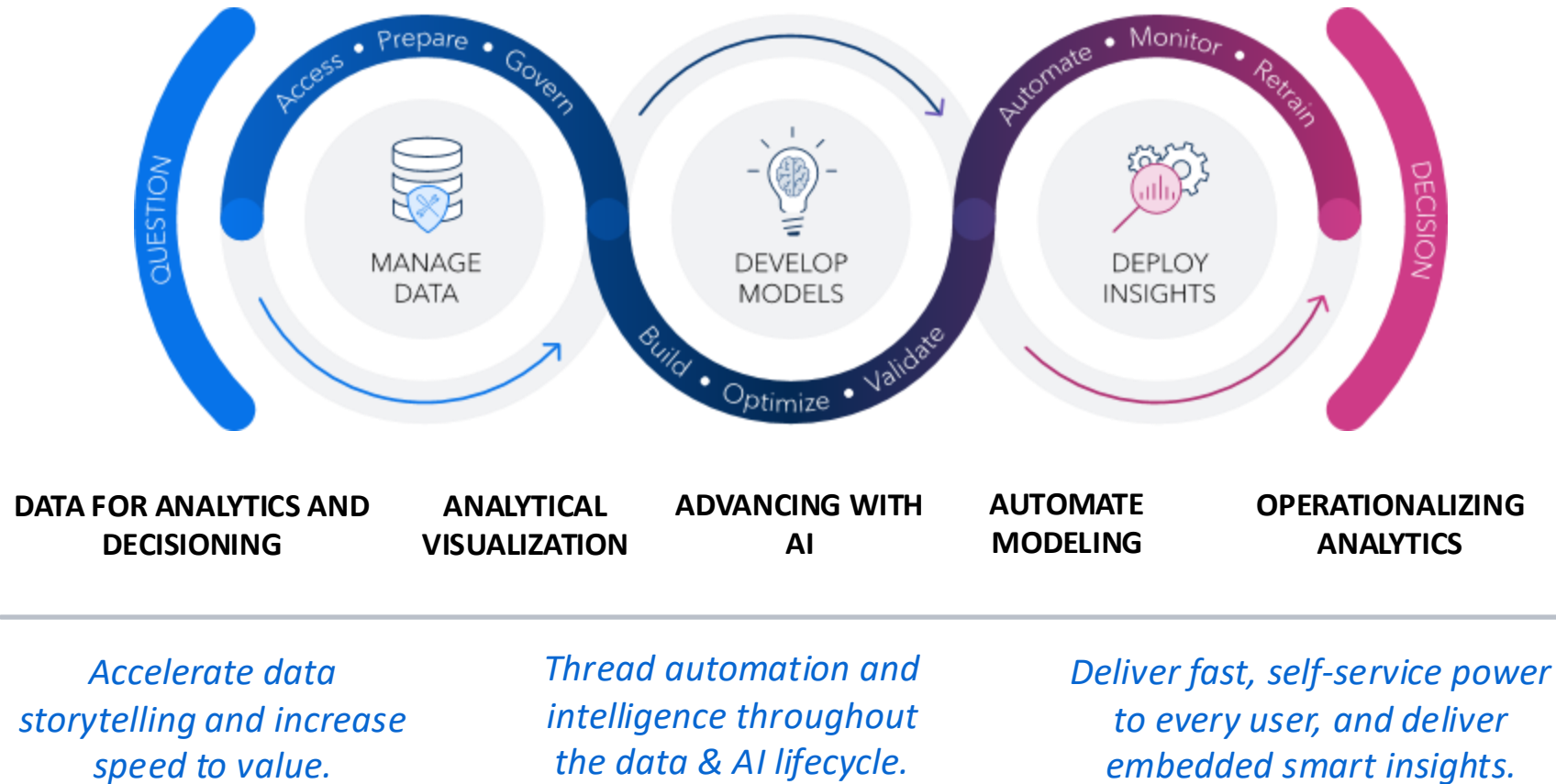


Harnessing Generative AI to Transform Clinical Trials

Soundarya Palanisamy, Sr Industry Advisor, Global Health and Life Sciences Practice, SAS Institute

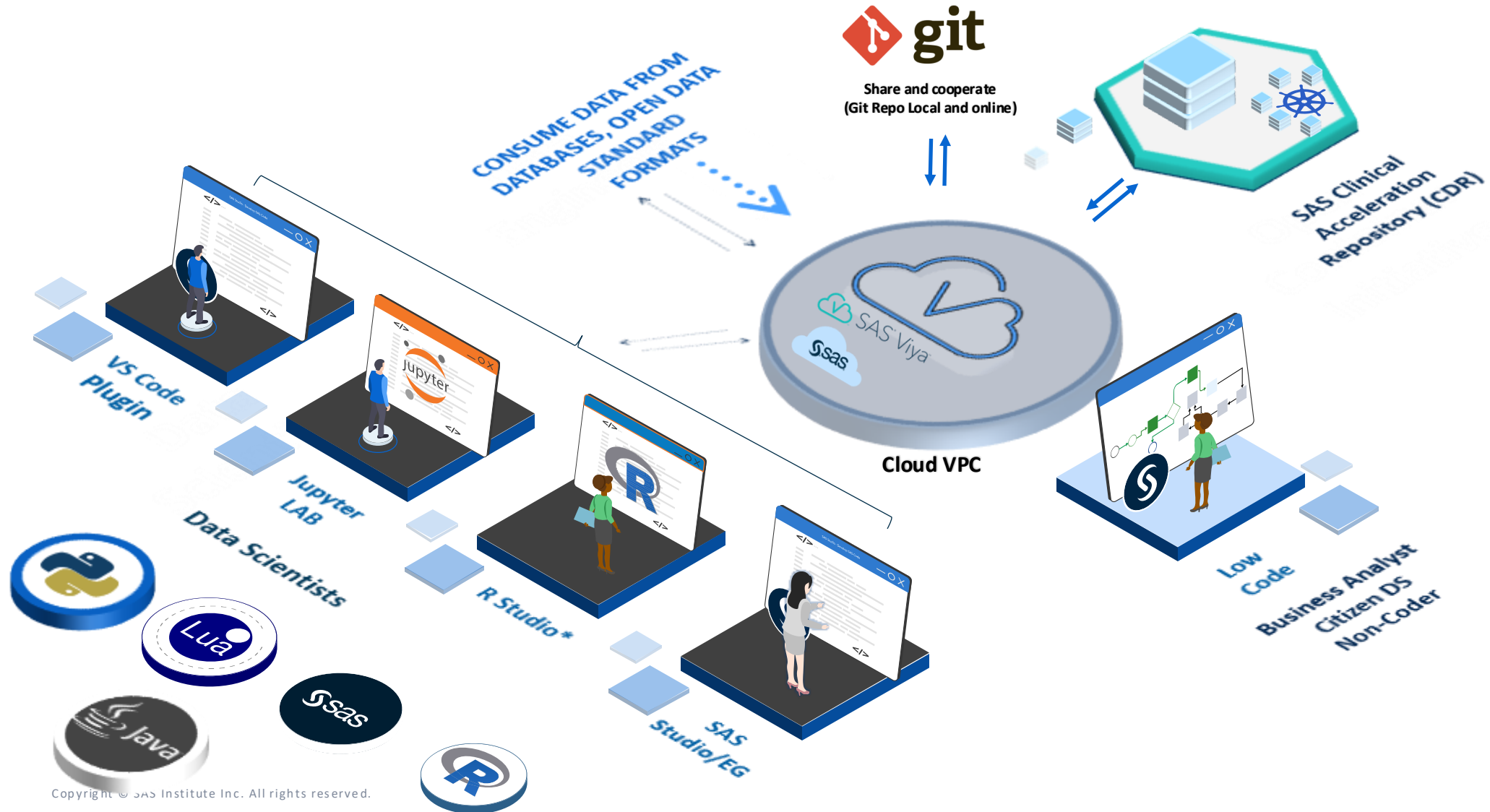


Viya is the Latest SAS Platform That Connects All Aspects of the Data and AI Lifecycle



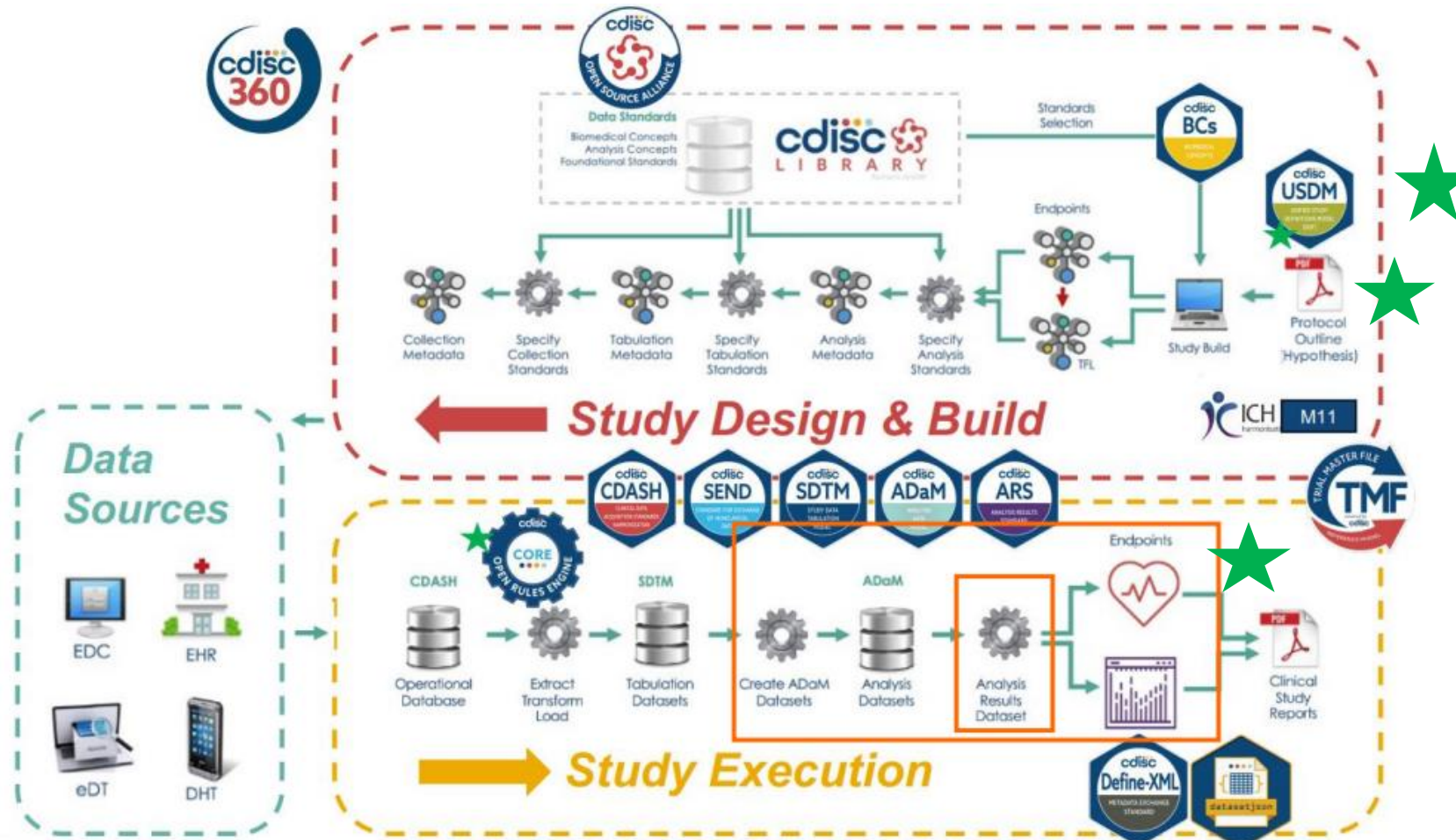
SAS Viya

SAS Viya is the latest release from SAS. It is your one-stop-shop for all your analytic needs. Think of it as suites of product that does things like data preparation, data visualization, model building, Programming (SAS\open-source) etc., all running on a very powerful in-memory & SAS 9 compute engine.



