PharmaSUG 2021 – Paper AI-012
Application to automate Clinical Study Report using AI

Farha Fathima Feroze, Symbiance Inc, Princeton, NJ
Ilango Ramanujam, Symbiance Inc, Princeton, NJ

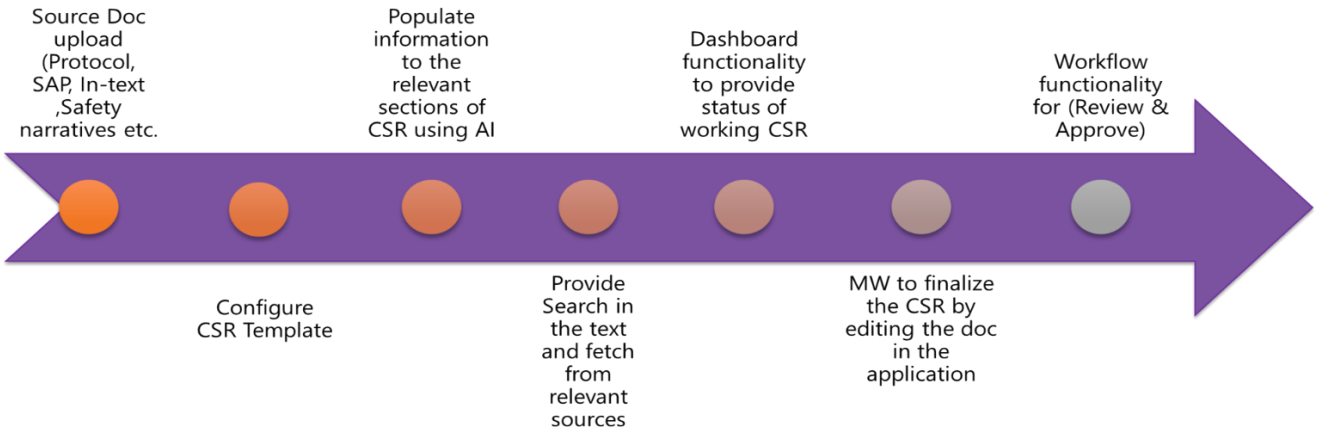
ABSTRACT

Creating Clinical Study Report (CSR) highly manual and time consuming, which require other medical writer to perform the quality control of numbers on the Table, safety narratives and information from other documents. Significant amount of information to the CSR comes from other sources such as Protocol, SAP, Safety Narratives, In-text tables etc. The appendices section is constructed from CRF, TLF etc. Automating the CSR creation by utilizing emerging technology such as machine learning, Natural language processing and Natural language generation (ML/NLP/NLG) will effectively reduce the manual efforts. This paper will discuss about one such Algorithm of Machine learning which was implemented in a tool. The tool is designed to produce pre-filled CSR with information from Protocol, SAP and other sources and interpretation of in-text tables in the respective sections of the template that follows ICH-E3 guideline. This approach could save 70%-80% of the time for the Medical writers so that they can focus on discussion points, interpretation of study results and conclusions.

BACKGROUND

- 70-80% of manual effort can be saved with automation, Medical Writers can focus on the Interpretation and discussion points and any texts that are missed by the AI engine to predict accurately.
- The AI engine smartly learns with more CSR writing using the application.
- 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 is created based on ICH –E3 guidelines
- The medical writers can focus more on **discussion points and interpretation of study results.**

SOLUTION



USE OF AI (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.

