

The Role of AI in Handling Clinical Data and Documents (Structured and Unstructured)

Farha Fathima Feroze, Symbiance Inc, Princeton, NJ

Abstract

Artificial Intelligence Techniques can be used easier to handle structured data vs unstructured data which are documents like MS word or pdf. challenges in structured data describes a method for producing TLGs from ADaM datasets. Identify the appropriate ADaM domain and variables using various models such as Logistic regression, Classifiers, XGBOOST etc. The AI engine is trained using historical data and annotates automatically TLG mock shell for the first time. Generate reports and stand-alone SAS program. The SAS code is generated using Python. The AI algorithm is used to assist the system in learning from the user when new data is encountered and re-using the accumulated knowledge. Challenges in unstructured data are extracting data from source documents, 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.

Role of AI

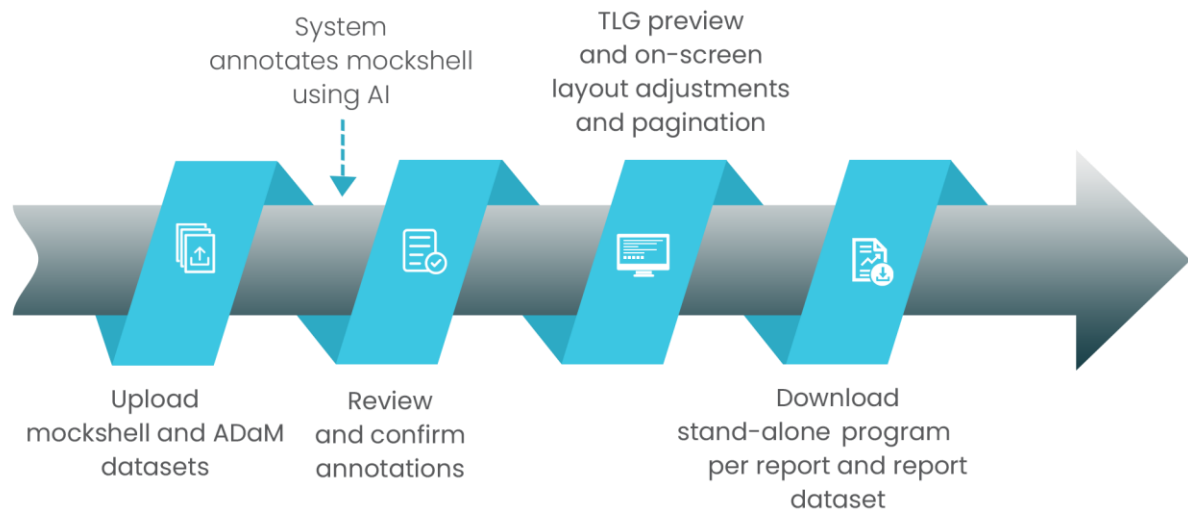
❖ Analysis & reporting tool (Structured data – SAS datasets)

- Application generates TLGs with assistance of AI (ML/NLP)
- Needs TLG shell & ADaM datasets.
- Easy adaption to any sponsor's TLG mock shells
- Downloadable stand-alone SAS programs without macros

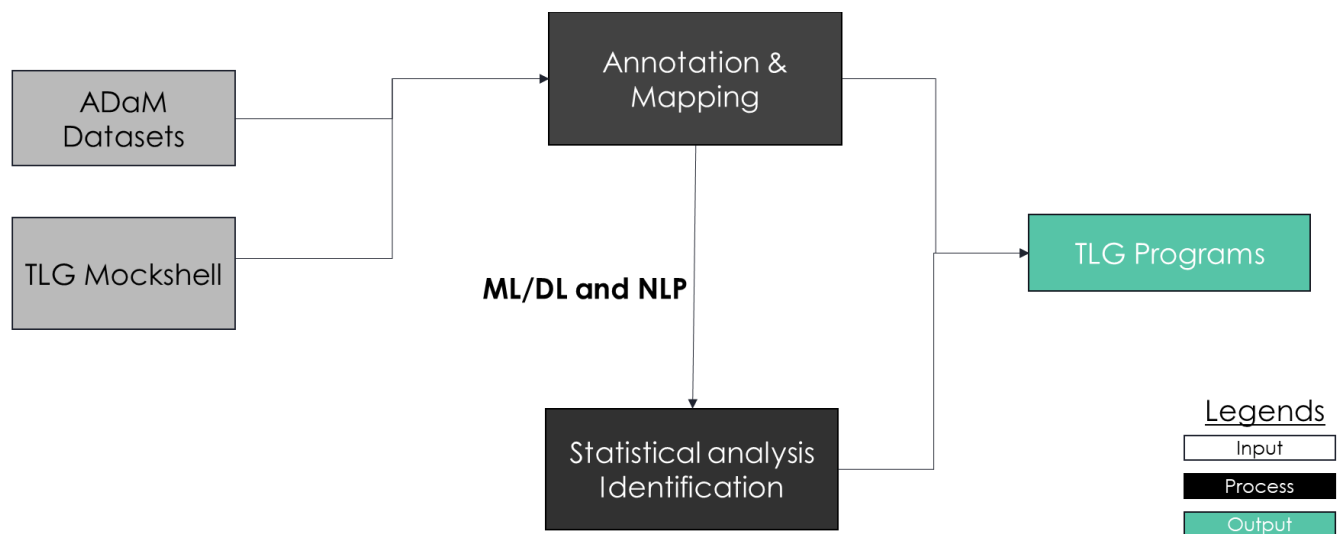
❖ Application to write scientific documents (Unstructured inputs & output)

