

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

STRUCTURED VS UNSTRUCTURED DATA

Structured Data

*Data that is collected from a known method and can be neatly arranged
(e.g. excel)*

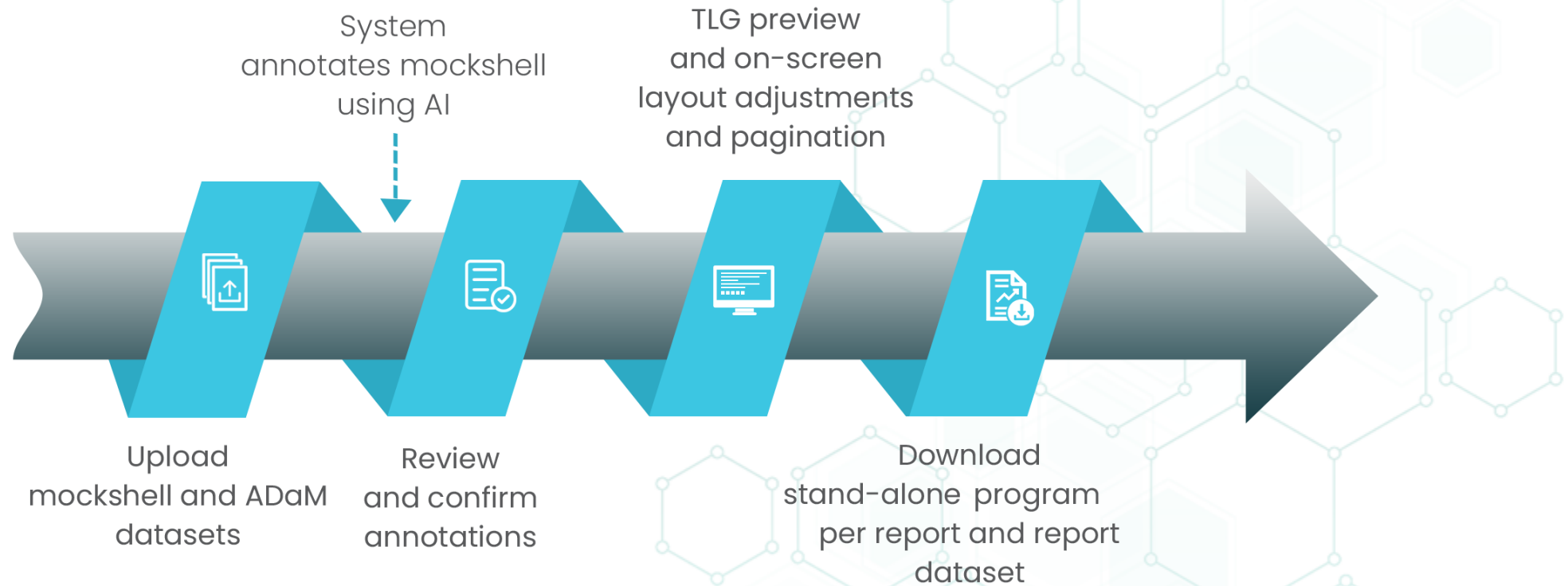
Unstructured Data

*Data that has no pre-defined format making analysis more difficult
(e.g. Msword)*

