phuse·share connect-share-advance

phuse.global

Single Day Events

argenx

Development of USDM through translation of human-readable protocols

Sandeep Juneja,

Clinical Solutions Lead (argenx)

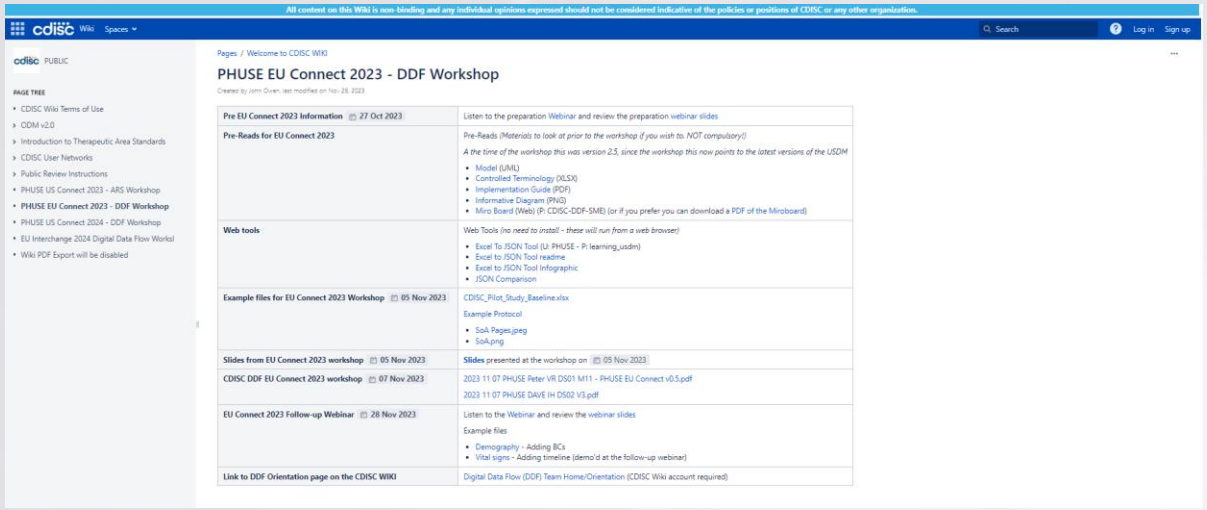
21 Nov 2024

Agenda

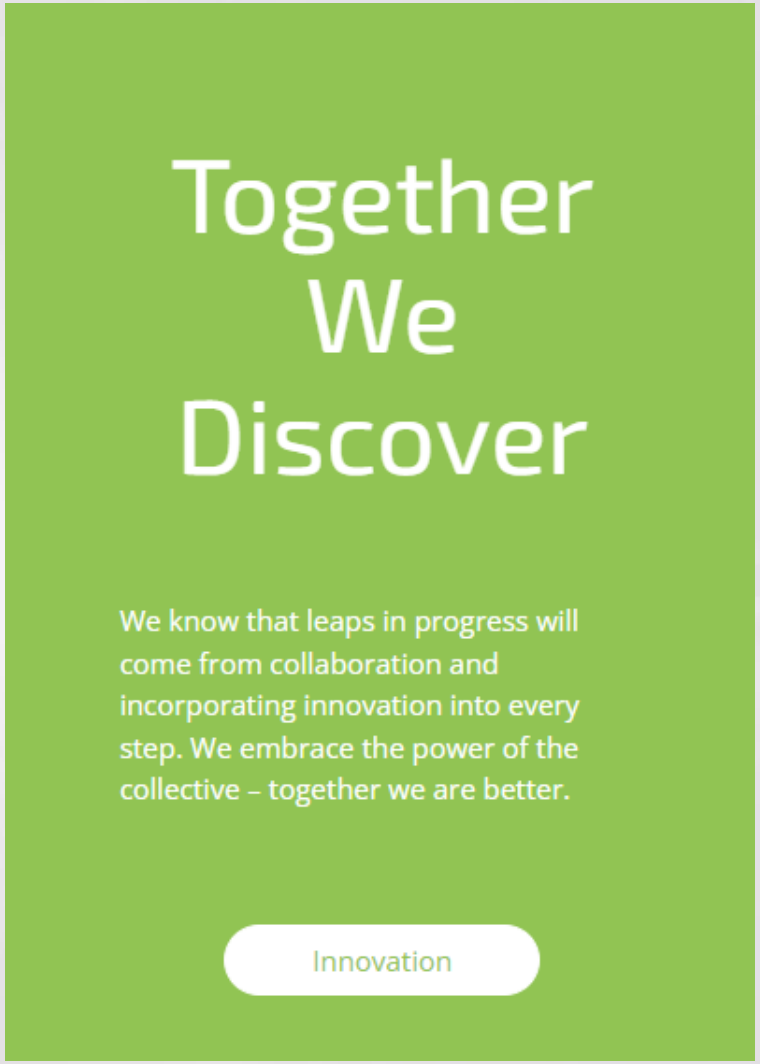
- **1. Background to the project**
- 2. The initial Journey
- 3. POC 2.0
- 4. POC 3.0 - Financial Considerations
- 5. The Journey continues... ?

How did we get here?

[PHUSE EU Connect 2023 - DDF Workshop - PUBLIC - Wiki (cdisc.org)](#)



[ICH M11 guideline, clinical study protocol template and technical specifications - Scientific guideline | European Medicines Agency (europa.eu)](#)



Co-Creation: We Create Through Collaboration
We trust in the power of the team and know that together we are better.

Ambition

- Historical protocols – machine readable – searchable library
  - Secondary goal: prove machine readable to human readable via existing protocol
    - Existing protocol → USDM → human readable

POC Scope

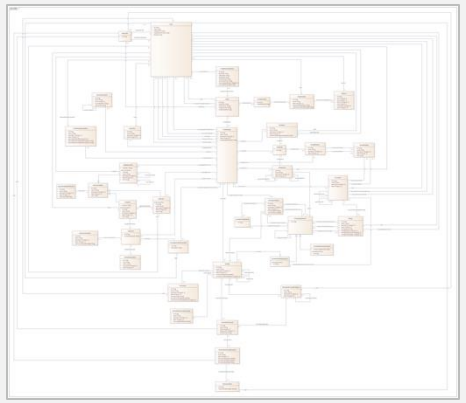
- Do **contextual information extraction** from 2 argenx clinical trial protocols (early and late phase) and **store the relevant info** into the USDM (excel) workbook … **in the right place** (in the right field).
- **Scalability:** Make sure the contextual information extraction is generic enough such that the text model used works on as many protocols as possible (including those never seen during training).