- Tool for Authoring of CSR, Redacting documents etc. – pdf or msword using AI (ML, NLP/NLU/NLP)
- Needs source documents such as protocol, SAP, post-text/in-text tables to produce CSR.
- Workflow for document management made easier.
- Muti-authoring, edit the word document within the tool are some of the highlights.
- “Find & Retrieve” feature lets the user train the System for AI models.

Analysis & reporting process flow



Uses of ML, DL & NLP for TLG generation



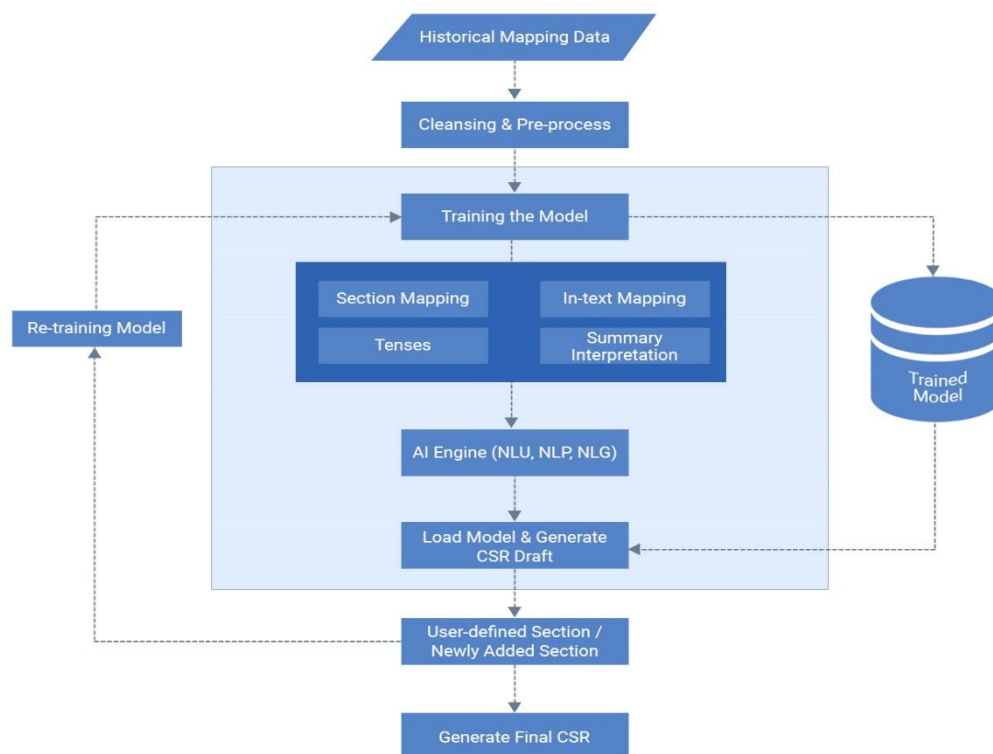
Analysis and reporting- use of ML/NLP

- ❖ To Extract table structure from TLG mock shell
- ❖ To Identify appropriate ADaM variable
- ❖ To identify required statistical methods for the given table.
- ❖ To train the system from users while it encounters new input information every time.
- ❖ To reuse the learning and produce more accurate results

Challenges handling unstructured data

- ❖ Extracting information from source documents such as PDF/MS Word
- ❖ Summarizing large information into data driven form of writing
- ❖ Read, understand and interpret the table in simple English.
- ❖ Converting tenses from present to past
- ❖ Section mapping