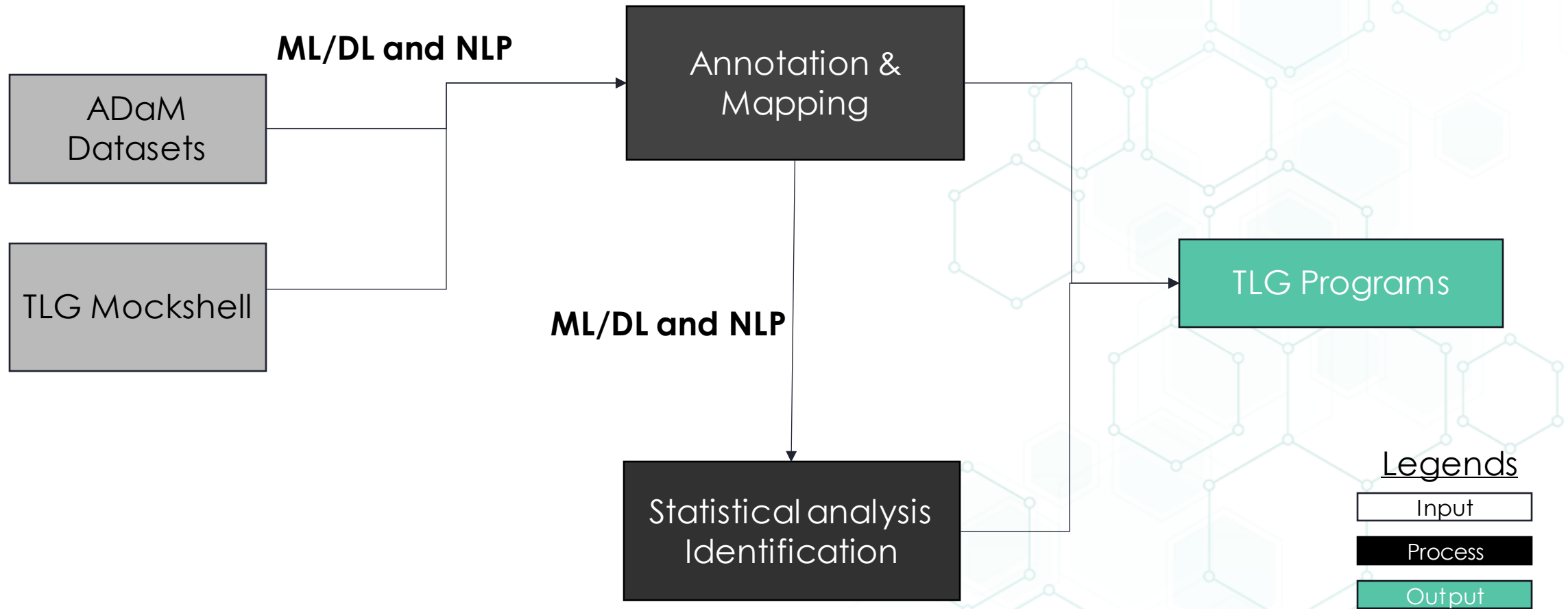
ROLE OF ARTIFICIAL INTELLIGENCE

- ❑ **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.

1. ANALYSIS & REPORTING PROCESS FLOW