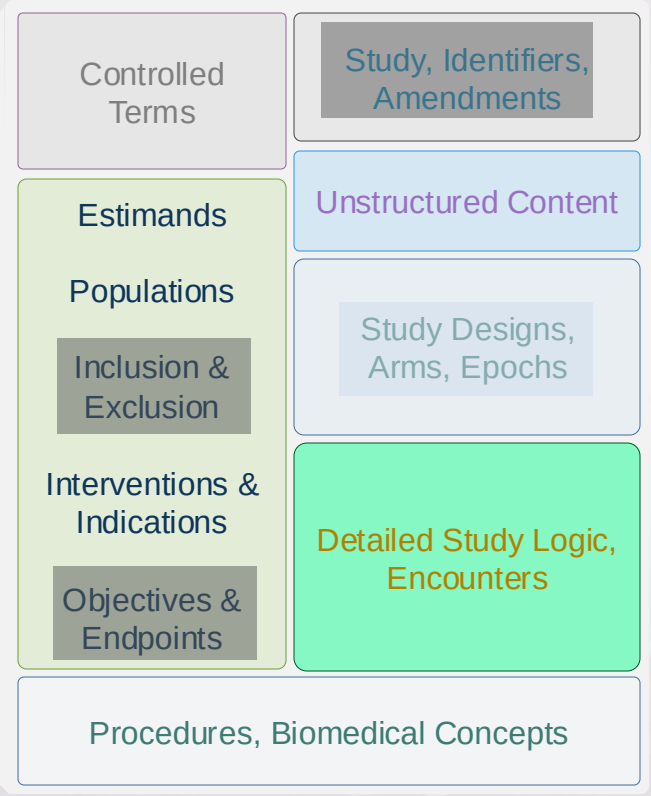
Co-Creation

- Team members

SAS:	Koen Knapen	(Principal Analytical Consultant , SAS),
	Rens Feenstra	(Principal Technology Solution Consultant , SAS),
	Jean-Charles Haillus	(Senior Project Manager, SAS),
	Fadi Glor	(Senior Account Executive , SAS)
argenx:	Jasmine Kestemont	(Head of Data Acquisition)
	Stijn Rogiers	(Head Data Integration & Standards)
	Sandeep Juneja	(Clinical Solutions Lead, DML, argenx)

USDM Content





Focus for PoC

1. Study Overview (Text)
   Study & StudyIdentifiers

2. Inclusion & Exclusion (Text)
   StudyDesignEligibilityCriteria

3. Study Objectives & Endpoints (Text)
   studyDesignOE

4. Investigational Plan:

   4.1 Study Design (Text + Image)
   studyDesign

   4.2 SoA (Table)
   mainTimeline

   4.3 Arms (Text)
   studyDesignArms

USDM excel

study Sheet

	A	B	C	D	E	F	G
1	**name**	SCOPE1					
2	**studyTitle**	Simple Test 1					
3	**studyVersion**	1					
4	**studyType**	Interventional Study					
5	**studyPhase**	C15602					
6	**studyAcronym**	SIMPLE					
7	**studyRationale**	A simple test					
8	**businessTherapeuticAreas**	SPONSOR: VAC=Vacines Group, SPONSOR: REG=Regulatory					
9	**briefTitle**	Something Brief					
10	**officialTitle**	Something Very Official					
11	**publicTitle**	Something Public					
12	**scientificTitle**	Somethign Clever But New					
13	**protocolVersion**	1					
14	**protocolStatus**	draft					
15							
16	**category**	**name**	**description**	**label**	**type**	**date**	**scopes**
17	study_version	Approval	Design approval date	Design Approval	Sponsor Approval Date	01/01/2023	country : GBR, country:FRA, region:ASIA, country :USA
18	protocol_document	Approval	Protocol document approval date	Protocol Approval	Protocol Effective Date	01/01/2023	Global
19	protocol_document	Approval	Protocol document approval date	Protocol Approval	Protocol Effective Date	01/02/2023	region:asia

studyIdentifiers Sheet

	A	B	C	D	E	F
1	**organisationIdentifierScheme**	**organisationIdentifier**	**organisationName**	**organisationType**	**studyIdentifier**	**organisationAddress**
2	USGOV	CT-GOV	ClinicalTrials.gov	Study Registry	NCT12345678	line|city|district|state|postal_code|GBR
3	DUNS	123456789	ACME Pharma	Clinical Study Sponsor	AP1234	Somewhere|In a City|In a District|In a big state|12345|FRA

studyDesignEligibilityCriteria Sheet

	A	B	C	D	E	F	G
1	**category**	**identifier**	**name**	**description**	**label**	**text**	**dictionary**
2	Inclusion	01	Age Criteria	The study age criterion		Subjects shall be between [min_age] and [max_age]	IE_Dict
3	Inclusion	02	Age Criteria Error	The study age criterion with error		Subjects shall be between [min_age] and [max_agexxx]	IE_Dict

USDM Excel Sheet Formats & Links Infographic
17th January 2024, USDM Package v0.43

Details the excel workbook format as used by the **USDM** python package.

Details of the package can be found at https://github.com/data4knowledge/usdm. Details for using the package and the sheet formats are detailed within the readme file within the repository.

USDM Excel (Cont'd)

studyDesignOE Sheet

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	objectiveName	objectiveDescription	objectiveLabel	objectiveText	objectiveLevel	objectiveDictionary	endpointName	endpointDescription	endpointLabel	endpointText	endpointPurpose	endpointLevel	endpointDictionary
2	OBJ1	Primary		The primary efficacy objective for this study is to evaluate the efficacy of TCZ compared with placebo in combination with SOC for the treatment of severe COVID-19 pneumonia	Study Primary Objective		END1	Day 28, 7 category scale		Clinical status assessed using a 7-category ordinal scale at Day 28		Primary Endpoint	
3	OBJ2	Secondary		The secondary efficacy objective for this study is to evaluate the efficacy of TCZ compared with placebo in combination with SOC for the treatment of severe COVID-19 pneumonia over the age of [min_age]	Study Secondary Objective	OE_Dict	END2	TTCI		Time to clinical improvement (TTCI) defined as a National Early Warning Score 2 (NEWS2) of <=2 maintained for 24 hours		Secondary Endpoint	
4							END3	Time to Imporvement		Time to improvement of at least 2 categories relative to baseline on a 7-category ordinal scale of clinical status		Secondary Endpoint	

USDM Excel Sheet Formats & Links Infographic
17th January 2024, USDM Package v0.43

Details the excel workbook format as used by the **USDM** python package.

