

Auto Generating List of Tables from a Statistical Analysis Plan - A Case Study

Benzi Mathew - VP & Head of Clinical AI & Analytics Solutions

Jaya Inampudi - Sr Director, Clinical AI & Analytics Solutions





Objective

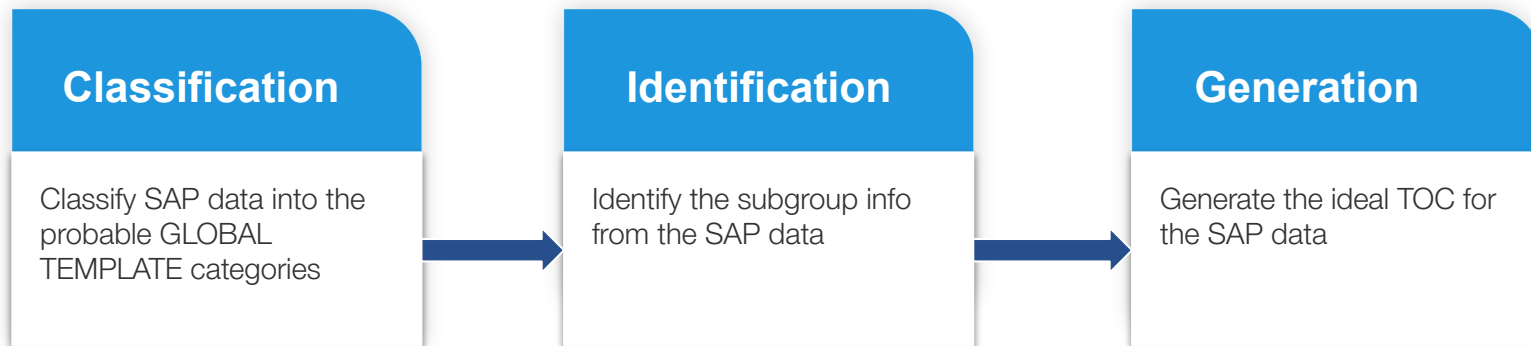
Use AI/ML techniques and heuristics based approach to develop a Proof Of Concept (POC) model, for the generation of Table of Contents/ List of Tables from statistical analysis plans associated with the clinical trial

- Utilize **Clinical trial artifacts such as Protocol, SAP, Endpoint library, sponsor Global standards**, etc. as inputs to the model and framework to create the recommendations
- Utilize **Sponsors historical study artifacts** for training and developing the AI/ ML model
- Protocol, SAP, Endpoint library, sponsor Global standards, TOC's etc.
- Utilize publicly available information as applicable for training and developing the AI/ ML model

Data Modeling



The task involves extracting ToC from SAP documents, requiring a solution that combines classification, identification, and generative techniques:





Approaches Explored

NLP models extract text data info; two approaches for TOC in SAP:

- Key Information Extraction and Related Pair Lines
- Global Template Classification using Classification/Matching and Embeddings approach

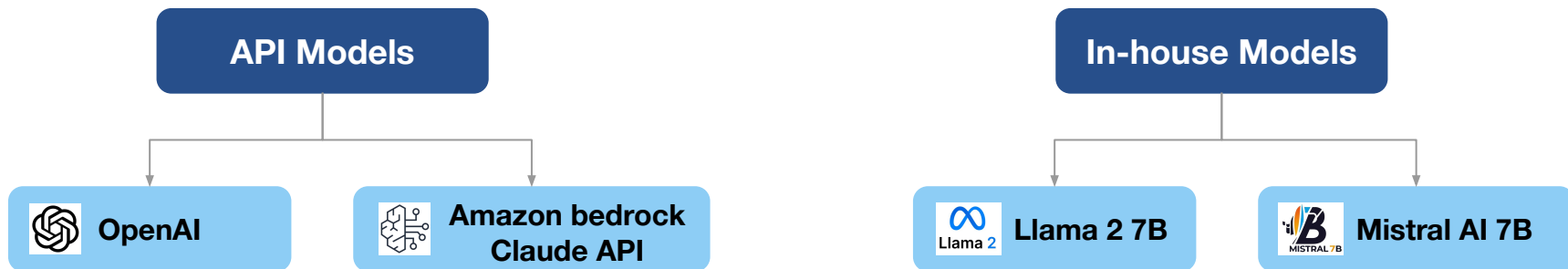
Creating SAP safety document TOC with LLM prompts requires pre-trained models:

- Zero Shot Generation
- One Shot Generation
- Chain of Thought Prompt generation

