. The top header includes the ZYLiq logo, the document name 'SYMBIANC... DEMO_SY...', and an 'Edit CSR' button. A navigation menu on the left lists document sections, with '1 TITLE PAGE' selected. The main editing area shows a 'Clinical Study Report' template with a table of key information. The table includes fields for Study Title, Investigational Drug Name, Indication, Protocol Number, Drug Development Phase, Study Initiation Date, and Study Completion Date. The 'Study Initiation Date' and 'Study Completion Date' rows are highlighted in yellow. The bottom status bar shows 'Page 1 of 109' and 'Zoom 100%'.

ZYLiq SYMBIANC... DEMO_SY... Edit CSR

File Home Insert Layout References Collaboration

Times new Rr 11 A A A A

toc 6 toc 7 Body Text **Title** List Paragraph Table Paragrap Heading 1

1 TITLE PAGE

2 STUDY SYNOPSIS

3 TABLE OF CONTENTS

4 LIST OF ABBREVIATIONS & DEFINIT...

5 ETHICS AND REGULATORY APPRO...

5.1 INDEPENDENT ETHICS COMMI...

5.2 ETHICAL CONDUCT OF THE ST...

5.3 PATIENT INFORMATION AND C...

6 INVESTIGATORS AND STUDY ADMI...

7 INTRODUCTION

7.1 BACKGROUND

7.2 RATIONALE FOR THE STUDY

8 STUDY OBJECTIVES

8.1 PRIMARY OBJECTIVE

8.2 SECONDARY OBJECTIVE

9 INVESTIGATIONAL PLAN

9.1 OVERALL STUDY DESIGN AND ...

9.1.1 STUDY TIMING

9.1.2 STUDY LOCATION

9.2 DISCUSSION OF STUDY DESIGN

9.3 SELECTION OF STUDY POPUL...

9.3.1 INCLUSION CRITERIA

9.3.2 EXCLUSION CRITERIA

DEVELOPER MODE

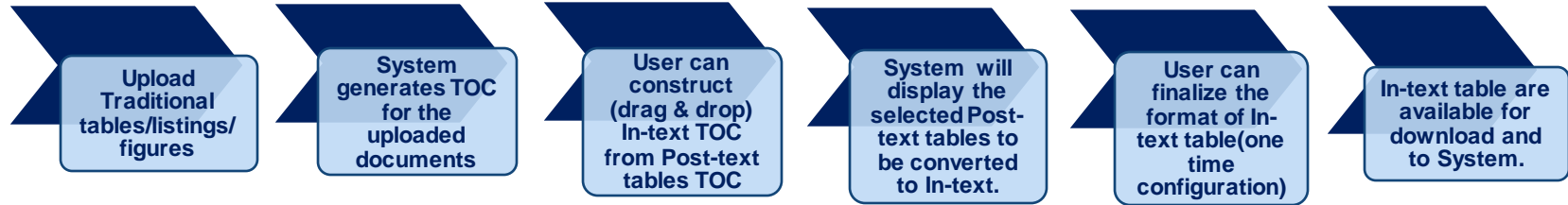
Symbiance SYMBIANC-21

1 TITLE PAGE

| Clinical Study Report | |
|----------------------------|--|
| Study Title: | An Open-Label, Multicenter Study with an Extension Phase to Evaluate the Safety, Tolerability, and Exposure-Efficacy Relationship of Test drug Oral Suspension when Administered as an Adjunctive Therapy in Pediatric Subjects (Age 4 to less than 12 years) with Inadequately Controlled Partial-Onset Seizures or Primary Generalized Tonic-Clonic Seizures |
| Investigational Drug Name: | EX2007/test drug |
| Indication: | Partial-Onset Seizures or Primary Generalized Tonic-Clonic Seizures |
| Protocol Number: | SYMBIANC-21 |
| Drug Development Phase: | 3 |
| Study Initiation Date: | |
| Study Completion Date: | |
| GCP Statement: | This study was conducted in compliance with the |

Page 1 of 109 English (United States) Zoom 100%

Converting Post-text tables to In-text tables



Application Edit Screen

Upload Source Document Table of Contents **In-Text Table Configuration** Final In-Text