Details of the package can be found at https://github.com/data4knowledge/usdm. Details for using the package and the sheet formats are detailed within the readme file within the repository.

studyDesignElements Sheet

	A	B	C	D	E
1	name	studyElementDescription	Label	transitionStartRule	transitionEndRule
2	EL1	Screening Element	Screening	Study Start	Screened
3	EL2	Baseline Element	Baseline	Screened	Radomized
4	EL3	Treatment Element 1	Treatment 1	Radomized	Completed treatment 1
5	EL4	Follow Up Element	Follow Up	Treated	Leave Study
6	EL5	Treatment Element 2	Treatment 2	Radomized	Completed treatment 2

studyDesign Sheet

	A	B	C	D	E
1	studyDesignName	Study Design 1			
2	studyDesignDescription	The main design for the study			
3	therapeuticAreas	SPONSOR:T2_DIABETES=Type 2 diabetes, SNOMED: 73211009=Diabetes mellitus (disorder)			
4	studyDesignRationale	Basic study			
5	studyDesignBlindingScheme	OPEN LABEL			
6	trialIntentTypes	BASIC SCIENCE, DEVICE FEASIBILITY			
7	trialTypes	Efficacy Study			
8	interventionModel	C82639			
9	mainTimeline	mainTimeline			
10	otherTimelines				
11					
12	Epoch/Arms	Screening	Baseline	Treatment	Follow-Up
13	Active	EL1	EL2	EL3,EL5	EL4
14	Placebo	EL1	EL2	EL5, EL3	EL4

studyDesignArms Sheet

	A	B	C	D	E	F
1	name	description	label	type	dataOriginDescription	dataOriginType
2	Active	Active Substance	Active Substance	Active Comparator Arm	Data collected from subjects	Data Generated Within Study
3	Placebo	Placebo	Placebo	Placebo Comparator Arm	Data collected from subjects	Data Generated Within Study

<timeline name> Sheet

	A	B	C	D	E	F	G	H	I	J
1	Name	Main Timeline	name	SCREEN	PRE DOSE	DOSE	D14	PROG	D28	FU
2	Description	This is the main timeline for the study design.	description	Screening	Pre Dose	Dosing	Day 14	Check Opt In	Day 28 Optional	Follow Up
3	Condition	Potential subject identified	label	Screen	Baseline	Treatment	Day 14	Check opt In by subject	Day 28	Follow Up
4			type	Activity	Activity	Activity	Activity	Decision	Activity	Activity
5			default	PRE DOSE	DOSE	D14	PROG	D28	FU	(EXIT)
6			condition					FU: if opted out		
7			epoch	Screening	Baseline	Treatment	Treatment		Treatment	Follow-Up
8			encounter	E1	E2	E3	E4		E5	E6
9	Parent Activity	Child Activity	BC/Procedure/Timeline							
10	-	Demographics	BC:Age, BC:Sex, BC:Race, BC:Body Weight	X						
11	-	Procedures	PR: PR1, PR: PR2	X	X	X	X			X
12	-	Optional Weight	BC:Weight	X			X			
13		Optional	BC:SYSBP, BC:DIABP						X	

Agenda

- 1. Background to the project
- **2. The initial Journey**
- 3. POC 2.0
- 4. POC 3.0 - Financial Considerations
- 5. The Journey continues... ?

POC approach

1) **Usage of LITI rules** for contextual information extraction

   LITI = **L**anguage **I**nterpretation for **T**extual **I**nformation

   LITI is proprietary syntax from SAS

   LITI includes concept rule types as well as fact rule types.

2) **Usage of LLM (RAG-for-LLMs)** for contextual information extraction

   RAG = **R**etrieval-**A**ugmented **G**eneration

3) PDF Table Extractor(s)
4) PDF Image Extractor(s)

5) Load from SAS tables into Excel and Word in an automated way

   *while retaining the link i.e. if the info in the SAS table changes, then you can just refresh the *.xlsx or *.docx file to see that new info reflected.*

Comparison of 2 ways for information retrieval / contextual extraction

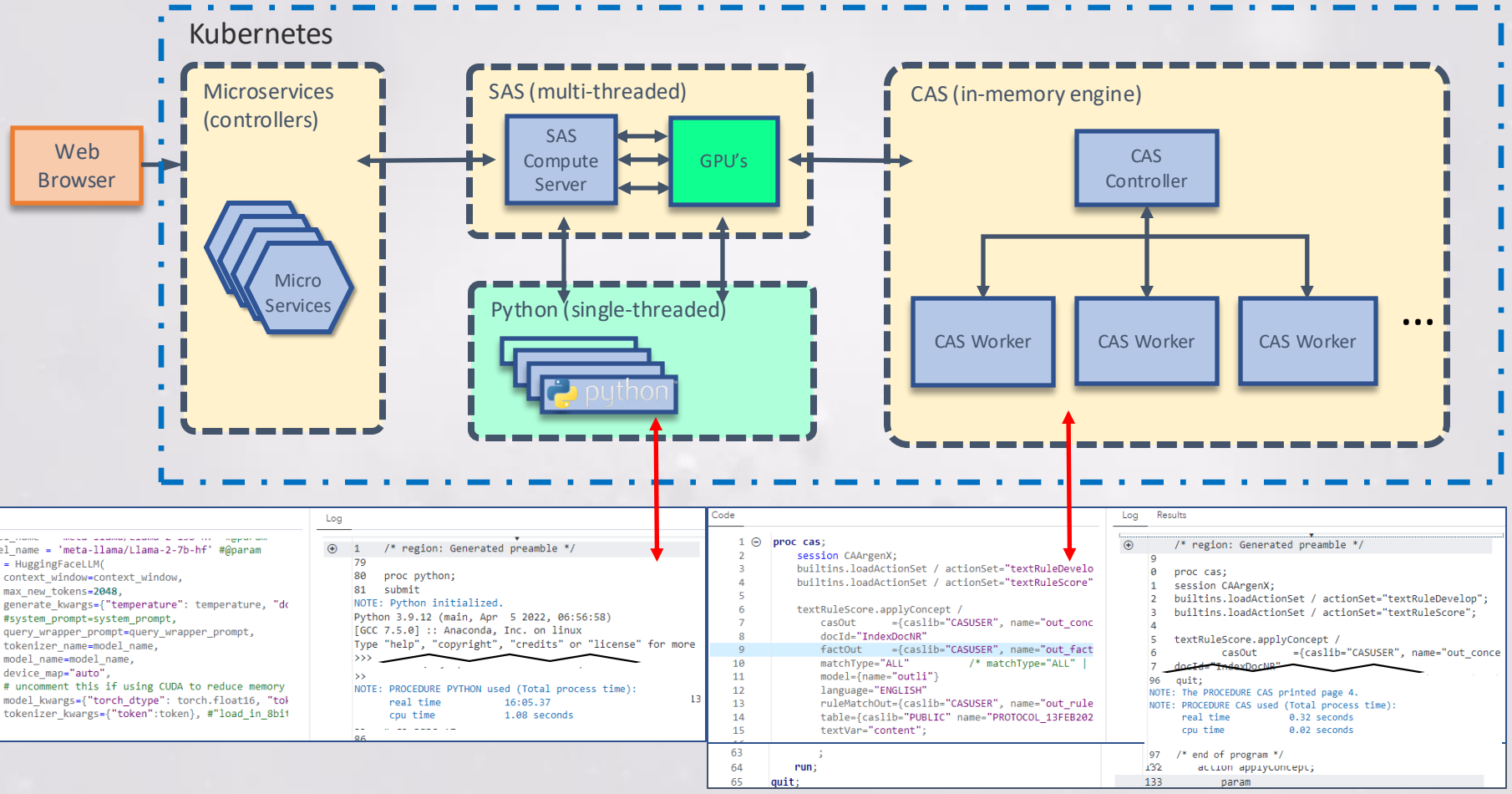
LITI – rules (SAS VTA)