Case Study

CDISC 360 Vision



Case Study: Development of USDM through translation of human-readable protocols



Development of USDM through translation of human-readable protocols

Jasmine Kestemont (Innovion)
Stijn Rogiers (argenx)

POC approach (Cont'd)

- 1) **Usage of LITI rules** for contextual information extraction
LITI includes concept rule types as well as fact rule types.
LITI is proprietary syntax from SAS.
LITI = **L**anguage **I**nterpretation for **T**extual **I**nformation
- 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.*



2024 Europe CDISC+TMF Interchange | #ClearDataClearImpact

16



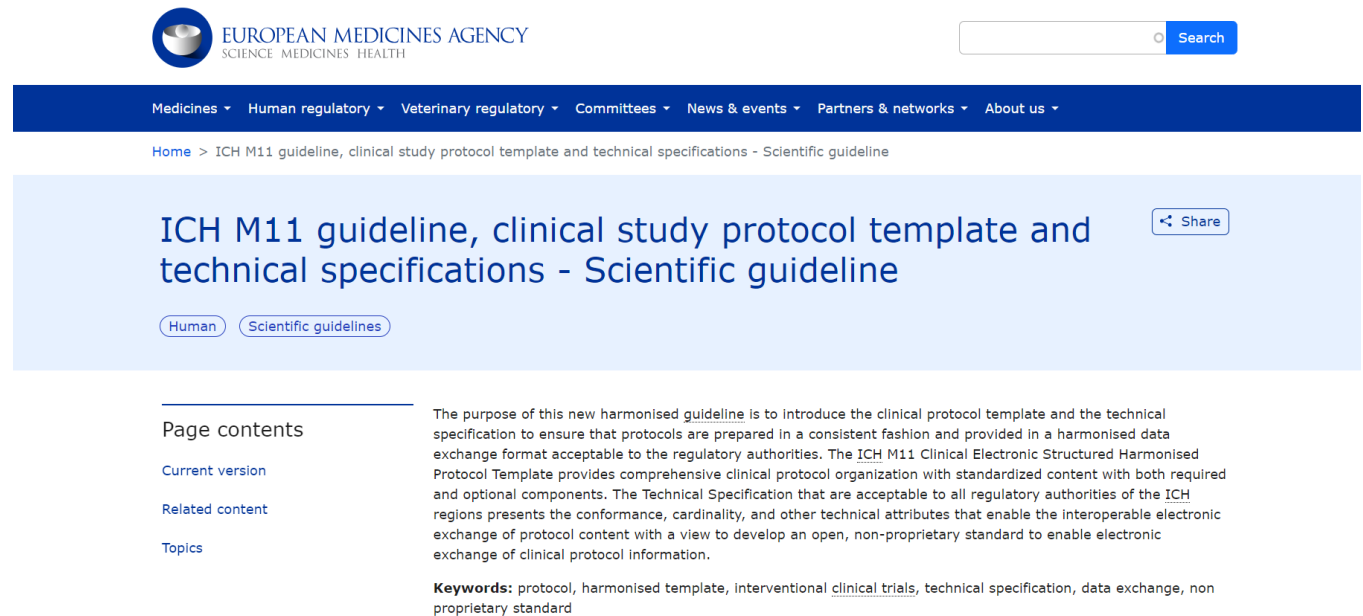
Motivation and background

Need to standardize protocols

[ICH M11 guideline, clinical study protocol template and technical specifications - Scientific guideline | European Medicines Agency \(europa.eu\)](https://www.ema.europa.eu/en/ich-m11-guideline-clinical-study-protocol-template-and-technical-specifications-scientific-guideline)

Explore new technology

ICH M11 guideline



The screenshot shows the official ICH M11 guideline page from the European Medicines Agency (EMA). The header includes the EMA logo and navigation links. The main title is "ICH M11 guideline, clinical study protocol template and technical specifications - Scientific guideline". Below the title, there are tabs for "Human" and "Scientific guidelines", with "Scientific guidelines" being the active tab. The page content describes the purpose of the guideline, which is to introduce a harmonized clinical protocol template and technical specification to ensure consistency and interoperability in clinical trials. The text mentions that the guideline provides comprehensive clinical protocol organization with standardized content and optional components, acceptable to all regulatory authorities of the ICH regions. It also includes a list of keywords: protocol, harmonised template, interventional clinical trials, technical specification, data exchange, non proprietary standard.

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Medicines ▾ Human regulatory ▾ Veterinary regulatory ▾ Committees ▾ News & events ▾ Partners & networks ▾ About us ▾

Home > ICH M11 guideline, clinical study protocol template and technical specifications - Scientific guideline

ICH M11 guideline, clinical study protocol template and technical specifications - Scientific guideline

Human Scientific guidelines

Page contents

- Current version
- Related content
- Topics

The purpose of this new harmonised guideline is to introduce the clinical protocol template and the technical specification to ensure that protocols are prepared in a consistent fashion and provided in a harmonised data exchange format acceptable to the regulatory authorities. The ICH M11 Clinical Electronic Structured Harmonised Protocol Template provides comprehensive clinical protocol organization with standardized content with both required and optional components. The Technical Specification that are acceptable to all regulatory authorities of the ICH regions presents the conformance, cardinality, and other technical attributes that enable the interoperable electronic exchange of protocol content with a view to develop an open, non-proprietary standard to enable electronic exchange of clinical protocol information.

Keywords: protocol, harmonised template, interventional clinical trials, technical specification, data exchange, non proprietary standard

AIM – Perform contextual extraction from clinical trial protocols and store relevant information into the USDM workbook in the right place (in the right field)

studyDesignOE Sheet												
	A	B	C	D	E	F	G	H	I	J	K	L
1	objectiveName	objectiveDescription	objectiveLabel	objectiveText	objectiveLevel	objectiveDictionary	endpointName	endpointDescription	endpointLabel	endpointText	endpointPurpose	endpointLevel
2	OB1	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	OB2	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 Improvement		Time to improvement of at least 2 categories relative to baseline on a 7-category ordinal scale of clinical statu		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				
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
13	Active	EL1	EL2	EL4, EL5
14	Placebo	EL1	EL2	EL5, EL3

studyDesignArms Sheet					
	A	B	C	D	E
1	name	description	label	type	dataOriginDescription
2	Active	Active Substance	Active Substance	Active Comparator Arm	Data collected from subjects
3	Placebo	Placebo	Placebo	Placebo Comparator Arm	Data collected from subjects

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

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}

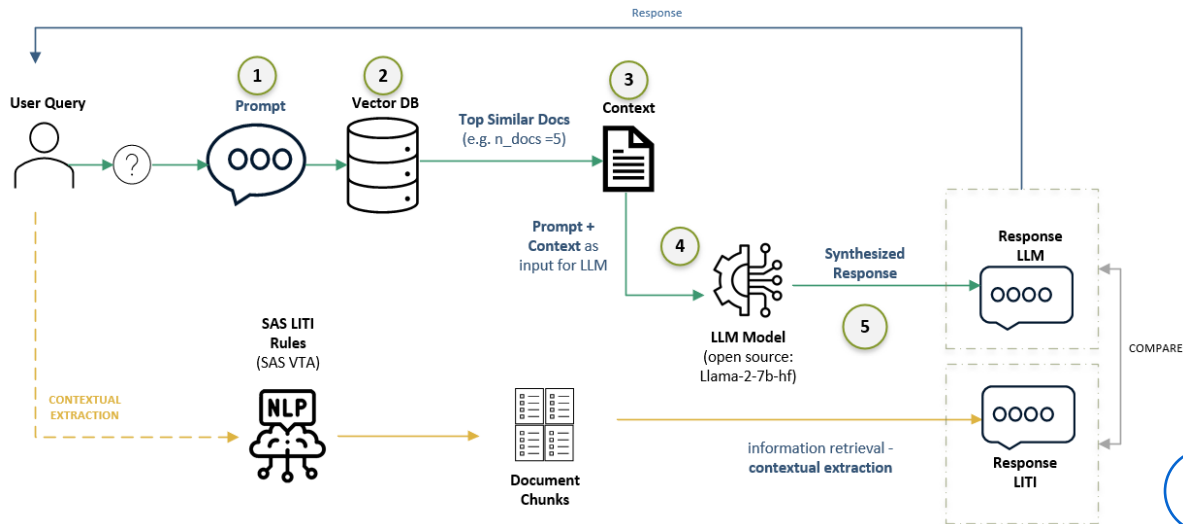
Primary Endpoint:

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

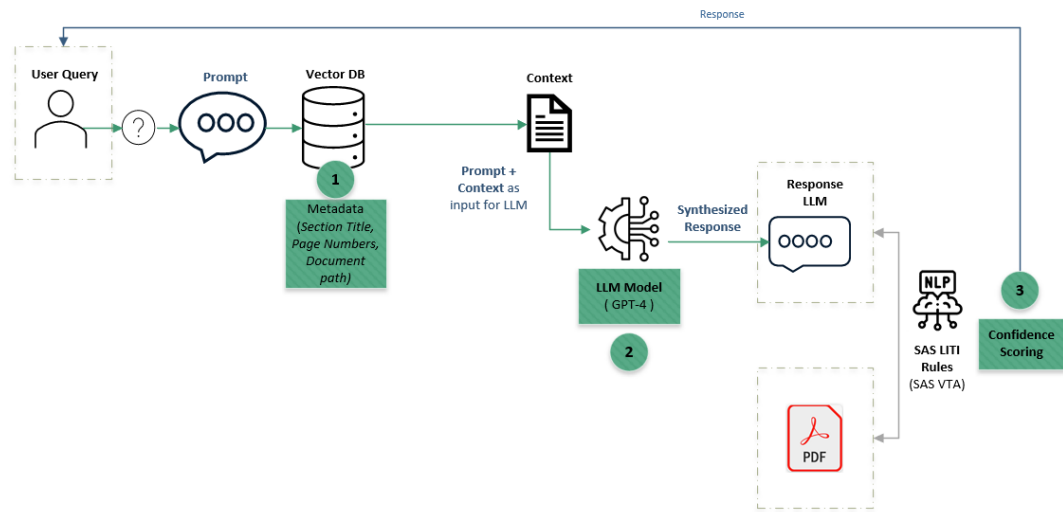
Secondary Endpoints:

- 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 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 SAS LITI (NLP) + LLM for contextual extraction



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



3 Evaluating Gen AI and Text Analytics for creating a USDM

Some initial results after pre processing, highlights the extractions of **inclusion** and **exclusion criteria** by LITI rules and LLMs.