USE 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

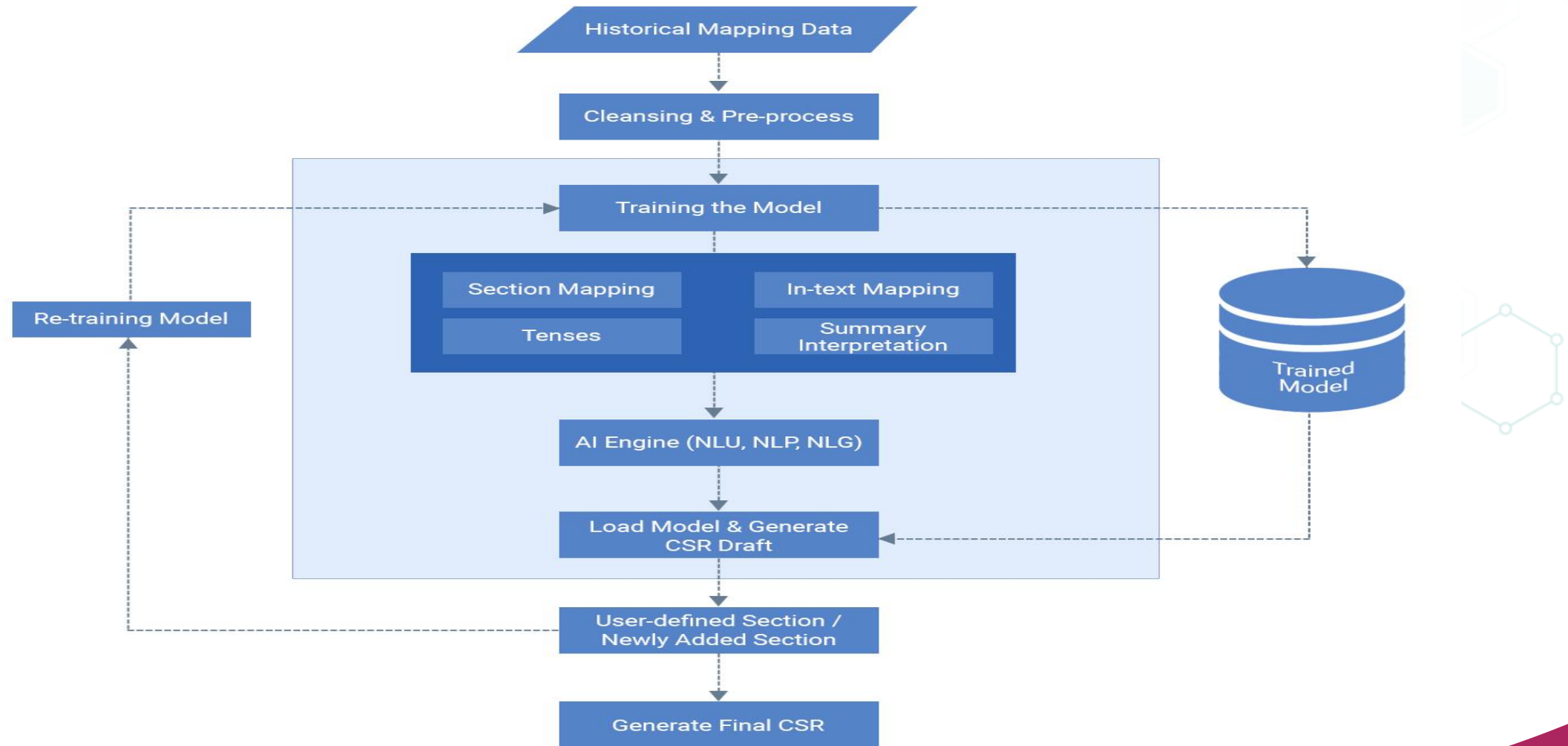
WHY AUTOMATION FOR PRODUCING CSR LIKE DOCUMENTS?