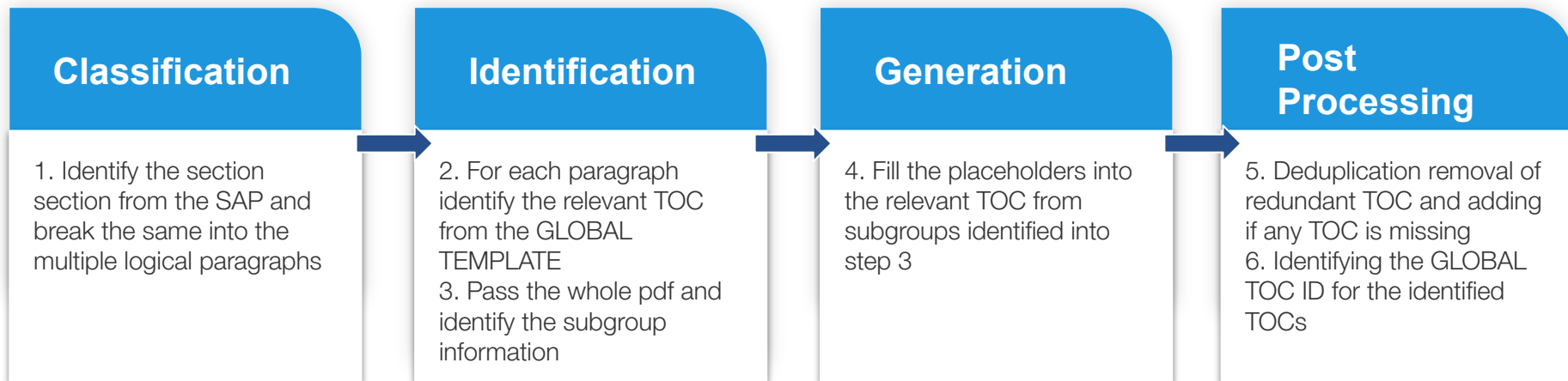


Model Performance

Generation using LLM prompts for a Safety Section TOC in SAP documents has drawbacks like inconsistency, unreliability, and longer inference times. Inhouse and SAS API models were tested for performance:



Process Overview





Evaluation

- Similarity Measure
- Weighted BERT Score using similarity score
- Precision or Global Matched rate
- Recall
- Mean Reciprocal Rank (MRR)
- Noval Generation Rate
- Human evaluation



Evaluation Metrics/Outcomes

Table 1: The table illustrates the results and metric using the proposed prompting approach for six studies in the Oncology domain. All below metrics have their corresponding number of generated TOCs as denominator for each respective SAP.

SAP Name	Generated TOCs	Matched Gen TOCs with Global Template	Not matched Gen TOCs vs Global	Precision (Match Gen TOCs in Global Template/ Gen TOCs)	Weighted Precision	Noval Generation Rate
Study 1	50	42	8	84.00%	66.60%	16.00%
Study 2	83	80	3	96.39%	74.21%	3.61%
Study 3	75	49	26	65.33%	44.93%	34.67%
Study 4	36	27	9	75.00%	49.44%	25.00%
Study 5	36	20	16	55.56%	38.89%	44.44%
Study 6	82	45	37	54.88%	38.89%	45.12%

Figure 2: Global Match Rate vs Generated TOCs

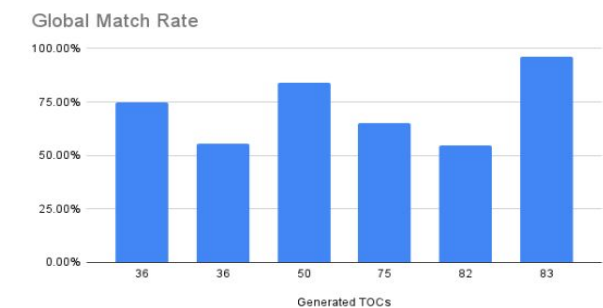


Figure 1: Global Match rate and Noval Generations vs Generated TOCs

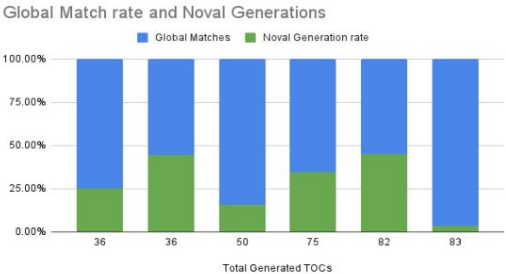
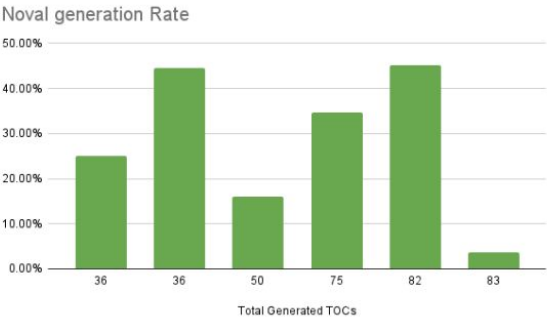


Figure 3: Noval Generations vs Generated TOCs





Analysis

Precision - While Study 2 achieved a high precision of 96.39%, the other studies show varying levels of accuracy, ranging from 54% to 84%. This suggests that the applications ability to generate TOCs that match the global template is inconsistent.