Legend :

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

Select any image to see a larger version.
Mobile users: To view the images, select the "Full" version at the bottom of the page.

		INCLUSION CRITERIA			EXCLUSION CRITERIA		
	document partitioning (training / validation / test)						
Documents (clinical trial protocols)		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*
Ell Lilly_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

There are several challenges to overcome, such as overfitting in LITI rules and the need for extensive pre-processing.

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)

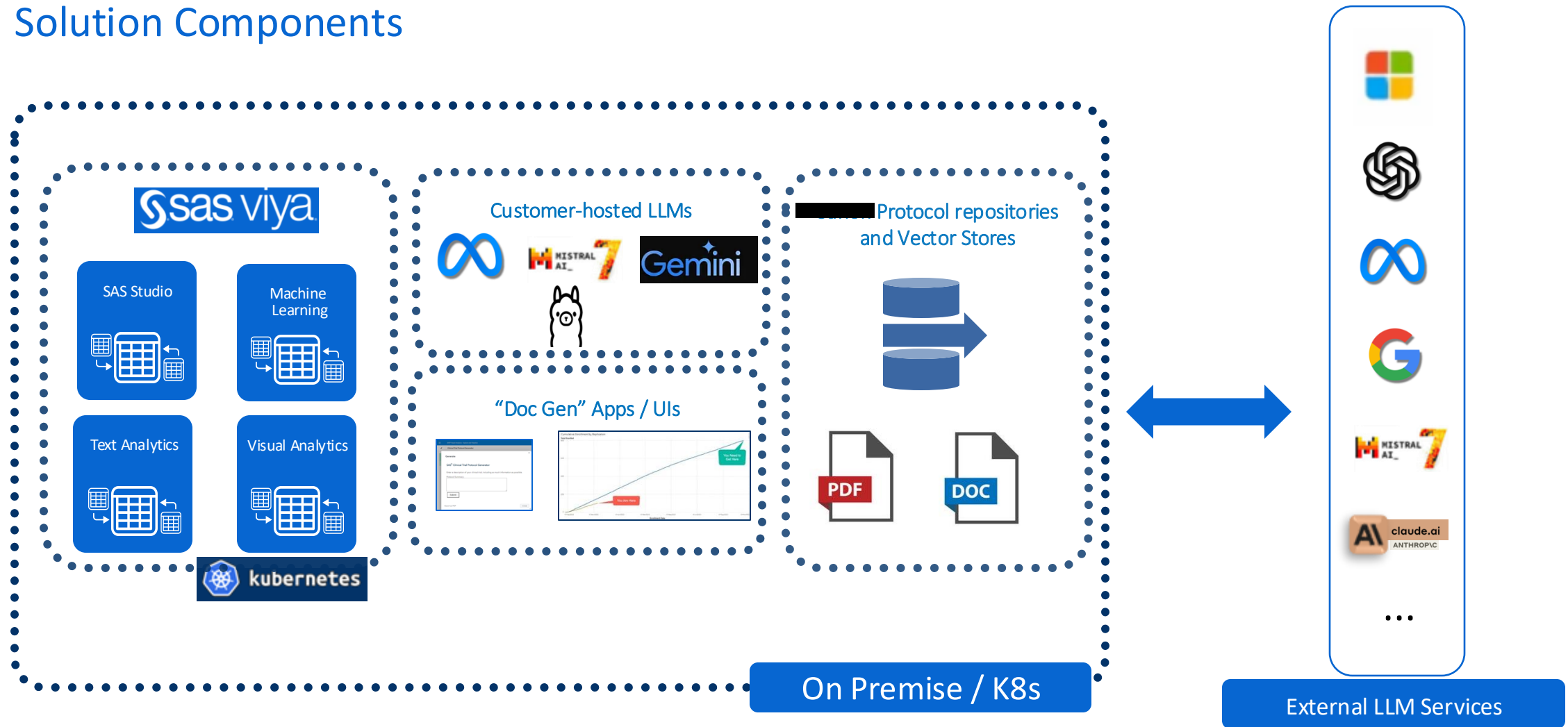
- It's generative AI.
- Knowledge required to do proper prompt engineering (designing the user query). But much less effort required. It's just an API-call to a pre-trained model.
- If the context around a particular topic can vary widely from document to document, there is a clear advantage.
- It's 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.

Case Study: AI-Powered Clinical Trial Protocol Benchmarking, Simulation, Development



This is the house that we built.

Solution Components



Note: Icons representing LLMs are purely indicative and not recommendations



Enter initial study characteristics (incl. objectives and endpoints)

Enter Initial Study Characteristics

SAS® Visual Analytics - Explore and Visualize

Protocol_Initial_Exploration

Opened reports (1)

All Protocols

Protocols_Structured

Structured Protocols

Study Characteristics

Filters: No selections

Provide Route of Administration

route

Provide Location of Trial

location

Select Trial Characteristics

☐ Comparator-controlled

☐ Modified double blind

☐ Multi-center

☐ Multi-country

☐ Observer-blind

☐ Open-label

Nbr of Trials

39

fileName	fileDate	fileType
NCT05568797_Prd SAS® Visual Analytics	Sunday, 15 September 2024, 12:36:54	txt
NCT05559476_Prot_000.txt	Sunday, 15 September 2024, 12:36:56	txt
NCT05144945_Prot_000.txt	Sunday, 15 September 2024, 12:36:56	txt
NCT05062759_Prot_000.txt	Sunday, 15 September 2024, 12:36:56	txt
NCT05050318_Prot_SAP_000.txt	Sunday, 15 September 2024, 12:36:54	txt
NCT05045144_Prot_000.txt	Sunday, 15 September 2024, 12:36:56	txt
NCT05044195_Prot_000.txt	Sunday, 15 September 2024, 12:36:55	txt
NCT05028361_Prot_003.txt	Sunday, 15 September 2024, 12:36:56	txt
NCT04622592_Prot_000.txt	Sunday, 15 September 2024, 12:36:56	txt
NCT04576702_Prot_000.txt	Sunday, 15 September 2024, 12:36:56	txt
NCT04551677_Prot_SAP_000.txt	Sunday, 15 September 2024, 12:36:56	txt
NCT04537234_Prot_000.txt	Sunday, 15 September 2024, 12:36:56	txt

Identify
similar studies
(internal and
external)



Identify Similar Studies

The screenshot displays the SAS Studio interface for developing code and flows. The top navigation bar includes 'New', 'Options', 'View', 'Open', and 'Save All'. The main workspace is divided into three panels:

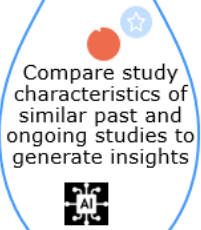
- Left Panel (Flow Canvas):** Shows a list of flow steps under the 'Flow' tab. The steps are:
 - Naive RAG with entire document corpus
 - Retrieval Augmented Generation - summarise cli
 - Retrieval Augmented Generation - summarise st
 - Retrieval Augmented Generation - identify similarThe fourth step is selected, showing a diagram with two nodes: 'LLM - Azure OpenAI RAG' and 'LLMOUTPUT'.
- Right Panel (Submitted Code and Results):** Displays a table with columns 'Code' and 'Log'. The table contains one row with the number '1'.
- Bottom Panel:** A large empty area for the flow canvas, with a green mouse cursor pointing at it.

On the flow canvas, select a node to view its details.

SAS® Studio

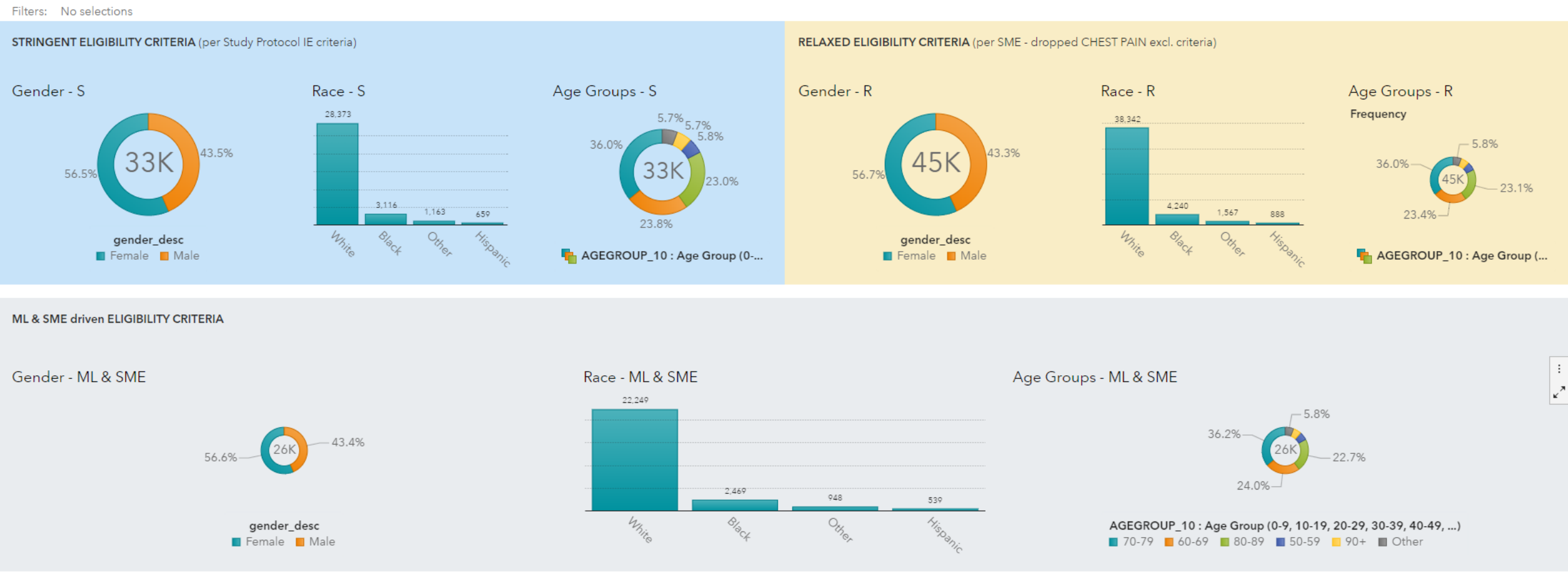
File: /Public/Protocols/code/Protocol_Data_RAG_Flow.flw

Line 1 Column 1 | Recover (4) | Submission (0)



Compare Study Characteristics

Baseline Characteristics (Stringent Vs. Relaxed Vs. ML/SME) : Geographical Distribution (Stringent Vs. Relaxed Vs. ML/SME) Automated Explanation Logistic Regression Decision Tree Gradient Boosting Neural Network Model Comparison





Protocol Draft Generation

SAS seamlessly integrates with LLMs to generate draft protocols

Our demo uses a custom Open AI model that is trained to write protocols based on the **ICH M11 Protocol Template**.