In-Text Table of Contents

Table - 4

Demographics and Other Baseline Characteristics Modified Intent-to-Treat Population

Sunovion Pharmaceuticals Inc.
SEP361-201
Final

Page 1 of 5

Table 14.1.2.1
Demographics and Other Baseline Characteristics
Modified Intent-to-Treat Population

| Statistic | Treatment Group | | | Total (N=245) |
|---|--------------------|-----------------------|--|------------------|
| | Placebo (N=125) | SEP-363856 (N=120) | | |
| Age (years) [a] | | | | |
| n | 125 | 120 | | 245 |
| Mean (SD) | 30.6 (6.07) | 30.0 (5.76) | | 30.3 (5.91) |
| Median | 32.0 | 32.0 | | 32.0 |
| Q1, Q3 | 25.0, 36.0 | 24.0, 34.0 | | 26.0, 35.0 |
| Min, Max | 15, 40 | 15, 40 | | 15, 40 |
| Age Group (years) [a] | | | | |
| n | 125 | 120 | | 245 |
| <18 years | 0 | 0 | | 0 |
| 18 - <25 years | 29 (23.2%) | 26 (21.7%) | | 55 (22.4%) |
| 25 - <40 years | 96 (76.8%) | 94 (78.3%) | | 190 (77.6%) |
| ≥40 years | 0 | 0 | | 0 |
| Sex | | | | |
| n | 125 | 120 | | 245 |
| Male | 79 (62.2%) | 77 (64.2%) | | 156 (62.7%) |
| Female | 46 (36.8%) | 43 (35.8%) | | 89 (36.2%) |
| Race | | | | |
| n | 125 | 120 | | 245 |
| American Indian or Alaska Native | 1 (0.8%) | 4 (3.2%) | | 5 (2.0%) |
| Asian | 0 | 0 | | 0 |
| Black or African American | 20 (16.0%) | 19 (15.8%) | | 39 (15.9%) |
| Native Hawaiian or Other Pacific Islander | 0 | 0 | | 0 |
| White | 104 (83.2%) | 96 (80.0%) | | 200 (81.4%) |
| Multiracial | 0 | 1 (0.8%) | | 1 (0.4%) |
| Other | 0 | 0 | | 0 |
| Ethnicity | | | | |
| n | 125 | 120 | | 245 |
| Hispanic or Latino | 6 (4.8%) | 5 (4.2%) | | 11 (4.5%) |
| Not Hispanic or Latino | 119 (95.2%) | 115 (95.8%) | | 234 (95.5%) |

Note: BMI = Body mass index. CGI-S = Clinical Global Impression - Severity Scale. Max = Maximum. Min = Minimum. N = Number of subjects in treatment group. n = Number of subjects in analysis. PANS = Positive and Negative Syndrome Scale. Q1 = 1st Quartile. Q3 = 3rd Quartile. SD = Standard deviation.
Note: Baseline is defined as the last non-missing measurement taken prior to the first dose of study medication.
Note: Percentages are based on the number of subjects with non-missing data in the MITT population in each treatment group.
Note: A subject is categorized as multiracial if more than one race is checked on the CRF form.
[a] Age at informed consent.
[b] Overall median value at baseline.
Cross Reference(s): Listings 16.2.4.1, 16.2.6.1, 16.2.6.2