- It's "Regular Expressions on steroids".
- Knowledge (and effort) required to write linguistic rules with correct syntax.
- Some patterns are hard to come by (because they are complicated, non-standard, with a lot of variety in the language …).
- You need some preliminary knowledge on the topic at hand and on what you want to catch!
- Rules are easily overfit (too specific) if n° of training docs is low.
- Library of LITI-rules can / should grow very big to guarantee a high hit rate.
- Results are "proven", non-debatable (there's a clear match in the text).
- Relatively light in terms of resource usage (and cost).
- page breaks and headers and footers are no gift.
- special and non-printable characters pose no problem.

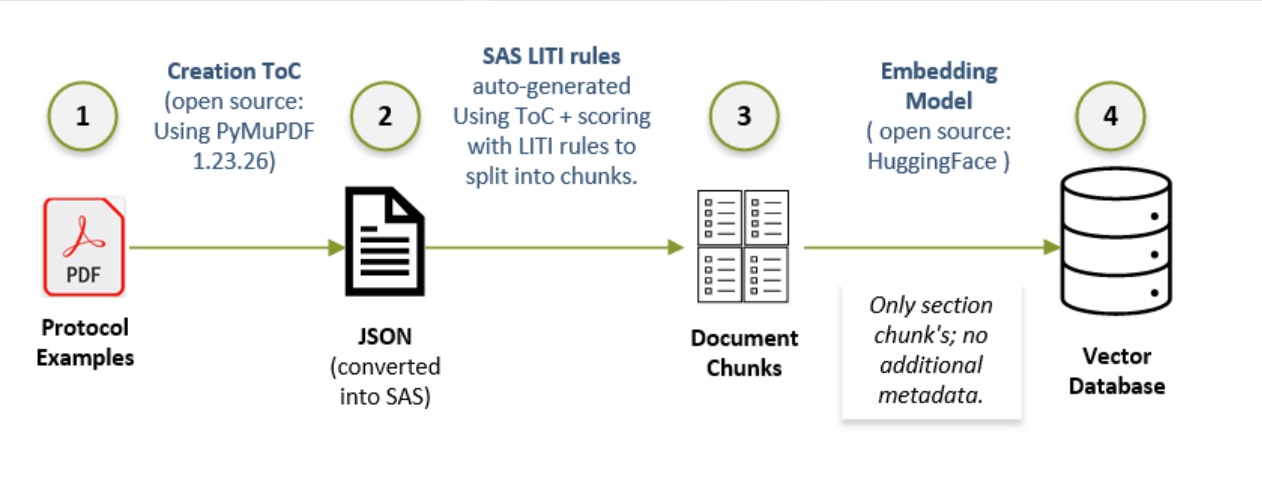
RAG – LLM (Open Source / Commercial Models)

- It's generative AI.
- Knowledge required to do proper prompt engineering (designing the user query). But much less effort required. It's just an API-call to a pre-trained model.
- If the context around a particular topic can vary widely from document to document, there is a clear advantage.
- It's generative AI, so some of the info returned can be made up (hallucination).
- Implementing offline models (internalization) and maintaining them is labor-intensive.
- Heavy and intensive in terms of resource usage (even for one or a few documents). Often, several GPUs are needed, and run-time is considerable.
- Scaling (expand the scope) is easier here.
- page breaks and headers and footers are no gift.
- special and non-printable characters pose no problem.

The environment: SAS Viya & Open Source



Pre-Processing to Overcome challenges



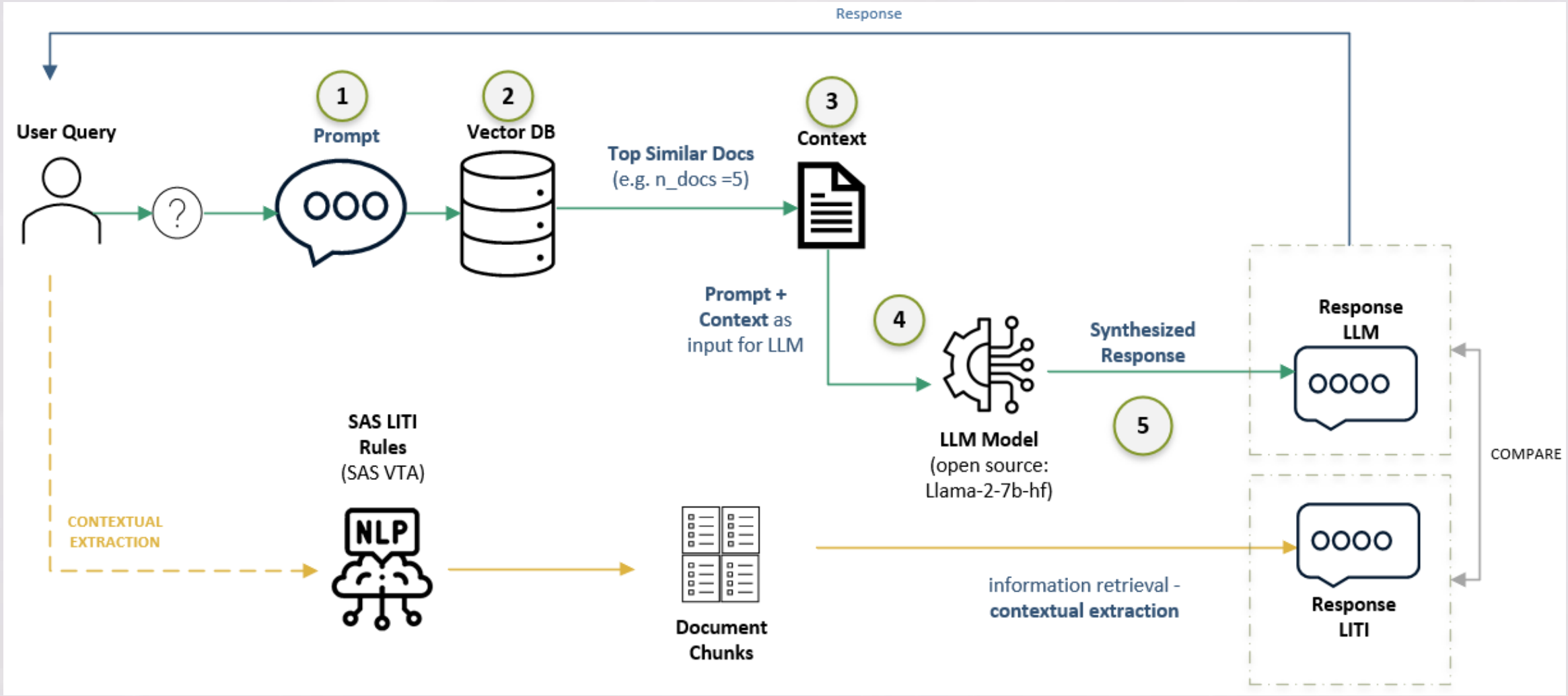
Log | Output Data (1)