SAS® Visual Analytics - Explore and Visualize

SAS Viya Clinical Trial Assistant

Opened reports (1)

Generate

SAS® Clinical Trial Protocol Generator

Enter a description of your clinical trial, including as much information as possible.

Protocol Summary:

This is a randomized, open-label, multicenter, prospective study to compare the efficacy and safety of the combination of VcR-CAP to that of R-CHOP in participants who have newly diagnosed mantle cell lymphoma grade II, III or IV and who are ineligible to undergo bone marrow transplantation.

Submit

Export as PDF

Close

Generate Full Protocol Package & Other Documents (Proposed)

SAS® Visual Analytics - Explore and Visualize

85 ?

Protocol Document Generation (DEMO)

Opened reports (1)

Source Documents

Add +

☒ Alzheimer's Disease Guide... 3.2

☐ CTCAE v5.0 2.1

☒ FDA 21 CFR Part 11 0.8

☒ FDA Guidance on Alzheim... 1.1

☒ GCP Principles 0.9

☒ ICH GCP E6 R2 Guideline 1.5

☐ Informed Consent Form v3.0 0.5

☒ Phase I Protocol 5.8

☒ Phase II Protocol 8.1

☐ Protocol Template v2.0 0.3

☒ PUBMED Research Paper 7.2

☒ Study Synopsis v1.0 1.2

☒ TransCelerate Template v1.2 1.8

Edit Protocol

Protocol Generation Guide

Help me create:

→
Generate
Protocol

New
Template

Generate
Section

Search
Similar
Studies

Generate
from
Template

Summary:

This section provides an automated summary of the loaded documents and the target study, generated using advanced Large Language Models (LLMs). The summary aims to capture the key aspects of the documents and study, including objectives, methodology, and relevant details.

Suggested Actions:

Suggest optimized study design...

Identify missing information

Flag regulatory compliance issues

[View History](#)

Enter Study Details Here...

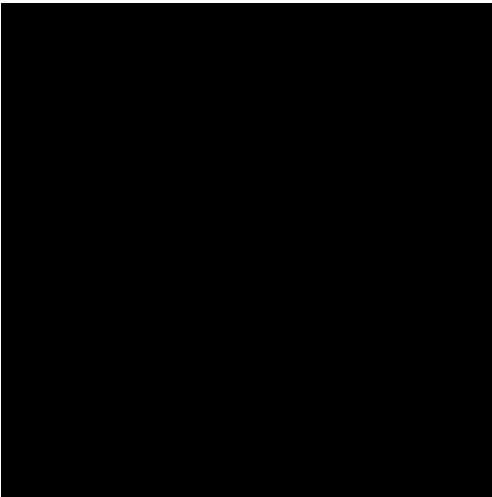
→

[Advanced](#)

Copyright © SAS Institute Inc. All rights reserved.

Sample Protocol Generation Results

A Phase III Randomized, Double-blind, Placebo-controlled Multicenter Study in Adults to Determine the Safety, Efficacy, and Immunogenicity



1.3 Relevant Clinical Data:

Phase I and II studies [REDACTED] demonstrated [REDACTED] with [REDACTED] in a Phase I study. These studies informed the dose selection and regimen for this Phase III trial.

References:

Grade	Description
1	Mild; asymptomatic or mild symptoms; clinical or diagnostic observations only; intervention not indicated.
2	Moderate; minimal, local or non-invasive intervention indicated; limiting age-appropriate instrumental activities of daily living.
3	Severe; inability to perform activities of daily living; hospitalization or prolongation of hospitalization indicated.
4	Life-threatening; urgent intervention indicated.

Relationship to Study Drug:

- The following table will be used to assess the relationship between AEs and study drug:

Relationship	Description
Unrelated	No reasonable possibility that the AE is related to study drug.
Possible	Cannot rule out a reasonable possibility that the AE is related to study drug.
Probable	A reasonable possibility that the AE is related to study drug.
Definite	A clear temporal relationship and/or known mechanism of action suggests a high probability that the AE is related to study drug.

11.2 Serious Adverse Events (SAEs)

Definition: An SAE is any AE that:

- Results in death
- Is life-threatening
- Requires hospitalization or prolongation of hospitalization
- Results in persistent or significant disability

8.3 Formulation of Test and Control Products

- AZD1222: non-replicating ChAdOx1 vector vaccine, 5×10^{10} viral particles (VP)/dose
- Placebo: sterile saline solution
- M [REDACTED]
- Formulation:

Product	Active Ingredient (mg/mL)	Other Ingredients (mg/mL)	pH
AZD1222	0.5 mg/mL	Histidine, sucrose, polysorbate 80	7.4
Placebo	-	Sodium chloride, water for injection	7.0

- Packaging and labeling: identical for AZD1222 and placebo, labeled by [REDACTED]

Label Mock-up: [Insert label mock-up]

8.4 Supply of Study Drug at Site

- Initial supply: shipped to site upon IRB approval
- Replacement kits: provided through IWRS upon notification of subject withdrawal
- Dosage/Dosage Regimen:
- Dose: 2 doses of 5×10^{10} VP/dose, administered IM on Day 1 and Day 29
- Route: intramuscular injection
- Timing: optimal timing between doses 28 days

8.5 Dispensing

- Authority: designated site pharmacist or investigator
- Requirements: verify subject eligibility, record dispensing in study records

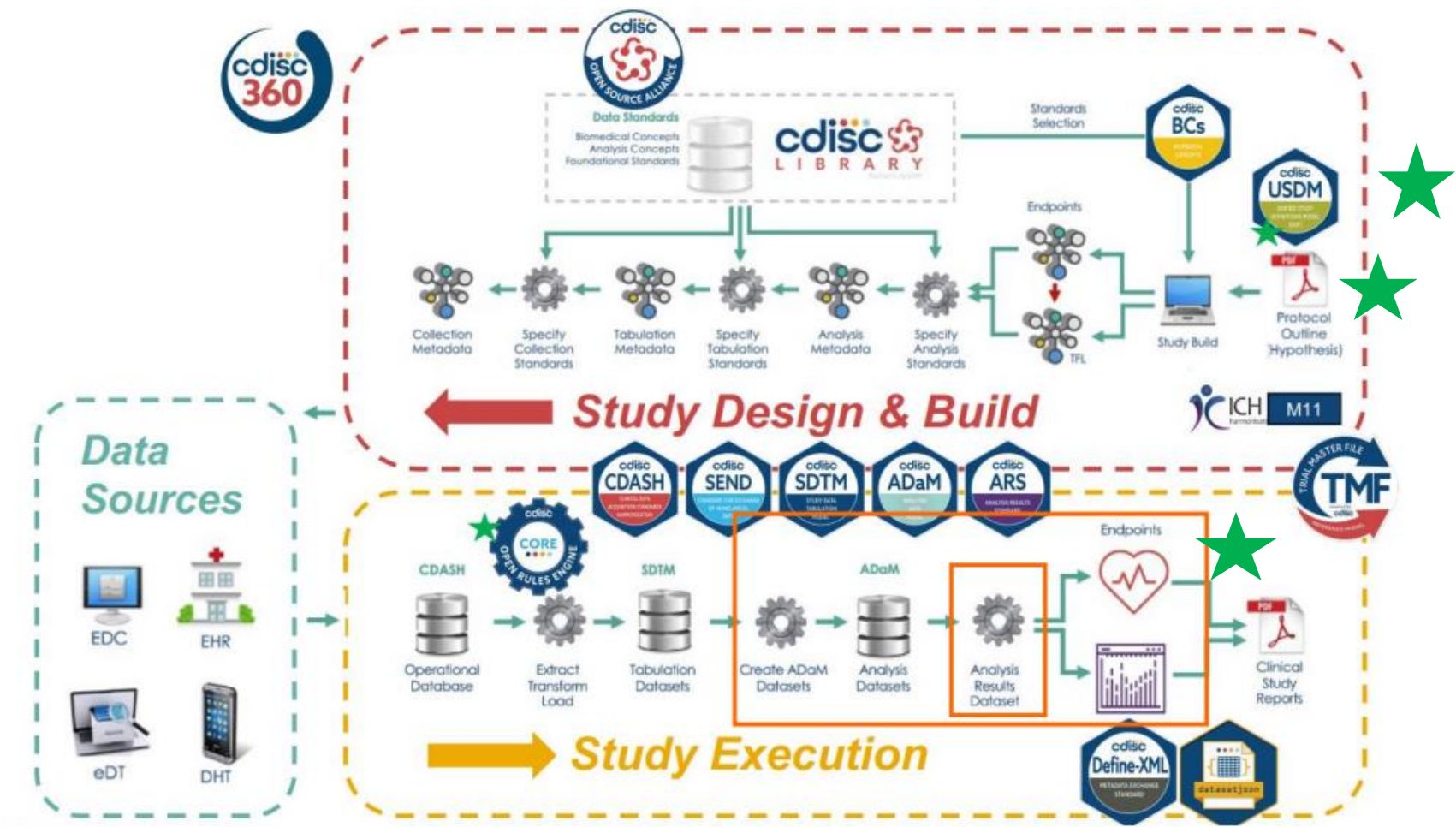
8.6 Administration Instructions

- Step-by-step instructions:
- 1. Verify subject identity and eligibility
- 2. Prepare study drug according to packaging instructions
- 3. Administer study drug via IM injection
- 4. Record administration in study records

