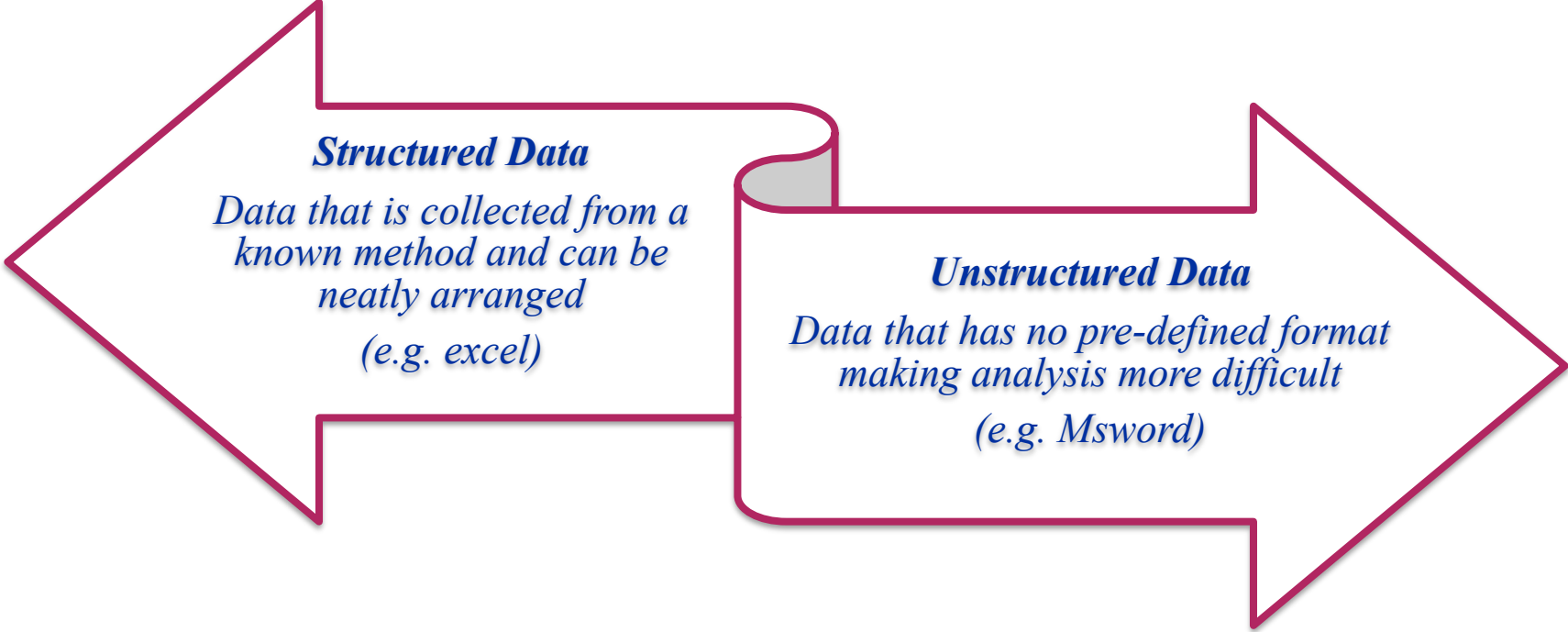


Single Day Event

Ilango Ramanujam,
Symbiance Inc

Role of AI in Clinical research functions such as Data standards, Bio-statistics & Medical writing

Data Type

A diagram illustrating two types of data. On the left, a large, hollow, maroon-outlined arrow points to the left. Inside this arrow, the text "Structured Data" is written in bold, italicized, dark blue font, followed by a description in a smaller, italicized, dark blue font: "Data that is collected from a known method and can be neatly arranged (e.g. excel)". On the right, a large, hollow, maroon-outlined arrow points to the right. Inside this arrow, the text "Unstructured Data" is written in bold, italicized, dark blue font, followed by a description in a smaller, italicized, dark blue font: "Data that has no pre-defined format making analysis more difficult (e.g. Msword)". A small, grey, rounded rectangular shape is positioned between the two arrows, overlapping the right side of the "Structured Data" arrow and the left side of the "Unstructured Data" arrow.