How AI is applied



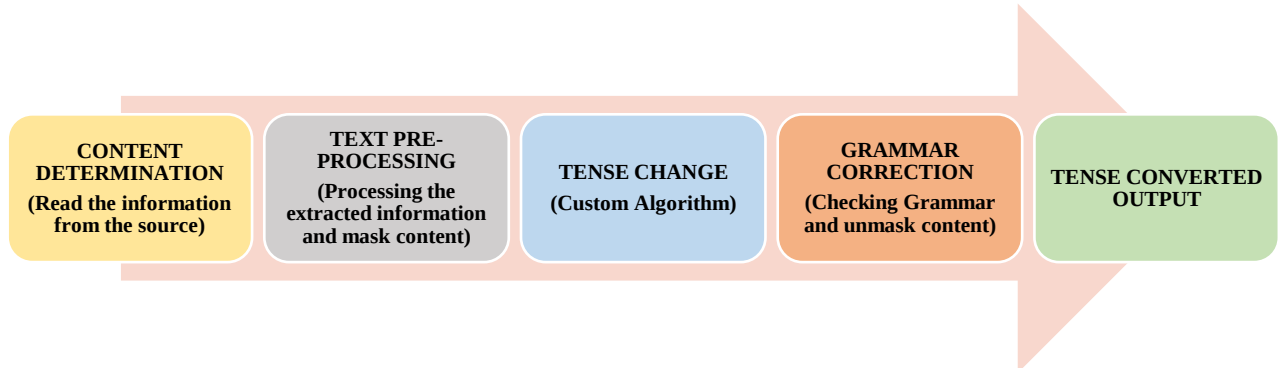
Our Approach with AI

- ❖ Lean writing
- ❖ Tense conversion.
- ❖ In-text table interpretation
- ❖ Section mapping

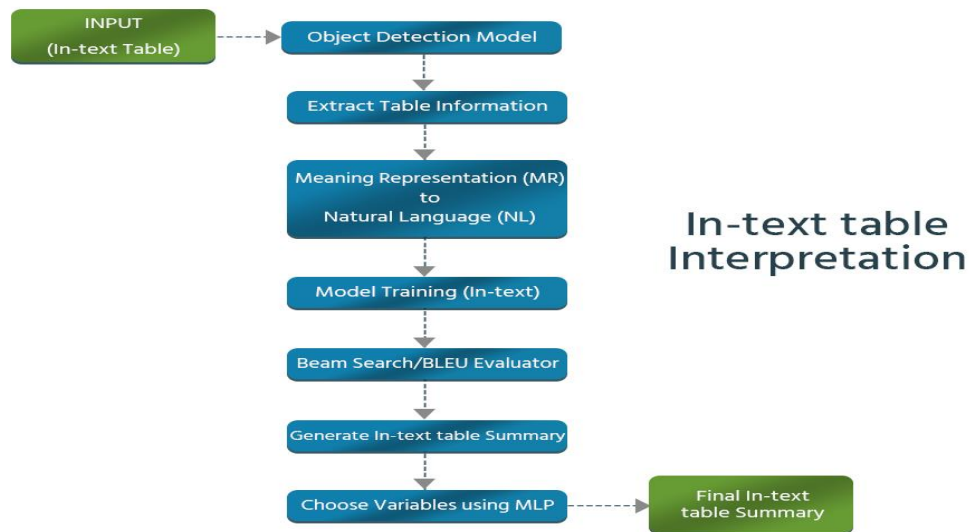
Lean writing

- ❖ Lean writing applies a data-driven approach focuses on the concrete “so what does this information mean” versus describing data and where the data came from.
- ❖ Regulatory documentation using lean writing principles is more effective than the expanded document that does not.
- ❖ Allows the user to be active and engaged in getting answers to their questions.
- ❖ Data-driven form of writing.

Tense Conversion



In-text table Interpretation



Section Mapping

- ❖ 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 content of CSR is constructed from Protocol, SAP, Safety narratives, In-text tables and medical writer's interpretation of study results.
- ❖ Copy and paste information from multiple source documents to relevant sections of the CSR.
- ❖ Creating Clinical Study Report (CSR) describes the end points, data collection and safety & efficacy analysis of the clinical study.
- ❖ Authoring of CSR is highly manual and time consuming.

Conclusion

Artificial Intelligence Techniques can be used easier to handle structured data such as TLG mock shell and unstructured data which are documents like MS word or pdf. For structure using the application it can be utilized well with any sponsor or type of TLG mock shell. Once annotations are finalized, the AI engine produces stand-alone programs and actual TLGs. And for structured using our tool it will help them in time saving for medical writers so that, they can focus on discussion points and interpretation of study results.

CONTACT INFORMATION

Farha Fathima Feroze
Product Manager
Symbiance Inc
500 College Road East Suite 220
Princeton NJ 08540
fferoze@symbiance.com
609 945 7431