TOC — Table rows: 113 | Columns: 3 of 3 | Rows 1 to 113

Enter expression

	toc	page	chapter
20	[3, '6.2.1. Primary Endpoint', 19]	19	6.2.1. Primary Endpoint
21	[3, '6.2.2. Secondary Endpoints', 19]	19	6.2.2. Secondary Endpoints
22	[1, '7. Investigational Plan', 20]	20	7. Investigational Plan
23	[2, '7.1. Overall Study Design', 20]	20	7.1. Overall Study Design
24	[2, '7.2. Number of Subjects', 20]	20	7.2. Number of Subjects
25	[2, '7.3. Treatment Assignment', 20]	20	7.3. Treatment Assignment
26	[2, '7.4. Dose Adjustment Criteria', 21]	21	7.4. Dose Adjustment Criteria
27	[2, '7.5. Criteria for Termination of Study', 21]	21	7.5. Criteria for Termination of Study
28	[1, '8. Selection and Withdrawal of subjects', 27]	27	8. Selection and Withdrawal of subjects
29	[2, '8.1. Inclusion Criteria', 27]	27	8.1. Inclusion Criteria
30	[2, '8.2. Exclusion Criteria', 28]	28	8.2. Exclusion Criteria
31	[2, '8.3. Subject Withdrawal Criteria', 30]	30	8.3. Subject Withdrawal Criteria
32	[1, '9. Treatment of Subjects', 31]	31	9. Treatment of Subjects
33	[2, '9.1. Description of IMP', 31]	31	9.1. Description of IMP
34	[2, '9.2. Restrictions', 31]	31	9.2. Restrictions
35	[3, '9.2.1. Concomitant Medication/Procedure(s)', 31]	31	9.2.1. Concomitant Medication/Procedure(s)
36	[3, '9.2.2. Alcohol', 32]	32	9.2.2. Alcohol
37	[3, '9.2.3. Physical Activities', 32]	32	9.2.3. Physical Activities
38	[3, '9.2.4. Dietary Aspects', 32]	32	9.2.4. Dietary Aspects
39	[3, '9.2.5. Smoking', 32]	32	9.2.5. Smoking

SAS LITI (NLP) + LLM



Results after pre-/processing

(example IN/EXCLUSION criteria)

Legend :

perfect
> 75 % hit rate
between 60 % and 75 % hit rate
< 60% hit rate

		INCLUSION CRITERIA			EXCLUSION CRITERIA		
Documents (clinical trial protocols)	document partitioning (training / validation / test)	Number	Number Captured With LITI	Number Captured with LLM	Number	Number Captured With LITI	Number Captured with LLM
Protocol - 13 Feb 2020.pdf	TRAINING	15	15	12*	22	22	19
Protocol_22Jul2020.pdf	TEST	7	7	6	20	20	34*
EliLilly_NCT03421379_Diabetes.pdf		10	10	10	26	26	26
Roche_NCT04320615_COVID.pdf		7	7	7	12	12	12
CDISC_Pilot_Study.pdf		8	8	8	23	23	23

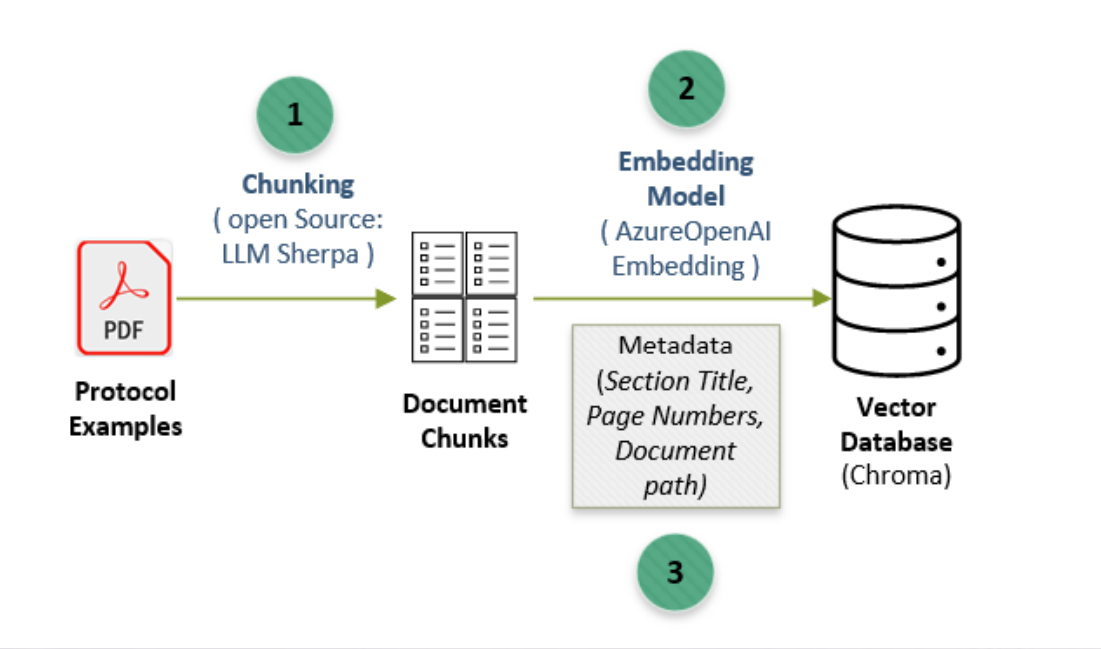
Agenda

- 1. Background to the project
- 2. The initial Journey
- **3. POC 2.0**
- 4. POC 3.0 - Financial Considerations
- 5. The Journey continues... ?

POC 2.0 - Accuracy Enhancements

Innovation: Our Core Mission
We live to innovate — and do so at every step.