Case Study: SAS agents for clinical trial submission

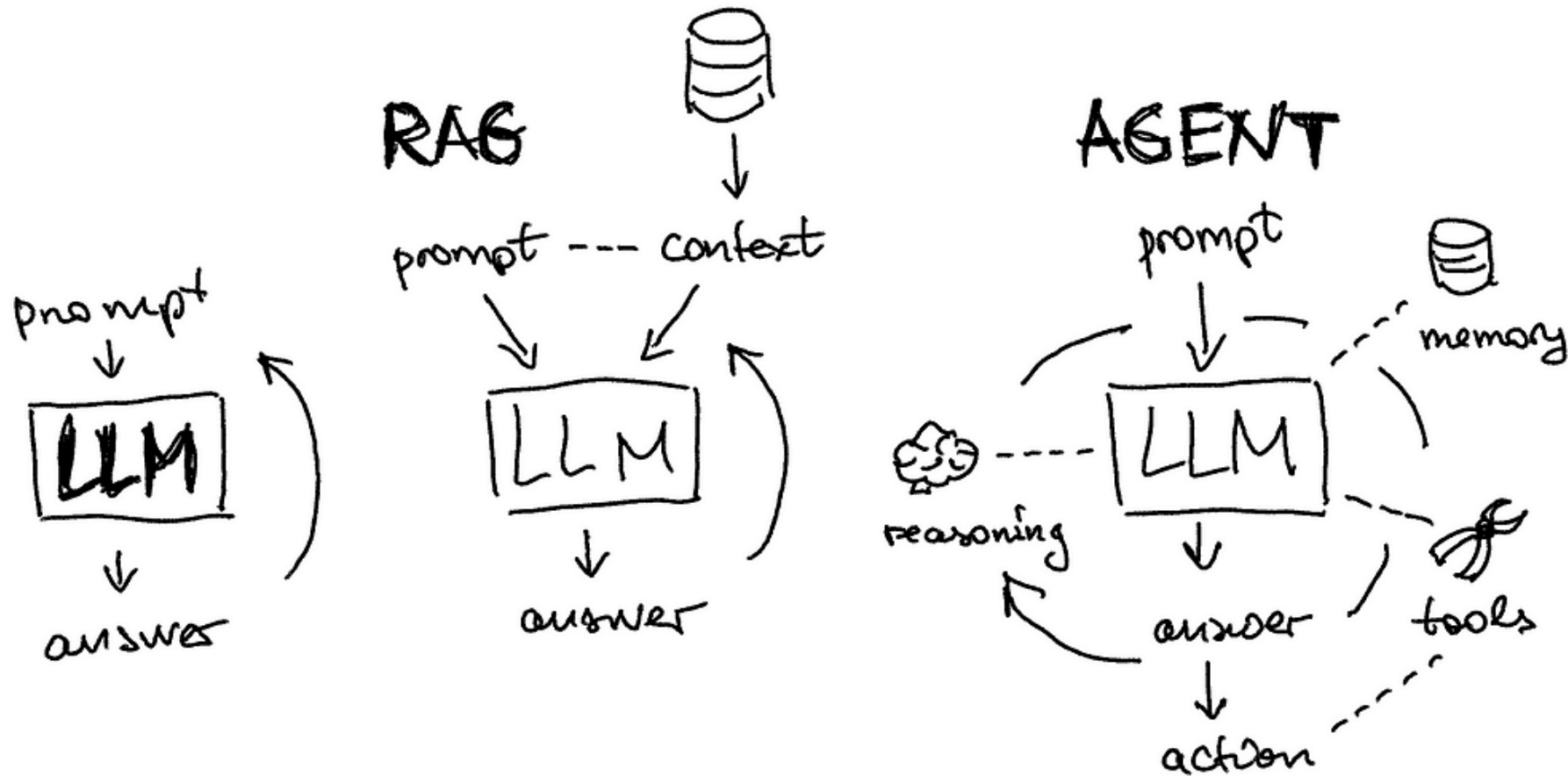


CDISC 360 Vision



The Art of Building Capable AI Assistants

From Foundational LLM to “LLM app”

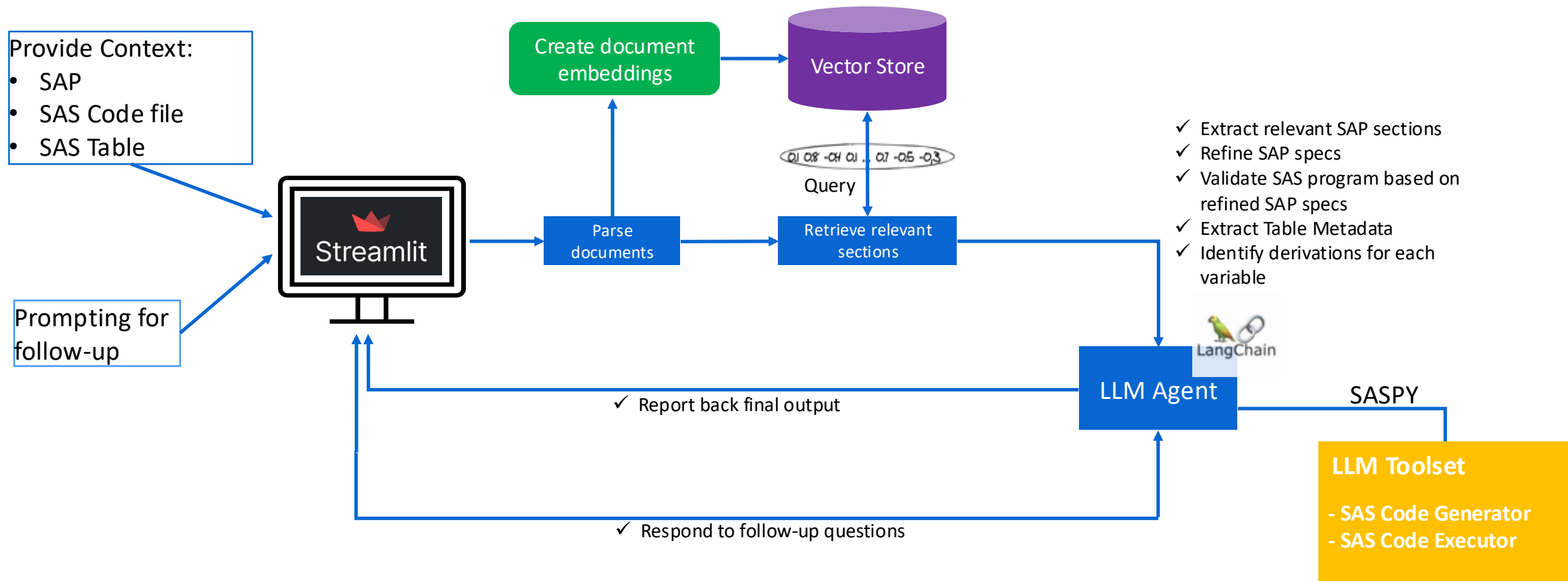


From:

[Intro to LLM Agents with Langchain: When RAG is Not Enough](#) | by Alex Honchar | Mar, 2024 | Towards Data Science

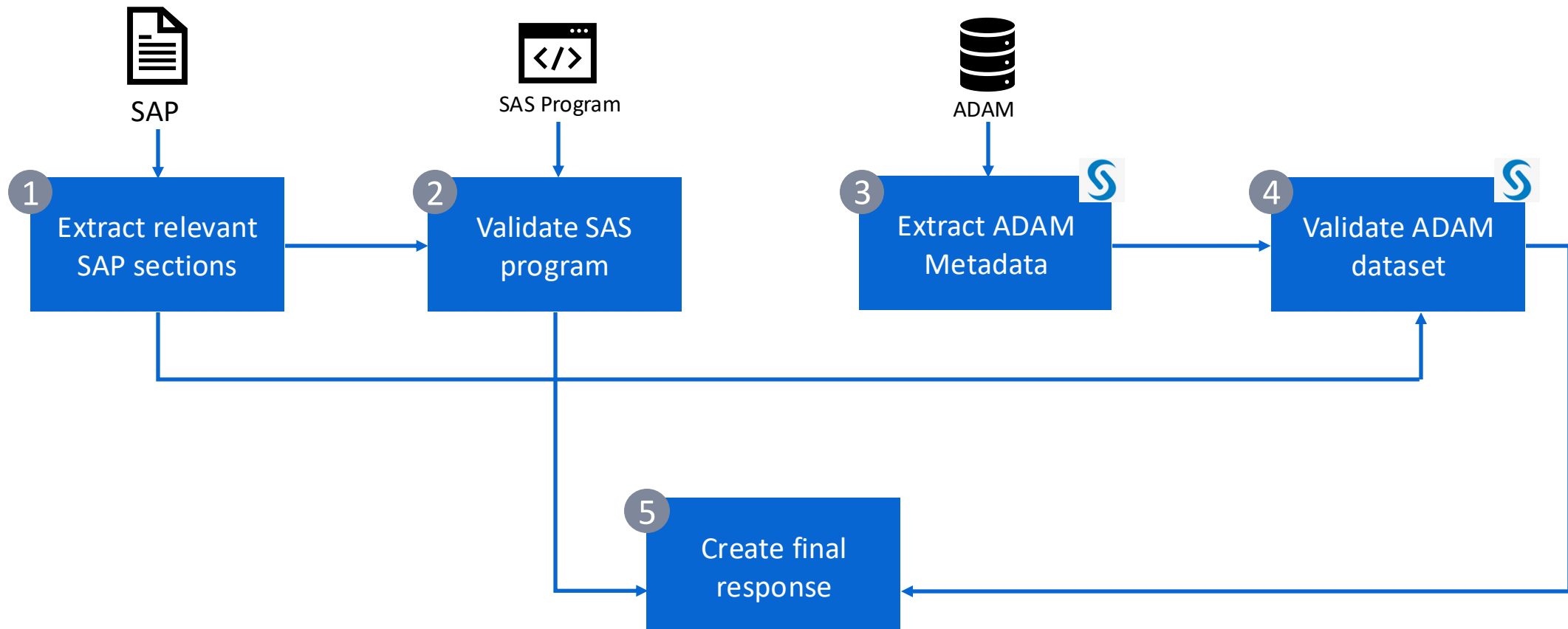
Copyright © SAS Institute Inc. All rights reserved.

High Level Design



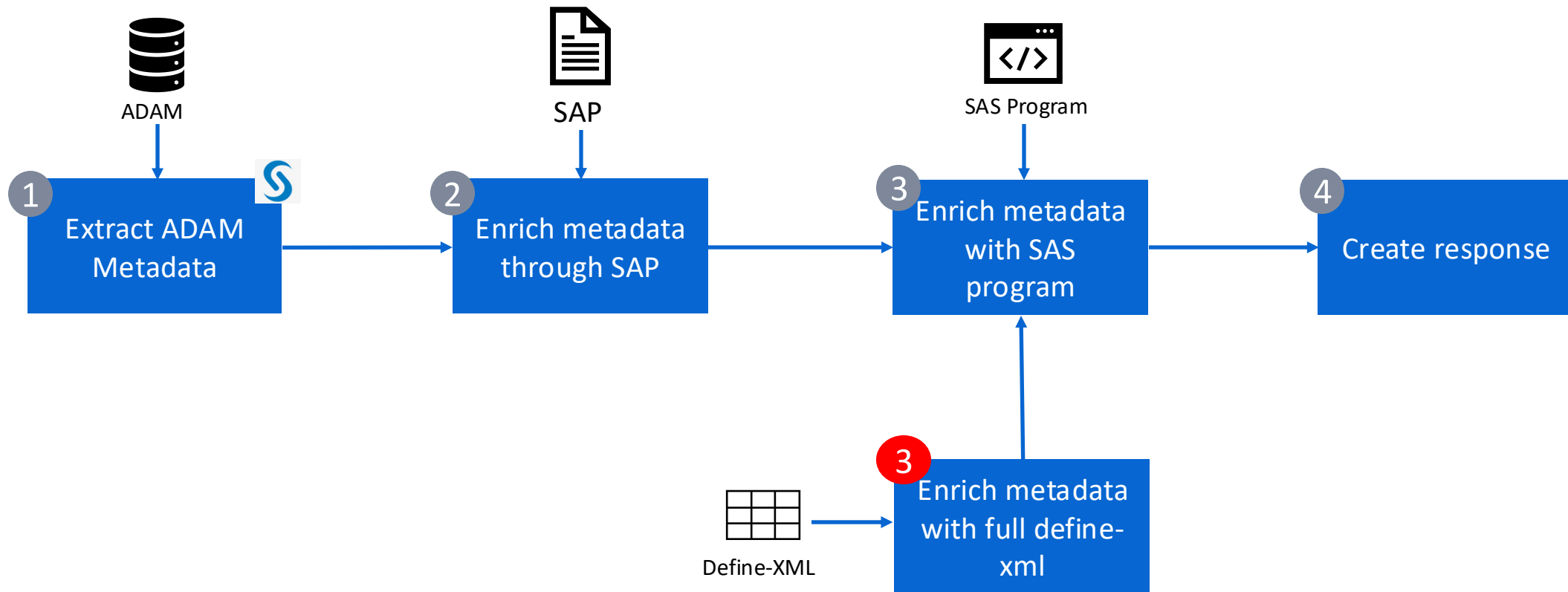
SAS Code Validation

Process flow



Define-XML

Process flow



Connect to SAS Viya

Username

sbxwev

Password

.....