- ❑ The goal is to bring 60-70% of efficiency by automating CSR writing, MW can focus on the interpretation and discussion points.
- ❑ 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

OUR APPROACH WITH ARTIFICIAL INTELLIGENCE

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

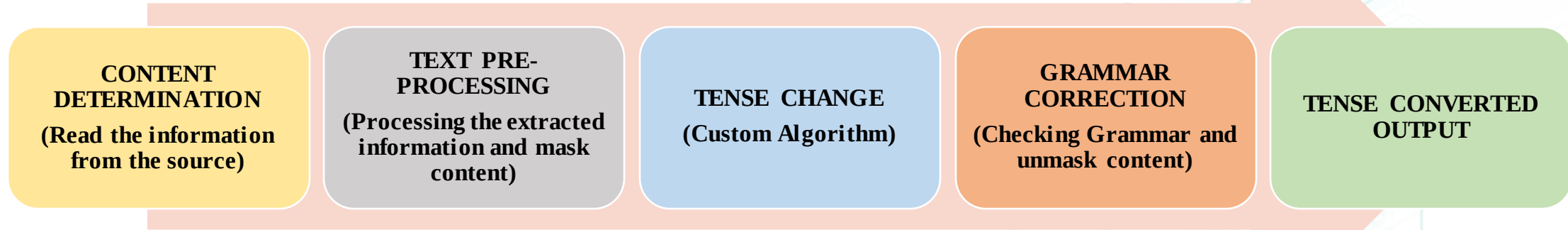
HOW AI IS APPLIED



LEAN WRITING

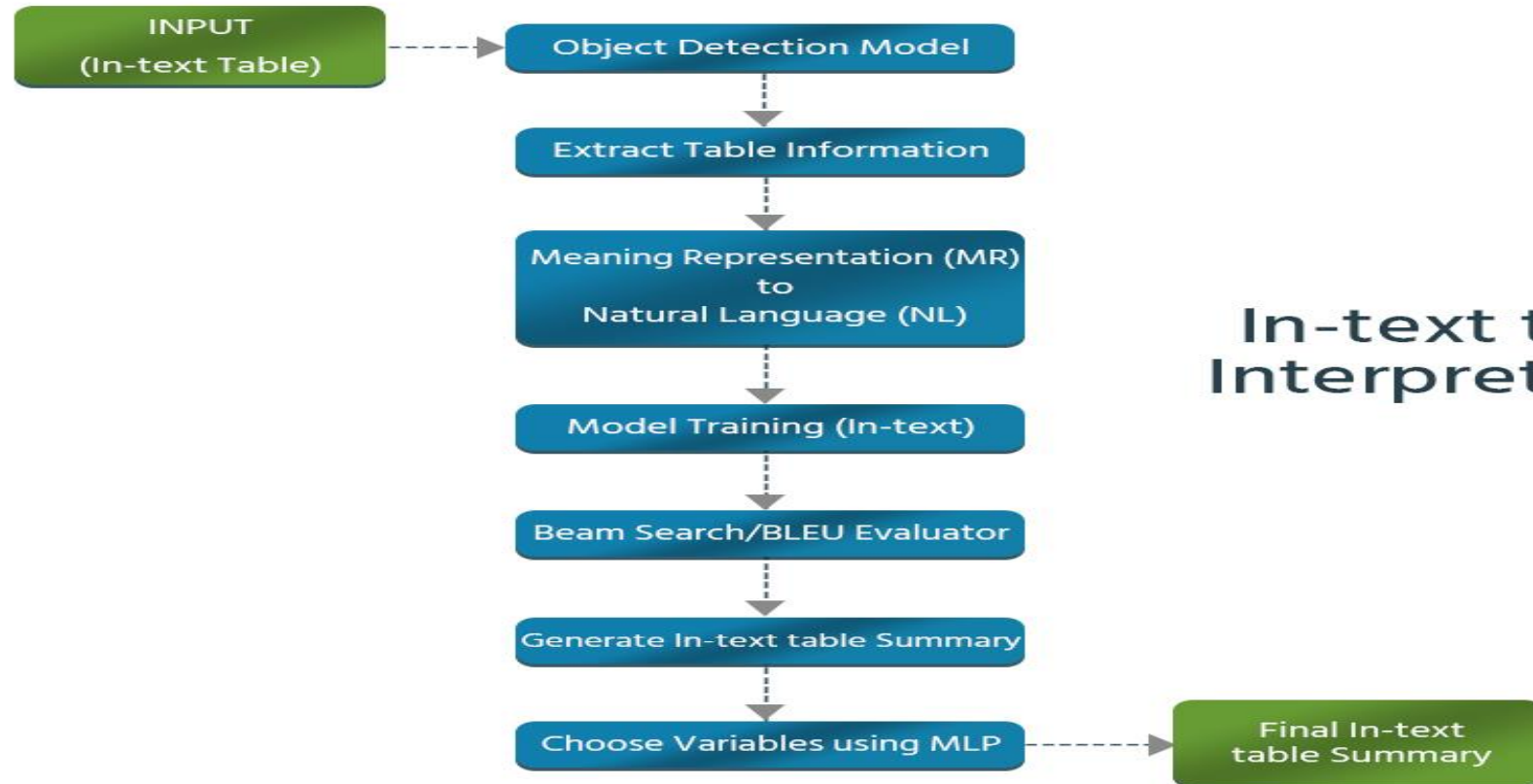
- ❑ Lean writing applies a data-driven approach focuses on the concrete “so what does this information mean” versus describing data and where the data came from.
- ❑ Regulatory documentation using lean writing principles is more effective than the expanded document that does not.
- ❑ Allows the user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

TENSE CONVERSION



- ❑ Natural language generation (NLG) is the use of artificial intelligence (AI) programming to produce written or spoken narratives from a data set
- ❑ Natural Language Generation model used acts as a translator which converts present tense to past tense.