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



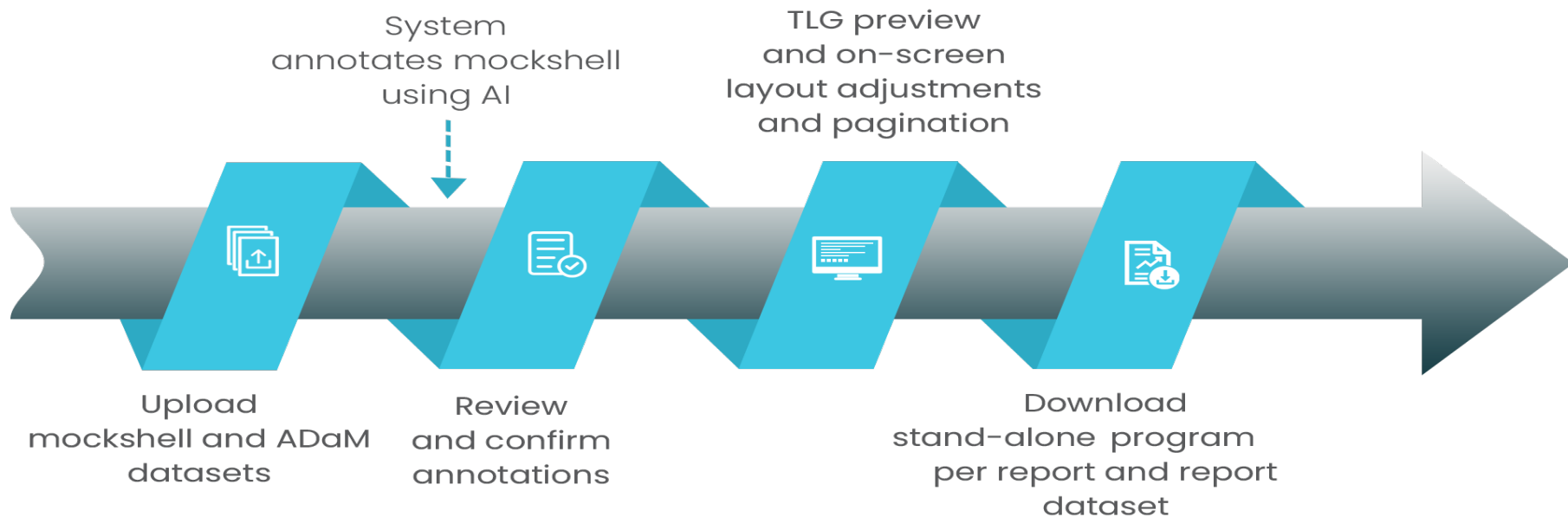
Role of Artificial Intelligence

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

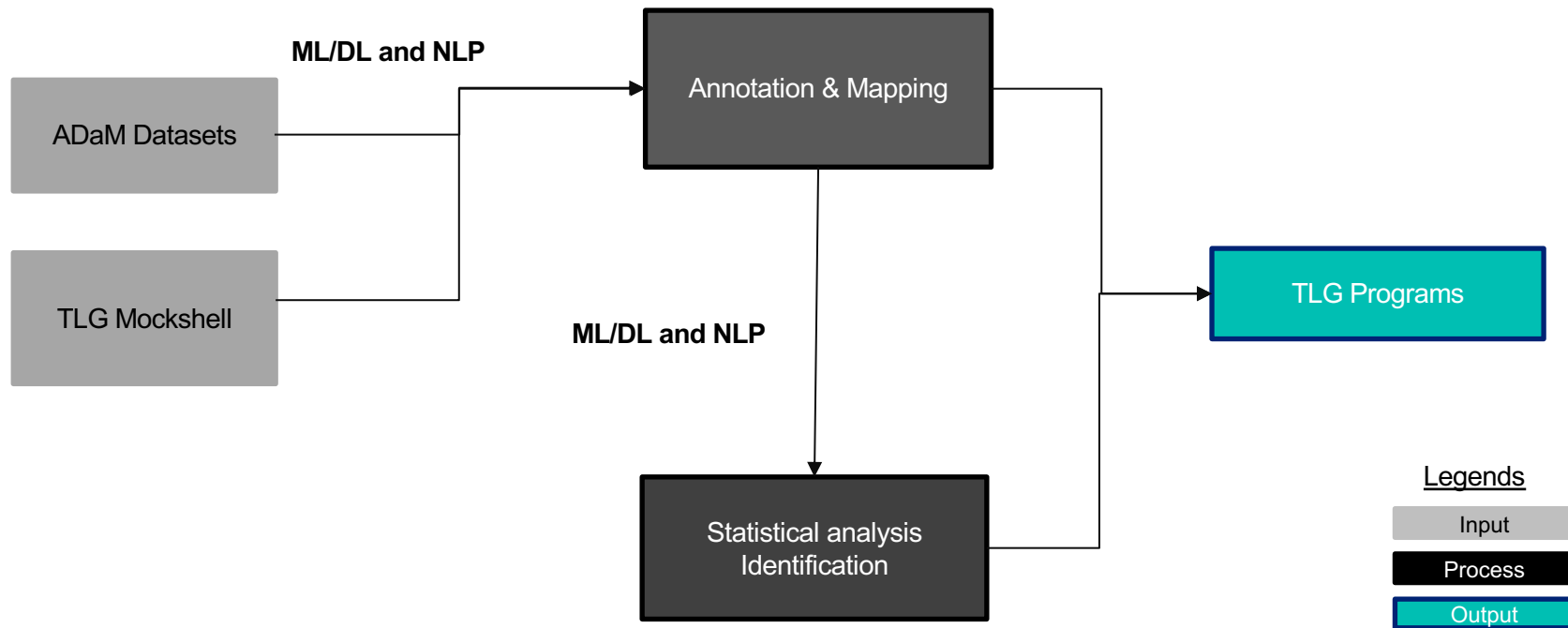


1. Analysis & Reporting Process Flow





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 re-use 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

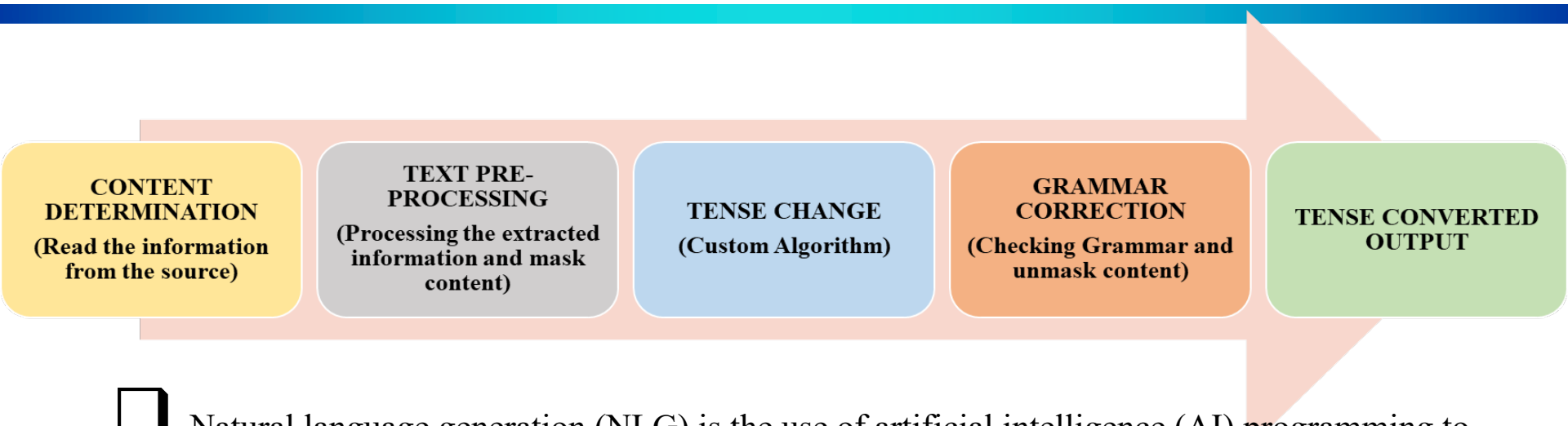


LEAN WRITING

- ☐ 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

- ☐ 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 sequece2sequence 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



IN-TEXT TABLE INTERPRETATION

- ☐ 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.



The Global Healthcare Data Science Community

Contact Channels

- 📞 UK +44 1843 609600
- ✉️ office@phuse.global
- 🌐 phuse.global

Social Media

- 🐦 [@phusetwitta](https://twitter.com/phusetwitta)
- 📘 [/phusebook](https://facebook.com/phusebook)
- ▶️ [/phusetube](https://youtube.com/phusetube)
- 🌐 [/company/phuse](https://linkedin.com/company/phuse)