Press Enter to submit form

Login

SAS Agents for Clinical Trials



Validation Findings for ADaM Dataset "EUETS.ADVS"

1. Frequency Distributions for Categorical Variables:

- All categorical variables were checked for frequency distributions. No unexpected categories or missing values were found in key variables like `STUDYID`, `USUBJID`, `SUBJGR1`, `FASFL`, `THTBIP`, `VSTESTCD`, `PARAMCD`, `AVISIT`, `ATPT`, `APHASE`, `DIYPE`, `ATOXDSCI`, `ATOXGR1`, `ATOXDSCI`, `ATOXGRH`, `ABLFL`, `AML01FL`, `CRIT1FL`, and `CRIT2FL`.

2. Summary Statistics for Continuous Variables:

- Continuous variables were summarized, showing expected ranges and distributions. Key variables like `VSSTRESN`, `AVAL`, `BASE`, and `CHG` were within expected ranges, except for some threshold violations noted below.

3. Outlier Detection in Continuous Variables:

- The `PROC UNIVARIATE` procedure identified extreme values in `VSSTRESN`, `AVAL`, and `CHG`. The minimum value for `VSSTRESN` and `AVAL` was 5, which is below the expected range for some parameters.

4. Validation of Derivations for Change from Baseline:

- The derivation of `CHG` (Change from Baseline) was validated. No mismatches were found between `CHG` and the calculated difference `AVAL - BASE`.

5. Threshold Violations for Abnormalities:

- Numerous threshold violations were detected for parameters like `DIABP` and `RESP`, indicating values outside the expected ranges specified in the SAP. For example, `DIABP` had values below 91 and above 100, and `RESP` had values below 17 and above 25.

[Disconnect from SAS Viya](#)

Navigation

[SAS Code Validation](#)[Define-XML](#)

7. Handling of Missing Values:

- Missing values were checked for key variables. `BASE` and `CHG` had missing values, which aligns with the SAP's specification of no imputation for missing values.

Summary:

- **Mismatches and Issues:**
 - **Threshold Violations:** Significant threshold violations were found for `DIABP` and `RESP`, which need to be addressed to align with SAP specifications.
 - **Outliers:** Extreme values in `VSSTRESN` and `AVAL` need further investigation to ensure data integrity.
- **Suggestions for Corrections:**
 - Review and correct the data for `DIABP` and `RESP` to ensure they fall within the specified thresholds.
 - Investigate the source of extreme values in `VSSTRESN` and `AVAL` to determine if they are valid or require correction.

Overall, the dataset requires adjustments to meet the SAP specifications, particularly concerning threshold violations and outlier management.



Combined Findings from Dataset and Code Validations

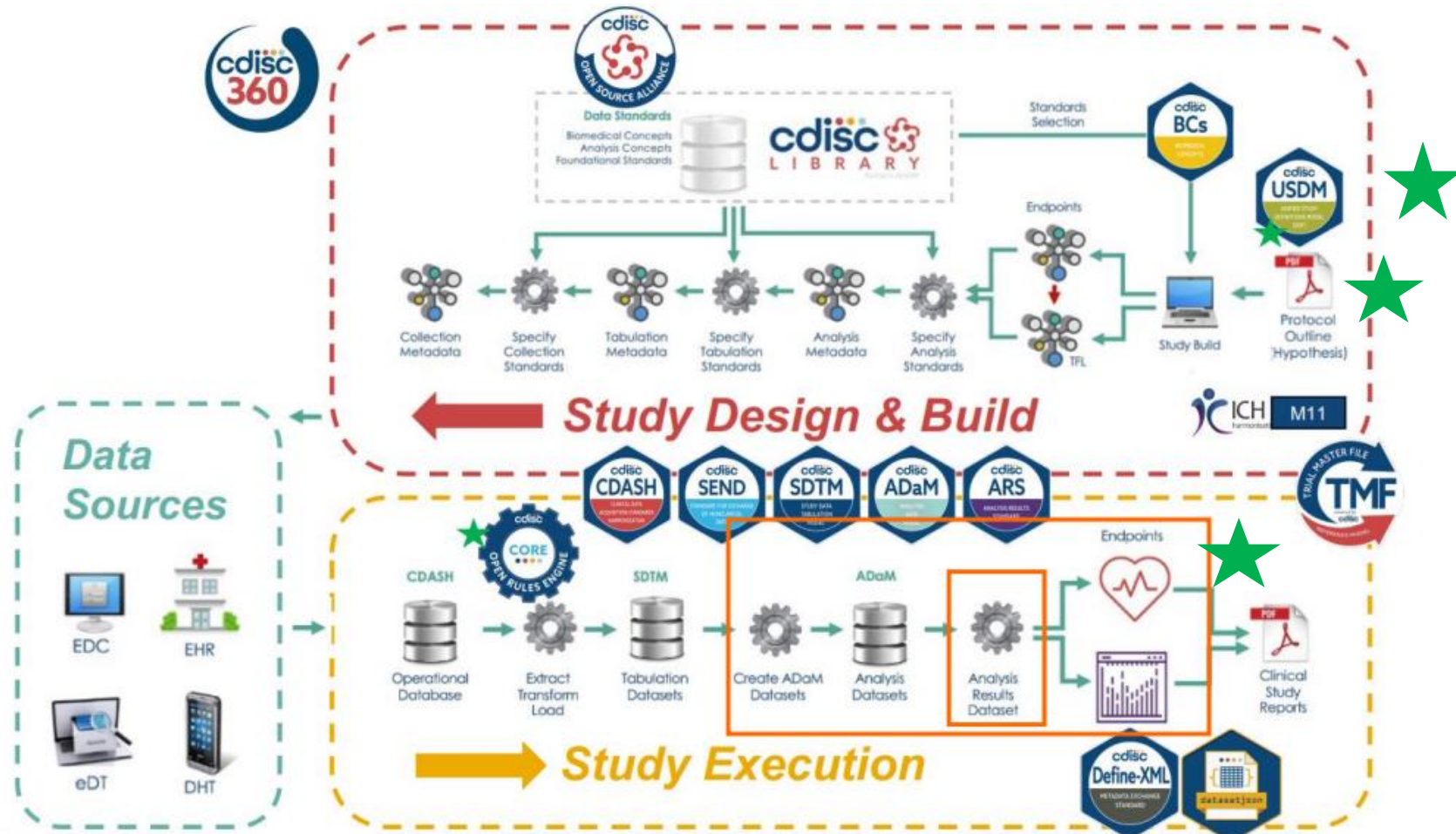
Areas of Compliance:

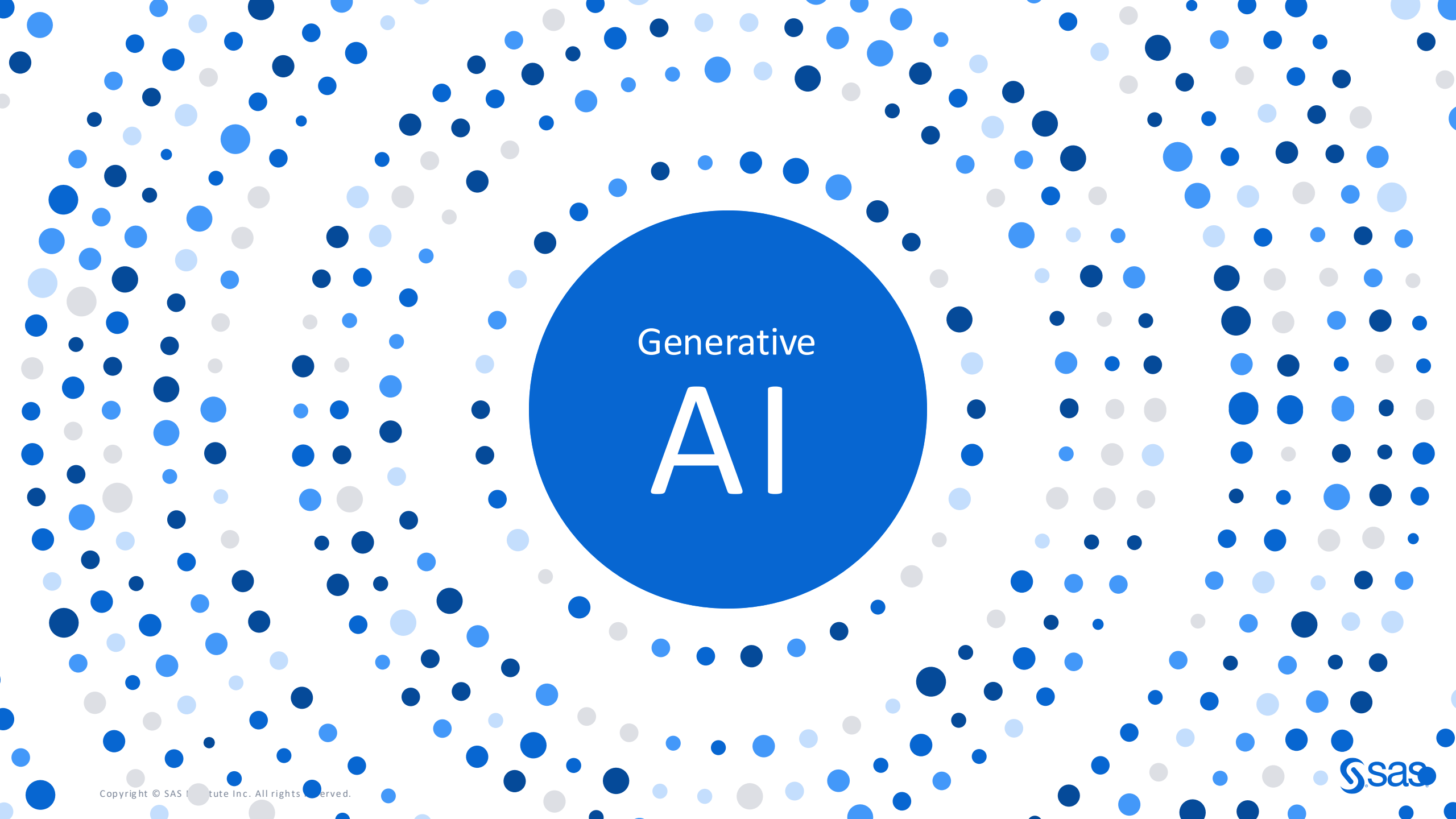
1. **Tachycardia and Bradycardia:**
 - Both the dataset and the SAS program correctly implement the grading for tachycardia and bradycardia as per the SAP specifications.