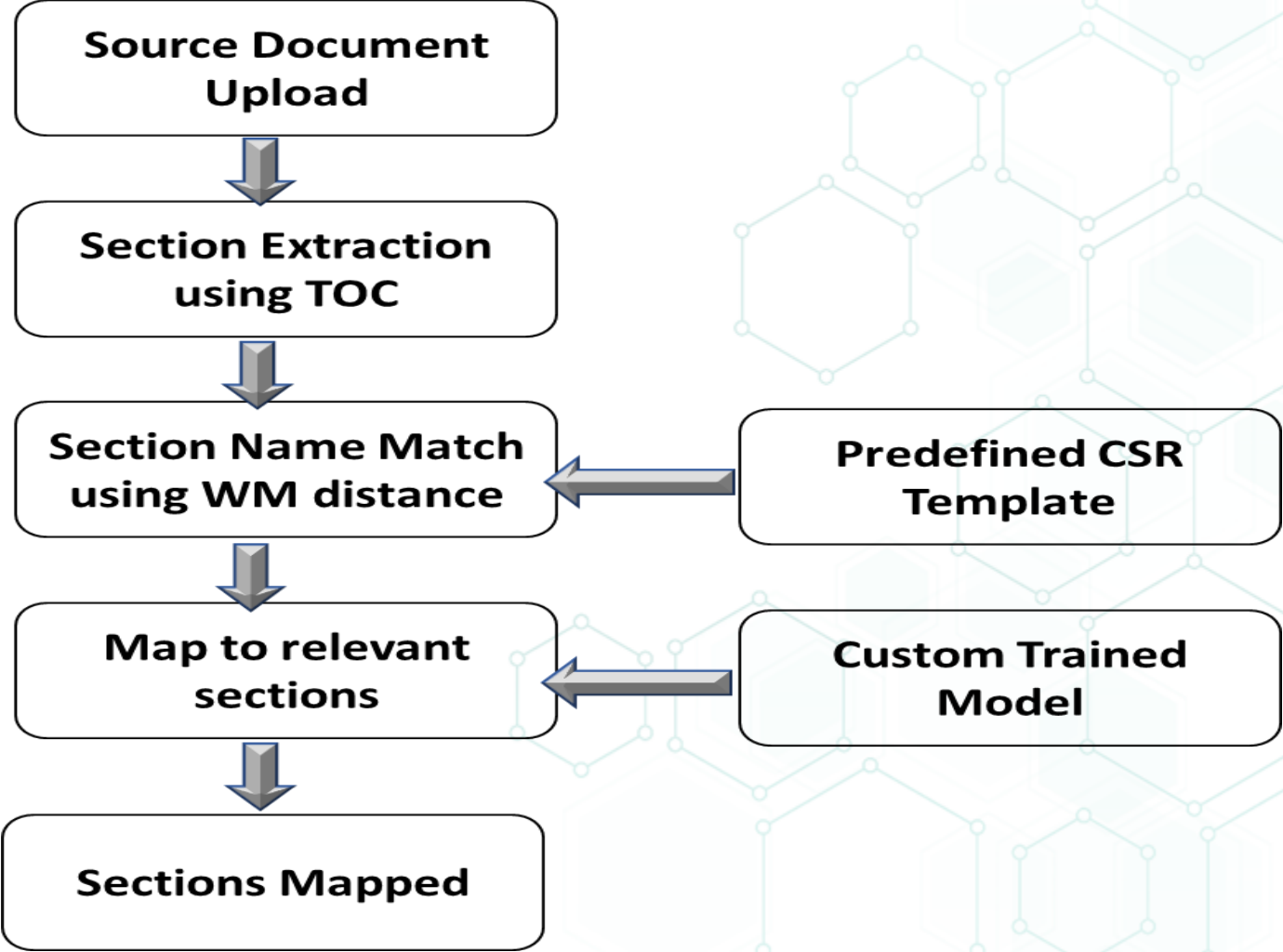
IN-TEXT TABLE INTERPRETATION



IN-TEXT TABLE INTERPRETATION – CONT'D

- ❑ Detect the in-text table structure using custom trained detectron2 object detection model
- ❑ Extract the table structure information's in data frames using custom algorithms
- ❑ Custom trained sequence2sequence model transferring meaning representations (MR) of a in-text table into a description in natural language (NL). Pytorch is used to build up the seq2seq neural model.
- ❑ Train model on the in-text table interpretation data set on a GPU
- ❑ Implement beam search for evaluation
- ❑ Implement a BLEU evaluator and report BLEU scores
- ❑ Generate the in-text table summary based on the template detected using seq2seq model
- ❑ Use a MLP Model to choose the variables whose values need to be replaced, and final summary text will be generated.

SECTION MAPPING



SECTION MAPPING – CONT'D

- ❑ 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

Questions and Comments?





Thank You!!

Farha Fathima Feroze
fferoze@symbiance.com
Product Manager
Mobile: +1 (609) 945 7431