Pre-Processing




```
56    os.environ["EMBED_MODEL"] = "text-embedding-ada-002-endpoint"
57
58    embedding_model_deployment = SAS.symget("embeddingModelDeployment")
```

LLM Sherpa → https://github.com/nlmatics/nlm-ingestor/blob/main/notebooks/test_llmsherpa_api.ipynb

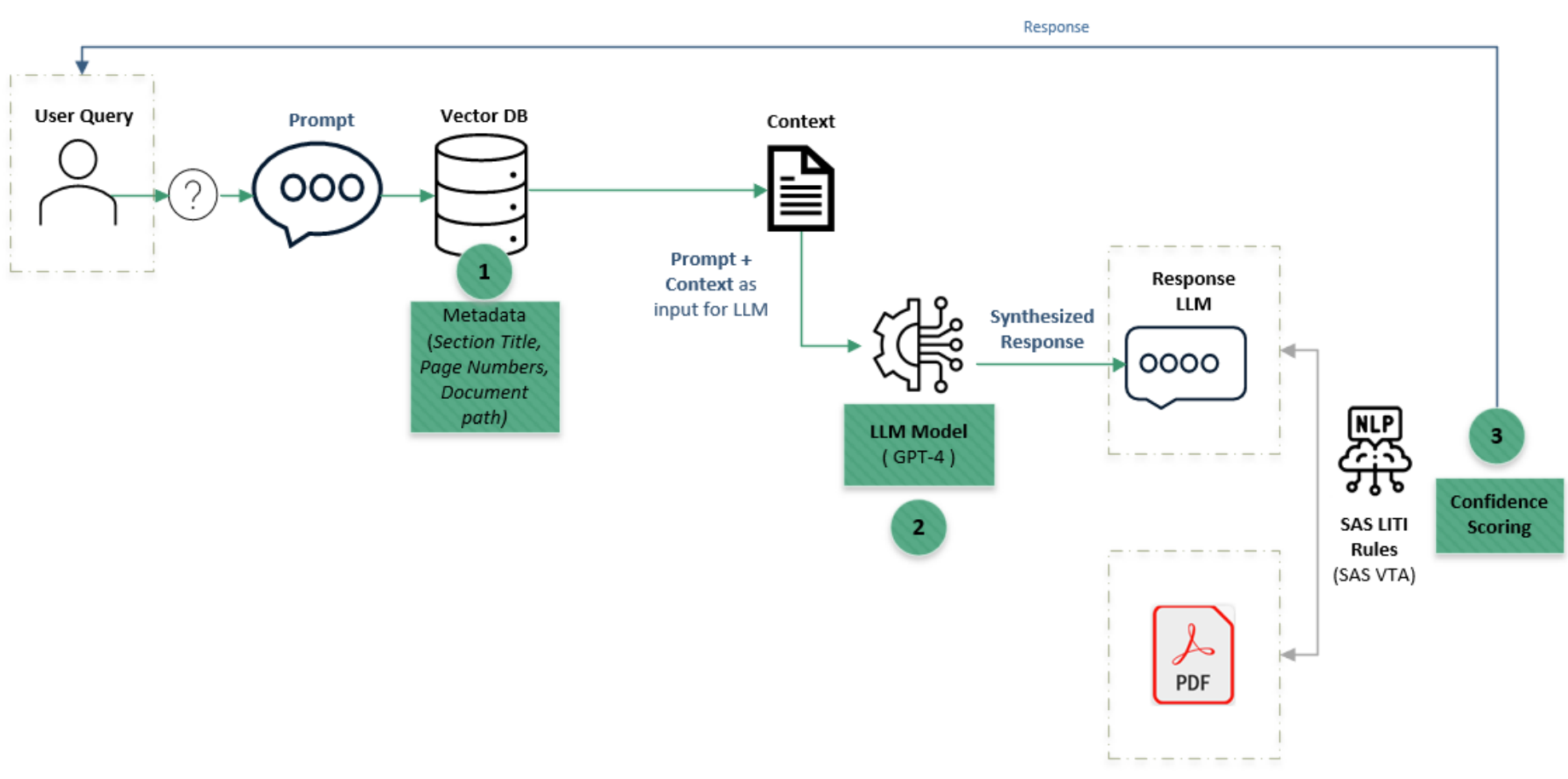
```
21
from langchain.docstore.document import Document
# Setup Model
section_pages  = []
for section in pdf_doc.sections():
    # text
    text = section.to_text(include_children=True, recurse=True)
    # Page Number
    page_id = section.page_idx
    section_title = section.title
    print(f"page:{page_id}, section:{section_title}")
#     print(f"context:{text}")
    page = Document(page_content=text,metadata={'page':page_id,'section':section_title})
    section_pages.append(page)

print(len(section_pages))
```

```
page:18, section:6.1. Objectives
page:18, section:6.1.1. Primary Objective
page:18, section:6.1.2. Secondary Objectives
page:18, section:Endpoints
page:18, section:6.2.1. Primary Endpoint
page:18, section:6.2.2. Secondary Endpoints
page:19, section:7. INVESTIGATIONAL PLAN
page:19, section:7.1. Overall Study Design
page:19, section:7.2. Number of Subjects
page:19, section:7.3. Treatment Assignment
page:20, section:7.4. Dose Adjustment Criteria
page:20, section:7.5. Criteria for Termination of Study
page:24, section:Study Period Screen inga
page:24, section:Study Day -28 to -2
page:24, section:b
page:26, section:8.1. Inclusion Criteria
page:27, section:8.2. Exclusion Criteria
page:29, section:8.3. Subject Withdrawal Criteria
page:30, section:9. TREATMENT OF SUBJECTS
page:30, section:9.1. Description of IMP
page:30, section:9.2. Restrictions
```

POC 2.0 - Accuracy Enhancements

LLM (Open AI**) + SAS LITI (**NLP + Confidence Scoring**)**




```
54    os.environ["AZURE_OPENAI_API_VERSION"] = "2024-02-01"
55    os.environ["CHATGPT_MODEL"] = "gpt-4-endpoint"
```

POC 2.0 - Accuracy Enhancements

Result

Example << primary / secondary endpoint >>

Endpoints

6.2.1. Primary Endpoint

- Percentage reduction in total IgG levels, compared to baseline, at day 29 (week 4), 7 days after the fourth IV or SC administration of efgartigimod

6.2.2. Secondary Endpoints

- Percentage reduction in total IgG levels at all other assessment timepoints as of week 4
- Percentage reduction in levels of IgG subtypes (IgG1, IgG2, IgG3, and IgG4) at all assessment timepoints
- Absolute values and changes from baseline in total IgG levels and levels of IgG subtypes (IgG1, IgG2, IgG3, and IgG4) at all assessment timepoints
- AUEC for percentage reduction in total IgG levels and for each subtype per weekly interval after each dose (week 1, week 2, week 3, and week 4), over the interval week 1 to week 4, and over the entire study period (week 1 to week 11)
- Serum levels of efgartigimod and derived PK parameters
- Clinical laboratory evaluations, vital sign measurements, ECG recordings, and incidence and characterization of TEAEs

```
query_text = "List all Primary and Secondary endpoints of the study"

context = retriever.get_relevant_documents(query_text)
prompt = prompt_template.format(context=context, question=query_text)

with get_openai_callback() as cb:
    response = llm.invoke(prompt)
    print(f"Total Cost (USD): ${format(cb.total_cost, '.6f')}")
    print(f"Model: {response.response_metadata['model_name']}, Response Metadata: {response.response_metadata['token_usage']}")
    print(response.content)
#     print(response)
```