Weighted Precision - This metric also indicates variability, with study 2 performing the best and study 5 showing the lowest scores. This suggests that the generated TOCs, while sometimes accurate, may not always align closely with the global template in terms of content or structure.

Novel Generation Rate - The rates vary significantly, with Study 3 and Study 5 showing the highest novelty rates. This suggests that the application is capable of generating unique TOC's, which could be beneficial in handling non-standard or innovative analysis plans.

Improvements

A **new version of Claude** is released - reduce hallucination and understand the context better

Reduce dependency on LLMs and use heuristic approaches where possible - better performance and consistency

Use **retrieval augmented generation (RAG)** architecture for subgroup information

Fine Tuning and data feedback incorporation - reduce hallucinations

Instead of single flow prompting, creating a **RAG architecture with a query involving the embedding** - improve performance to an extent

As the PoC present only supports the SAP Safety section, an extended method is required to support other analysis sets like efficacy and PK set.



 Patient Narratives

 SAP Insights

 Cognito

 Sherpa

 PIL Generator

 CSR Generator

 DOC AI



SAP Input



Para to Probable TOC Map

Sub Group Info

Final TOC



+
Upload SAP Data

Submit



No data

 Patient Narratives

 SAP Insights

 Cognito

 Sherpa

 PIL Generator

 CSR Generator

 DOC AI

SAP Input

[Para to Probable TOC Map](#)
[Sub Group Info](#)
[Final TOC](#)


Text from SAP (Safety Section)

Safety analyses will be based on the Safety Analysis Set. No inferential statistics will be performed; only descriptive statistics will be provided.

During the study participants will be assessed for response (to determine necessity of additional courses of CPX-351) by the investigator. Participants may receive up to four cycles of treatment between induction and consolidation cycles at the investigator's discretion. The numbers and percentages of participants receiving 1 or 2 inductions, and 1, 2 or 3 consolidations will be provided. The number and percentage of participants with dose interruptions, if any, will also be summarized. Descriptive statistics will be provided for the cumulative dose received.

Adverse events recorded in the case report form will be coded SOC and PT using the MedDRA, current or most recent version. The investigator will assess the relatedness of each AE to CPX-351. The severity of AEs will be recorded using the Common Terminology Criteria for Adverse Events (CTCAE).

Adverse events are graded by the investigator as Grade 1 (mild), Grade 2 (moderate), Grade 3 (severe), Grade 4 (life threatening), or Grade 5 (fatal). The severity grade of events for which the severity was not recorded will be categorized as "missing" for

Probable TOC from Global Template

 Patient Narratives

 SAP Insights

 Cognito

 Sherpa

 PIL Generator

 CSR Generator

 DOC AI

SAP Input

> Para to Probable TOC Map

[Sub Group Info](#)

Final TOC



SAP Data

treatment : CPX-351

subgroup : Normal,Moderate,Severe

panel name : Chemistry,Hematology

AE Type : Serious,Treatment-Related

Custom Category Name or Standardized MedDRA Query :

 Patient Narratives

 SAP Insights

 Cognito

 Sherpa

 PIL Generator

 CSR Generator

 DOC AI

>

SAP Input

Para to Probable TOC Map

Sub Group Info

Final TOC



0 - Summary of Screen Failures

☆ 1 | 👍 5 | 💬 2 | [Edit](#) | [Delete](#)

1 - Study Disposition

☆ 1 | 👍 5 | 💬 2 | [Edit](#) | [Delete](#)

2 - Summary of Major Protocol Deviations

☆ 1 | 👍 5 | 💬 2 | [Edit](#) | [Delete](#)

3 - Summary of Analysis Sets

☆ 1 | 👍 5 | 💬 2 | [Edit](#) | [Delete](#)

4 - Summary of Demographic Data

☆ 1 | 👍 5 | 💬 2 | [Edit](#) | [Delete](#)

5 - Summary of Demographic Data by Cohort

☆ 1 | 👍 5 | 💬 2 | [Edit](#) | [Delete](#)

6 - Summary of Baseline Characteristics

☆ 1 | 👍 5 | 💬 2 | [Edit](#) | [Delete](#)

7 - Summary of Baseline Characteristics by Cohort

☆ 1 | 👍 5 | 💬 2 | [Edit](#) | [Delete](#)

8 - Summary of Demographic Data and Baseline Characteristics

☆ 1 | 👍 5 | 💬 2 | [Edit](#) | [Delete](#)

9 - Summary of Demographic Data and Baseline Characteristics by Cohort

☆ 1 | 👍 5 | 💬 2 | [Edit](#) | [Delete](#)

Thank You

Scan the QR code
to learn more about
Saama



Benzi Mathew
Head of Clinical AI & Analytics
Solutions & Services
benzi.mathew@saama.com



Jaya Inampudi
Sr Director, Clinical AI & Analytics
Solutions & Services
jaya.inampudi@saama.com