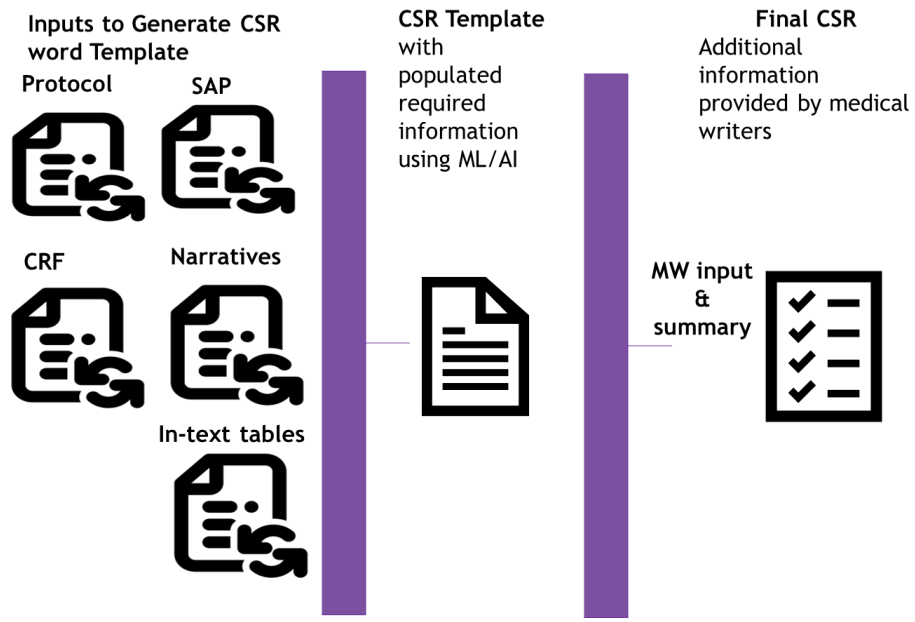
- Title Page & Synopsis:

  Customized ML Algorithm used in title page and study synopsis.

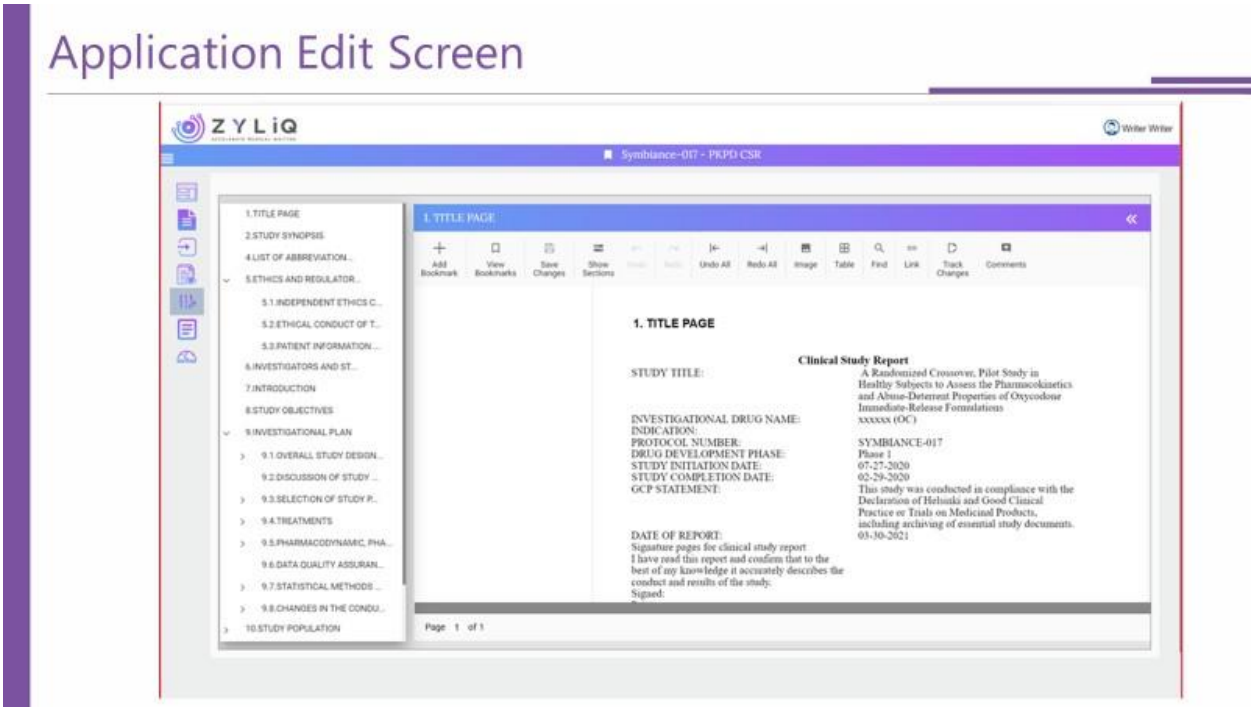
- Deep Learning Model usage:

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

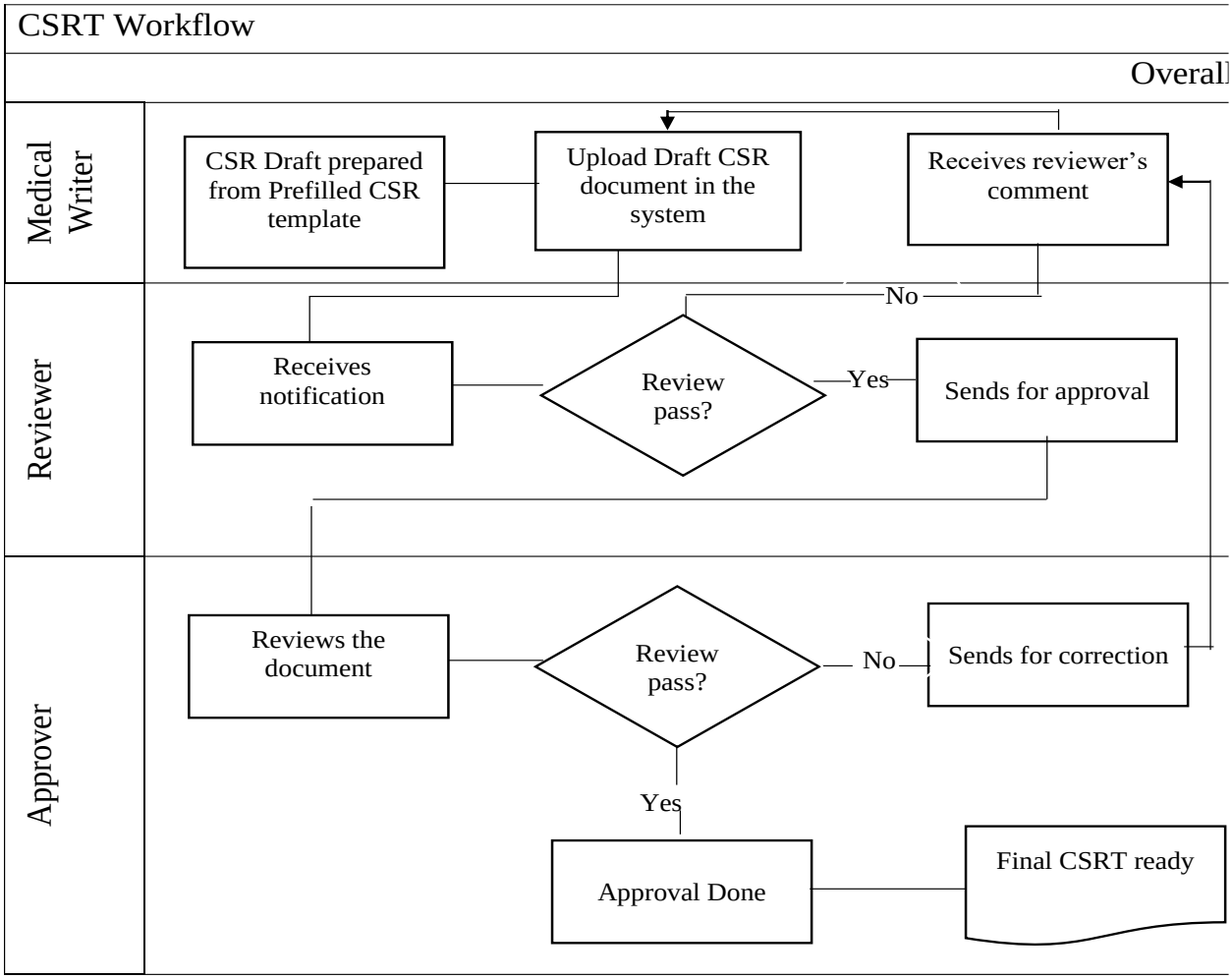
PROCESS SUMMARY



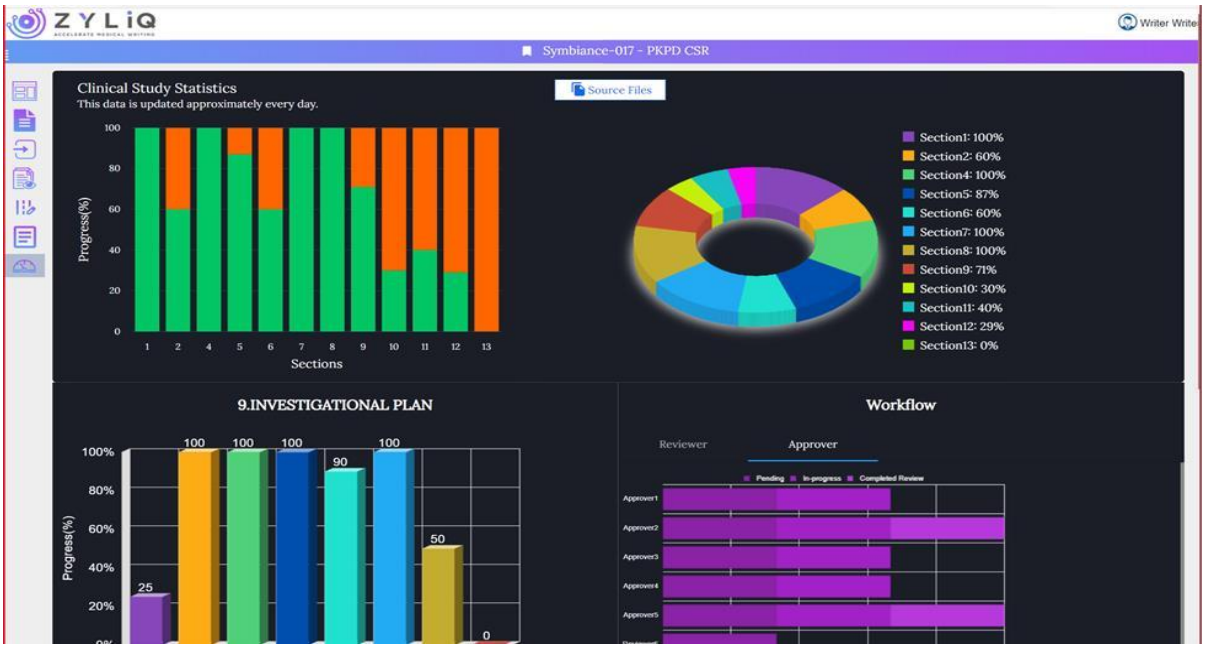
HIGHLIGHTS OF TOOL



WORKFLOW INTEGRATION



DASHBOARD



CONCLUSION

The generation of Clinical study Report (CSR) can be automated to a significant part with help of AI techniques such as Machine Learning (ML), Natural language processing (NLP) and Natural language Generation (NLG). The use of standard templates (early phase or late stage) and a flexibility to configure it based on study objectives & the information collected will save good percentage of medical writer's effort spent on creating the document. Prediction accuracy can be improved with more CSR created by using the application.

CONTACT INFORMATION

Farha Fathima Feroze
Product Manager
Symbiance Inc
500 College Road East Suite 220
Princeton NJ 08540
609 243 9050 EXT: 115