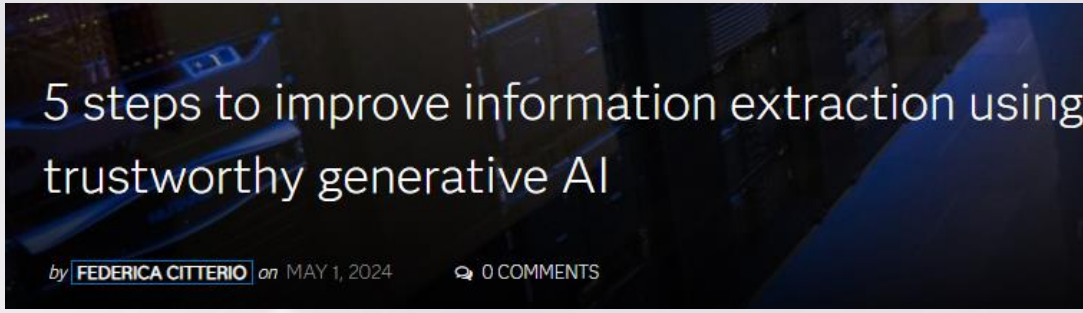
```
Total Cost (USD): $1.602420
Model: gpt-4, Response Metadata: {'completion_tokens': 256, 'prompt_tokens': 52902, 'total_tokens': 53158}
Primary Endpoint:
- Percentage reduction in total IgG levels, compared to baseline, at day 29 (week 4), 7 days after the fourth IV or SC administration of ef
gartigimod.

Secondary Endpoints:
- Percentage reduction in total IgG levels at all other assessment timepoints as of week 4.
- Percentage reduction in levels of IgG subtypes (IgG1, IgG2, IgG3, and IgG4) at all assessment timepoints.
- Absolute values and changes from baseline in total IgG levels and levels of IgG subtypes (IgG1, IgG2, IgG3, and IgG4) at all assessment t
imepoints.
- Area under the effect curve (AUEC) for percentage reduction in total IgG levels and for each subtype per weekly interval after each dose
(week 1, week 2, week 3, and week 4), over the interval week 1 to week 4, and over the entire study period (week 1 to week 11).
- Serum levels of efgartigimod and derived PK parameters.
- Clinical laboratory evaluations, vital sign measurements, ECG recordings, and incidence and characterization of TEAEs.
```

$1.6 * 20 (calls/day) * 3 (days/week) * 26 (weeks/year) * 10 (users) = $24,960

POC 2.0 - Accuracy Enhancements

Confidence Scoring (SAS Viya VTA – LITI)



“Step 5: Rule inference and confidence score calculation”

…the same corpus that was used to extract information with the LLM can be scored against the LITI rule to check if the information is actually present. A confidence score is calculated based on the number of relevant terms matched by the rule.

- Confidence is a number between 0 and 1, where a higher value indicates a better quality of the extraction.
- The confidence score decreases if the LLM model hallucinates undesirable effects that were not present in the input document.

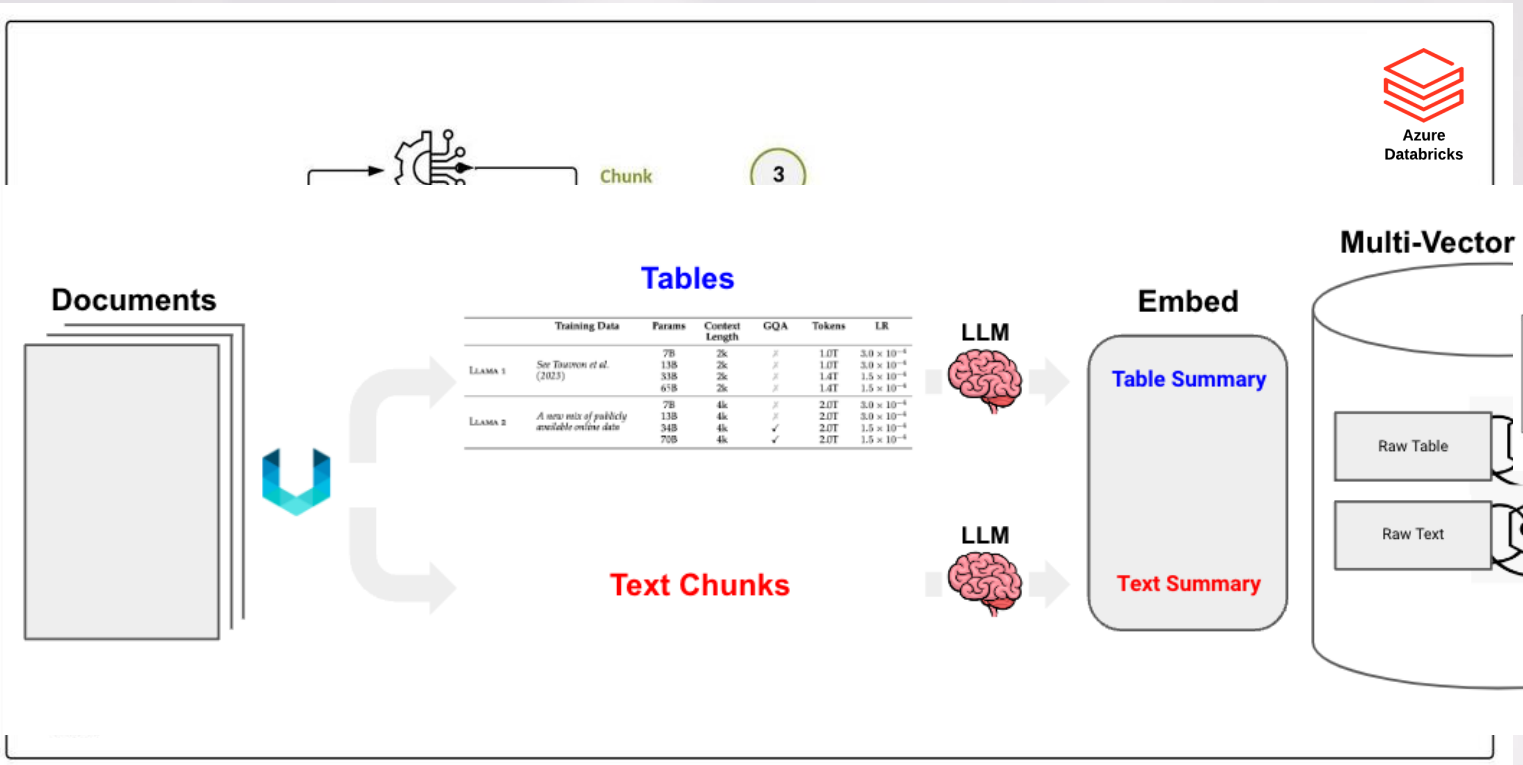
As a result, this methodology can also be used to evaluate different LLM models or different prompts by providing a quantifiable and robust metric of accuracy.

https://blogs.sas.com/content/subconsciousmusings/2024/05/01/five-steps-to-improve-information-extraction-using-trustworthy-generative-ai/

Agenda

- 1. Background to the project
- 2. The initial Journey
- 3. POC 2.0
- **4. POC 3.0 - Financial Considerations**
- 5. The Journey continues... ?