Ask a follow-up question about the validation



CDISC 360 Vision



The background of the slide is a light blue-grey color, densely populated with numerous small circles in various shades of blue and grey. These circles are scattered across the entire frame, creating a textured, data-like background. In the center of the slide is a large, solid blue circle. Inside this circle, the words "Generative" and "AI" are written in white. "Generative" is in a smaller, sans-serif font, and "AI" is in a much larger, bold, sans-serif font.

Generative AI

Generative AI in action with SAS



SAS Viya

Explain, Govern,
Orchestrate LLMs



Viya Copilots

Accelerate AI
Lifecycle / Industry/
Business Tasks



Synthetic Data Generation

Mitigate Data
Quality and Scarcity



Generative AI in Action with SAS



SAS Viya

Explain, Govern,
Orchestrate LLMs



Viya Copilots

Accelerate AI
Lifecycle / Industry/
Business Tasks



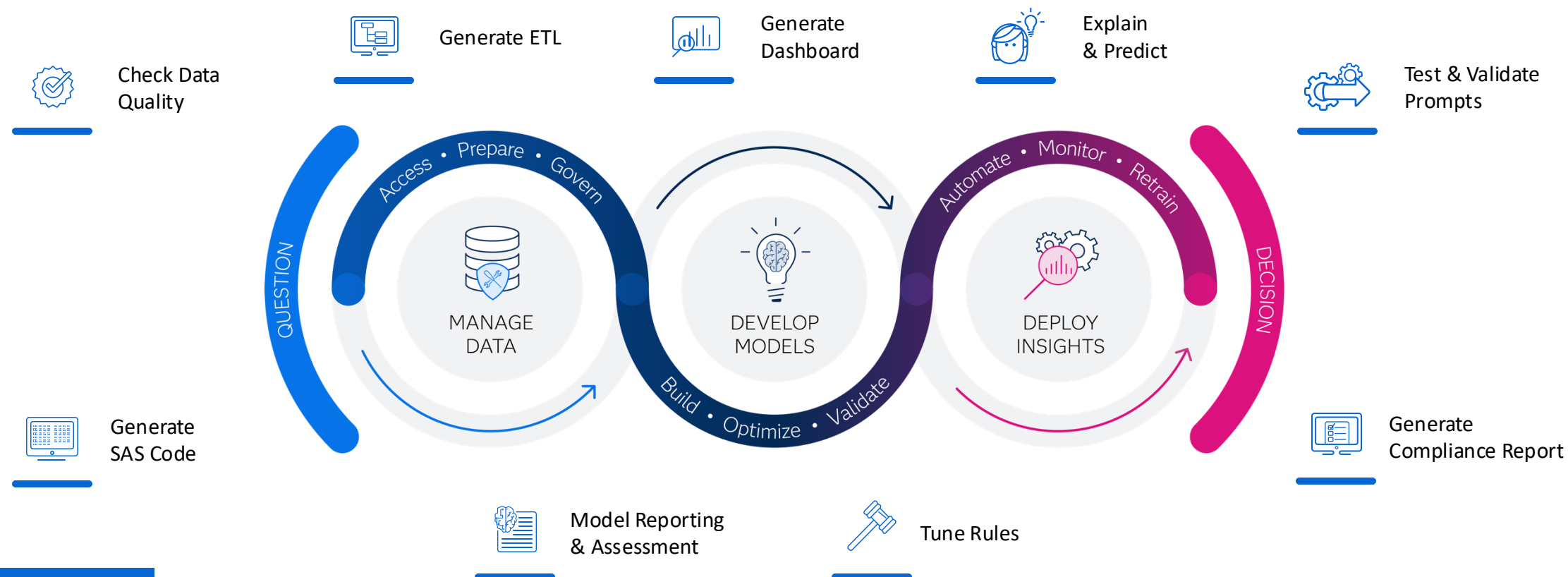
Synthetic Data Generation

Mitigate Data
Quality and Scarcity



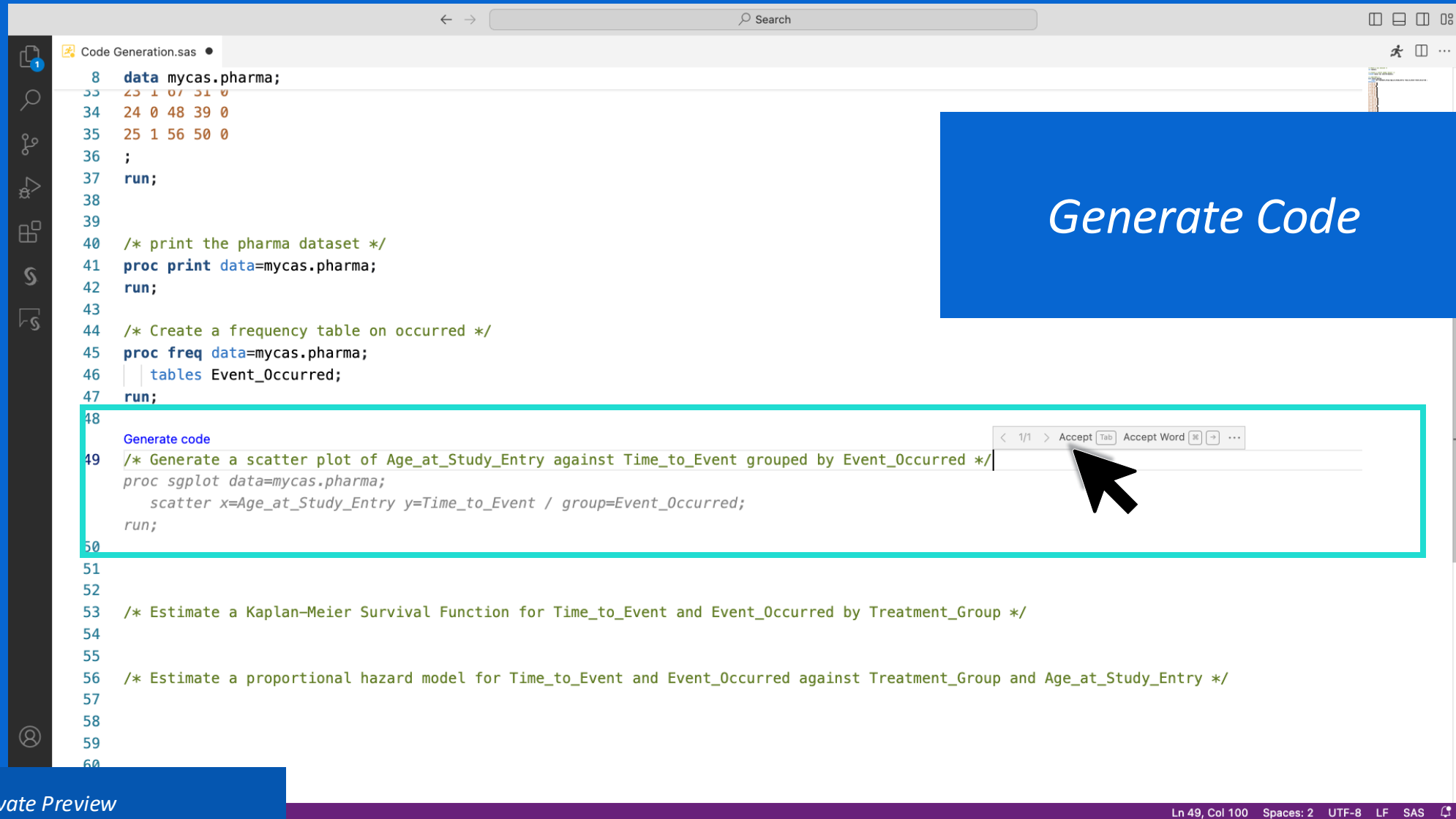
SAS® Viya® Copilot to help you use SAS Viya

Viya Copilot to further democratize analytics as a suite of GenAI productivity tools to help you conversationally query the end-to-end SAS Viya AI & Analytics Life Cycle platform



*Proof of concept
Subject to change*

SAS Code Generation



The screenshot displays the SAS Viya Copilot interface. The main window shows a SAS code editor with the following code:

```
8 data mycas.pharma;  
33 23 1 07 31 0  
34 24 0 48 39 0  
35 25 1 56 50 0  
36 ;  
37 run;  
38  
39  
40 /* print the pharma dataset */  
41 proc print data=mycas.pharma;  
42 run;  
43  
44 /* Create a frequency table on occurred */  
45 proc freq data=mycas.pharma;  
46 | tables Event_Occurred;  
47 run;  
48  
49 /* Generate a scatter plot of Age_at_Study_Entry against Time_to_Event grouped by Event_Occurred */  
50 proc sgplot data=mycas.pharma;  
51 scatter x=Age_at_Study_Entry y=Time_to_Event / group=Event_Occurred;  
52 run;  
53  
54 /* Estimate a Kaplan-Meier Survival Function for Time_to_Event and Event_Occurred by Treatment_Group */  
55  
56 /* Estimate a proportional hazard model for Time_to_Event and Event_Occurred against Treatment_Group and Age_at_Study_Entry */  
57  
58  
59  
60
```

A blue box highlights the 'Generate code' button, and a black arrow points to the 'Accept' button in the code completion menu.

Private Preview

Ln 49, Col 100 Spaces: 2 UTF-8 LF SAS

Generate Code

SAS VIYA COPILOT: CHAT

You

Explain the code.

SAS Viya Copilot

This SAS code is using the PROC LIFETEST procedure to perform survival analysis on the dataset named "mycas.pharma".

The "Time_to_Event" variable is the time variable that represents the duration until the event of interest occurs, and the "Event_Occurred" variable is a binary variable indicating whether the event has occurred (1 for event occurrence, 0 for no event occurrence). This is specified using the "Time_to_Event*Event_Occurred(0)" syntax.

The "strata" statement is used to stratify the analysis by the variable "Treatment_Group", which means that the analysis will be performed separately for each unique value of the "Treatment_Group" variable.

Overall, this code is performing survival analysis using the Cox proportional hazards model to assess the relationship between the time to event and the treatment group, while taking into account any stratification specified by the "Treatment_Group" variable.