In-Text Configuration

Font Size

Title Edit

Configuration Details

Calibri

14

Title

Header Column 5

Table Border? Yes No

Footer? Yes No

Copy from the table

Update & Save

Cancel

| | Treatment Group | | | Total (N=245) |
|---|--------------------|-----------------------|--|------------------|
| | Placebo (N=125) | SEP-363856 (N=120) | | |
| Age (years) ^a - n | 125 | 120 | | 245 |
| Mean (SD) | 30.6 (6.07) | 30.0 (5.76) | | 30.3 (5.91) |
| Median | 32.0 | 32.0 | | 32.0 |
| Min, Max | 18, 40 | 19, 40 | | 18, 40 |
| Age Group (years) ^a - n | 125 | 120 | | 245 |
| <18 years - n (%) | 0 | 0 | | 0 |
| 18 - <25 years - n (%) | 29 (23.2%) | 26 (21.7%) | | 55 (22.4%) |
| 25 - <40 years - n (%) | 96 (76.8%) | 94 (78.3%) | | 190 (77.6%) |
| >40 years - n (%) | 0 | 0 | | 0 |
| Sex - n | 125 | 120 | | 245 |
| Male - n (%) | 79 (62.2%) | 77 (64.2%) | | 156 (63.7%) |
| Female - n (%) | 46 (36.8%) | 43 (35.8%) | | 89 (36.3%) |
| Race - n | 125 | 120 | | 245 |
| American Indian or Alaska Native - n (%) | 1 (0.8%) | 4 (3.3%) | | 5 (2.0%) |
| Asian - n (%) | 0 | 0 | | 0 |
| Black or African American - n (%) | 20 (16.0%) | 19 (15.8%) | | 39 (15.9%) |
| Native Hawaiian or Other Pacific Islander - n (%) | 0 | 0 | | 0 |
| White - n (%) | 104 (83.2%) | 96 (80.0%) | | 200 (81.4%) |
| Multiracial - n (%) | 0 | 1 (0.8%) | | 1 (0.4%) |
| Other - n (%) | 0 | 0 | | 0 |
| Ethnicity - n | 125 | 120 | | 245 |
| Hispanic or Latino - n (%) | 6 (4.8%) | 5 (4.2%) | | 11 (4.5%) |
| Not Hispanic or Latino - n (%) | 119 (95.2%) | 115 (95.8%) | | 234 (95.5%) |
| Country - n | 125 | 120 | | 245 |



Benefits of Automation using AI



Benefits of AI

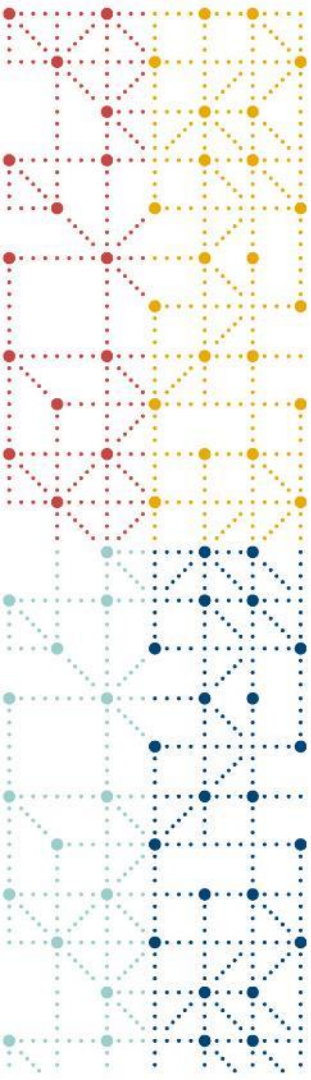
- The system is designed to generate pre-filled CSR with information from Protocol, SAP and other sources as per ICH-E3 guidelines in the respective sections of the CSR
- The system will save 60-70% of time of the medical writers when they write the CSR using the application.
- Medical writers can focus more on interpretation of study results and their discussion points.
- System can be configured as per the sponsor's needs in the workflow integration



Conclusion

Conclusions

- MW can add or modify the content by section within the system.
- Multi-Authoring (can have more than one authors for same study)
- Conversion of post-text TLF to In-text tables
- ZYLiQ editor functions as MS word and have the same look and feel.
- Traceability Report(source file to final CSR mapping), Audit Log & Version History
- Workflow integration(Author(s), Reviewer(s) & Approver)
- Narrative writing from Patient profiles/ Subject Listing .
- In-text table interpretation in simple English.
- Automatic Tense conversion (Present/future tense to past tense)



Thank You!

Contact:

Farha Fathima Feroze

Product Manager

Mobile: +1 (609) 945 7431

The logo for CDISC, featuring the lowercase letters "cdisc" in a dark blue font. Above the letter "i" are three small colored dots: a red one on the left, a yellow one in the middle, and a light green one on the right.