Financial Considerations - Accuracy Enhancements & Cost Optimizations




```
query_text = "List all Primary and Secondary endpoints of the study"

context = retriever.get_relevant_documents(query_text)
prompt = prompt_template.format(context=context, question=query_text)

with get_openai_callback() as cb:
    response = llm.invoke(prompt)
    print(f"Total Cost (USD): ${format(cb.total_cost, '.6f')}")
    print(f"Model: {response.response_metadata['model_name']}, Response Metadata: {response.response_metadata['token_usage']}")
    print(response.content)
#     print(response)
```

```
Total Cost (USD): $1.602420
Model: gpt-4, Response Metadata: {'completion_tokens': 256, 'prompt_tokens': 52902, 'total_tokens': 53158}
Primary Endpoint:
- Percentage reduction in total IgG levels, compared to baseline, at day 29 (week 4), 7 days after the fourth IV or SC administration of ef
gartigimod.

Secondary Endpoints:
- Percentage reduction in total IgG levels at all other assessment timepoints as of week 4.
- Percentage reduction in levels of IgG subtypes (IgG1, IgG2, IgG3, and IgG4) at all assessment timepoints.
- Absolute values and changes from baseline in total IgG levels and levels of IgG subtypes (IgG1, IgG2, IgG3, and IgG4) at all assessment t
imepoints.
- Area under the effect curve (AUEC) for percentage reduction in total IgG levels and for each subtype per weekly interval after each dose
(week 1, week 2, week 3, and week 4), over the interval week 1 to week 4, and over the entire study period (week 1 to week 11).
- Serum levels of efgartigimod and derived PK parameters.
- Clinical laboratory evaluations, vital sign measurements, ECG recordings, and incidence and characterization of TEAEs.
```

```
12:15 PM (9s)                                                    30                                                    Python

query_text = "List all Ednpoints - Primary and Secondary endpoints of the study"

context = retriever.invoke(query_text)
prompt = prompt_template.format(context=context, question=query_text)


with get_openai_callback() as cb:
    response = model.invoke(prompt)
    print(f"Total Cost (USD): ${format(cb.total_cost, '.6f')}")
    print(f"Model: {response.response_metadata['model_name']}, Response Metadata: {response.response_metadata['token_usage']}")
    print(response.content)
#     print(response)
```

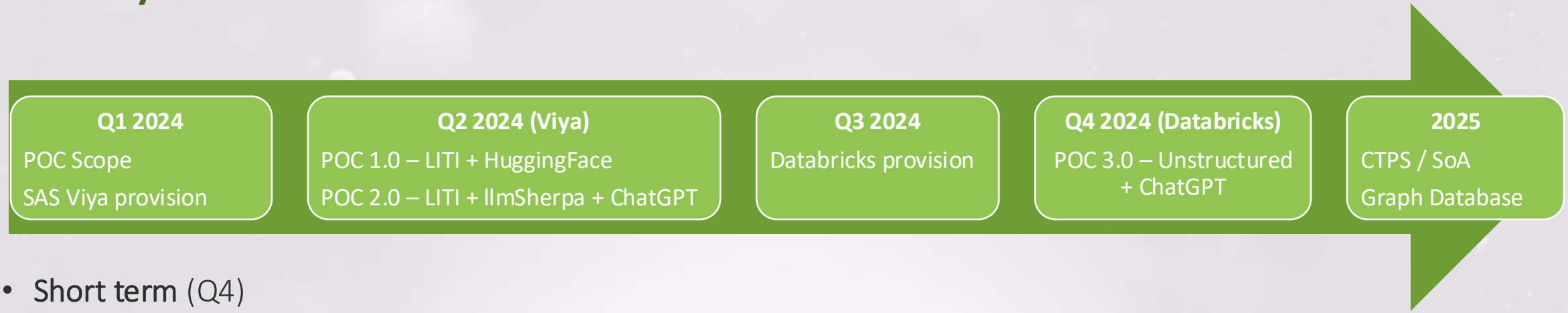
```
Total Cost (USD): $0.170640
Model: gpt-4, Response Metadata: {'completion_tokens': 256, 'prompt_tokens': 5176, 'total_tokens': 5432, 'completion_tokens_details': None, 'prompt_tokens_details': None}
Primary Endpoint:
- Percentage reduction in total IgG levels, compared to baseline, at day 29 (week 4), 7 days after the fourth IV or SC administration of efgartigimod.

Secondary Endpoints:
- Percentage reduction in total IgG levels at all other assessment timepoints as of week 4.
- Percentage reduction in levels of IgG subtypes (IgG1, IgG2, IgG3, and IgG4) at all assessment timepoints.
- Absolute values and changes from baseline in total IgG levels and levels of IgG subtypes (IgG1, IgG2, IgG3, and IgG4) at all assessment timepoints.
- Area under the effect curve (AUEC) for percentage reduction in total IgG levels and for each subtype per weekly interval after each dose (week 1, week 2, week 3, and week 4), over t
he interval week 1 to week 4, and over the entire study period (week 1 to week 11).
- Serum levels of efgartigimod and derived PK parameters.
- Clinical laboratory evaluations, vital sign measurements, ECG recordings, and incidence and characterization of TEAEs.
```

Agenda

- 1. Background to the project
- 2. The initial Journey
- 3. POC 2.0
- 4. POC 3.0 - Financial Considerations
- **5. The Journey continues... ?**

Journey Continues……



- Short term (Q4)
  - Leverage Open AI service (commercial models)
  - Explore multi-modal model approach
  - Reflect on POC and Define Next Steps

- Mid term (2025)
  - focus on CTPS (concept sheet)
  - Schedule of Assessments (SoA)

- Longer term - protocol builds
  - Cost
  - Technology (no off-the shelf available yet?), we are biotech … not software builders
  - Change management
  - New roles

Thank You