Code Generation.sas

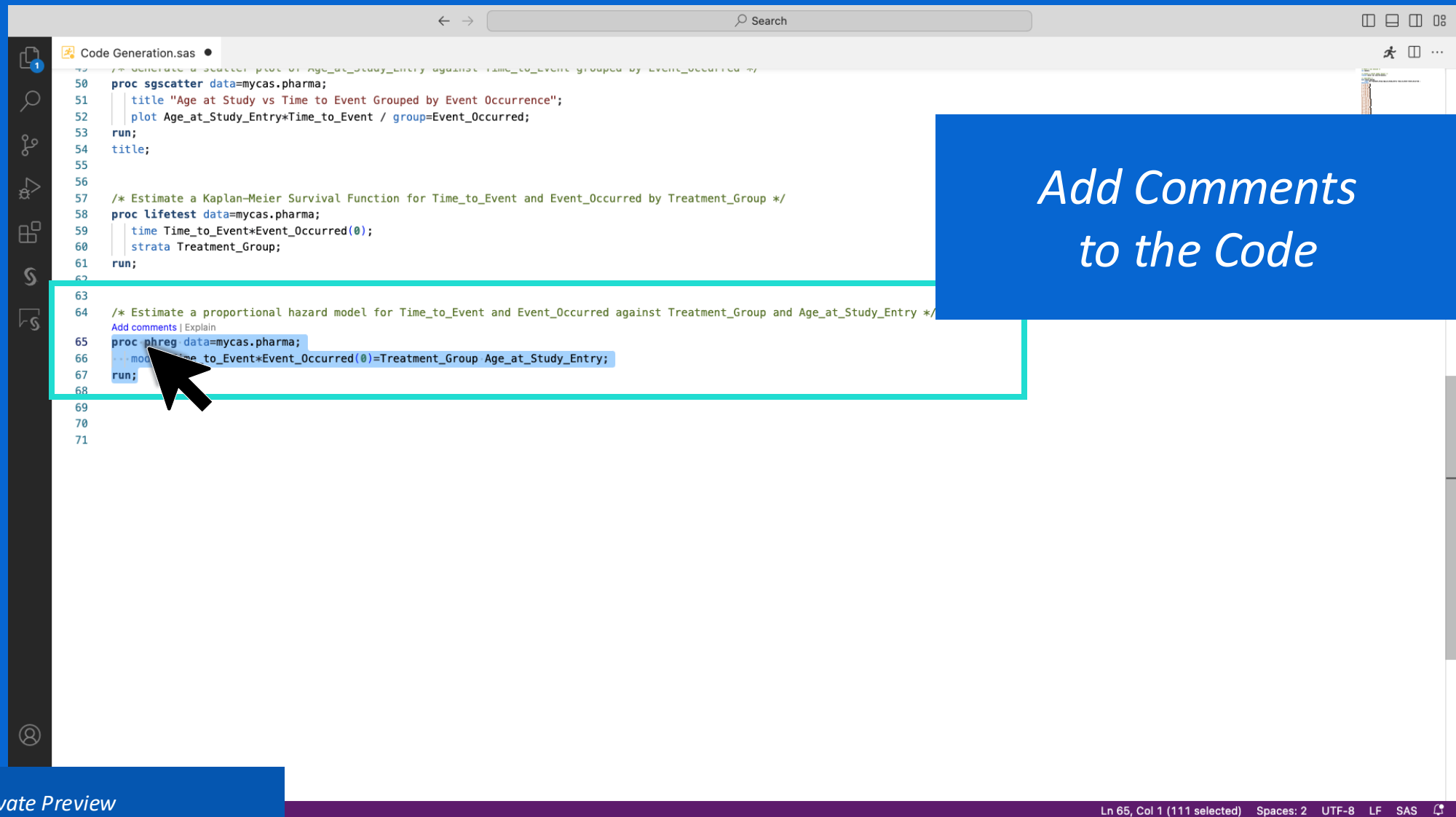
```
43
44 /* Create a frequency table on occurred */
45 proc freq data=mycas.pharma;
46 | tables Event_Occurred;
47 run;
48
49 /* Generate a scatter plot of Age_at_Study_Entry aga
50 proc sgscatter data=mycas.pharma;
51 | title "Scatter Plot: Age vs. Time to Event (Group
52 | plot Age_at_Study_Entry*Time_to_Event / group=Eve
53 run;
54 title;
55
56
57 /* Estimate a Kaplan-Meier Survival Function for Time_to_Event and Event_Occurred by
   Treatment_Group */
   Add comments | Explain
58 proc lifetest data=mycas.pharma;
59 | time Time_to_Event*Event_Occurred(0);
60 | strata Treatment_Group;
61 run;
62
63
64 /* Estimate a proportional hazard model for Time_to_Event and Event_Occurred against
   Treatment_Group and Age_at_Study_Entry */
65 proc phreg data=mycas.pharma;
66 | model Time_to_Event*Event_Occurred(0)=Treatment_Group Age_at_Study_Entry;
67 run;
68
69
70
71
```

Explain the Code

Private Preview

Ln 58, Col 1 (105 selected) Spaces: 2 UTF-8 LF SAS

SAS Code Generation



The screenshot displays the SAS Viya Copilot interface. The main editor shows SAS code for generating a scatter plot and survival functions. A section of the code, starting with a comment about a proportional hazard model, is highlighted in light blue. A black arrow points to the first line of this highlighted section. To the right of the code, a blue callout box contains the text "Add Comments to the Code". The status bar at the bottom indicates the current line and column (Ln 65, Col 1) and that 111 characters are selected.

```
Code Generation.sas •  
50 proc sgscatter data=mycas.pharma;  
51 | title "Age at Study vs Time to Event Grouped by Event Occurrence";  
52 | plot Age_at_Study_Entry*Time_to_Event / group=Event_Occurred;  
53 run;  
54 title;  
55  
56  
57 /* Estimate a Kaplan-Meier Survival Function for Time_to_Event and Event_Occurred by Treatment_Group */  
58 proc lifetest data=mycas.pharma;  
59 | time Time_to_Event*Event_Occurred(0);  
60 | strata Treatment_Group;  
61 run;  
62  
63  
64 /* Estimate a proportional hazard model for Time_to_Event and Event_Occurred against Treatment_Group and Age_at_Study_Entry */  
65 proc phreg data=mycas.pharma;  
66 | model Time_to_Event*Event_Occurred(0)=Treatment_Group Age_at_Study_Entry;  
67 run;  
68  
69  
70  
71
```

Add Comments to the Code

Ln 65, Col 1 (111 selected) Spaces: 2 UTF-8 LF SAS

Private Preview

Prep & Analyze Data Feature

Your Dataset

Data

BAD	LOAN	MORTDAYS	VALUE	REASON	JOB	YOJ	DEROG	DELINQ	CAGE	
1	1100	25860	39025	Homelmp	Other	10.5	0	0	94.366666667	
1	1300	70053	68400	Homelmp	Other	7	0	2	121.83333333	
1	1500	13500	16700	Homelmp	Other	4	0	0	149.46666667	
1	1500	.	.	.	.	.	.	.	.	
0	1700	97800	112000	Homelmp	Office	3	0	0	93.333333333	
1	1700	30548	40320	Homelmp	Other	9	0	0	101.46600191	
1	1800	48649	57037	Homelmp	Other	5	3	2	77.1	
1	1800	28502	43034	Homelmp	Other	11	0	0	88.766029879	
1	2000	32700	46740	Homelmp	Other	3	0	2	216.93333333	
1	2000	.	62250	Homelmp	Sales	16	0	0	115.8	
1	2000	22608	.	.	.	18	.	.	.	
1	2000	20627	29800	Homelmp	Office	11	0	1	122.53333333	
1	2000	45000	55000	Homelmp	Other	3	0	0	86.066666667	
0	2000	64536	87400	.	Mgr	2.5	0	0	147.13333333	
1	2100	71000	83850	Homelmp	Other	8	0	1	123	
1	2200	24280	34687	Homelmp	Other	.	0	1	300.86666667	
1	2200	90957	102600	Homelmp	Mgr	7	2	6		
1	2200	23030	.	.	.	19	.	.	.	
1	2300	28192	40150	Homelmp	Other	4.5	0	0		
0	2300	102370	120953	Homelmp	Office	2	0	0		
1	2300	37626	46200	Homelmp	Other	3	0	1		
1	2400	50000	73395	Homelmp	ProfExe	5	1	0		
1	2400	28000	40800	Homelmp	Mgr	12	0	0		
1	2400	18000	.	Homelmp	Mgr	22	.	2	121.73333333	

Proof of concept

with mock data

Prep & Analyze Data Feature

Your Dataset

(1)

Data

Filter

Check Data Quality

BAD	LOAN	MORTD	VALUE	REASON	JOB	YOJ	DEROG	DELINQ	CLAGE
1	1100	25860	39025	Homelmp	Other	10.5	0	0	94.366666667
1	1300	70053	68400	Homelmp	Other	7	0	2	121.833333333
1	1500	13500	16700	Homelmp	Other	4	0	0	149.466666667
1	1500	.	.	.	.	.	.	.	.
0	1700	97800	112000	Homelmp	Office	3	0	0	93.333333333
1	1700	30548	40320	Homelmp	Other	9	0	0	101.46600191
1	1800	48649	57037	Homelmp	Other	5	3	2	77.1
1	1800	28502	43034	Homelmp	Other	11	0	0	88.766029879
1	2000	32700	46740	Homelmp	Other	3	0	2	216.933333333
1	2000	.	62250	Homelmp	Sales	16	0	0	115.8
1	2000	22608	.	.	.	18	.	.	.
1	2000	20627	29800	Homelmp	Office	11	0	1	122.533333333
1	2000	45000	55000	Homelmp	Other	3	0	0	86.066666667
0	2000	64536	87400	.	Mgr	2.5	0	0	147.133333333
1	2100	71000	83850	Homelmp	Other	8	0	1	123
1	2200	24280	34687	Homelmp	Other	.	0	1	300.866666667
1	2200	90957	102600	Homelmp	Mgr	7	2	6	121.733333333
.	.	.	.	.	.	19	.	.	.
1	2400	28000	40800	Homelmp	Mgr	12	0	0	121.733333333
1	2400	18000	.	Homelmp	Mgr	22	.	2	121.733333333

Check Data Quality

Enter the data quality task you want to perform, or type '+' for recommended data quality tasks....

Proof of c

with moc

Proof of concept

with mock data



Proof of concept

with mock data

Proof of concept

with mock data

Prep & Analyze Data Feature

Your Dataset

Data

Filter

Check Data Quality

BAD	LOAN	MORTDUE	VALUE	REASON	JOB	YOJ	DEROG	DELINQ	CLAGE
1	1100	25860	39025	Homelmp	Other	10.5	0	0	94.36666667
		70053	68400	Homelmp	Other	7	0	2	121.83333333
		13500	16700	Homelmp	Other	4	0	0	149.46666667
		.	.	.	.	.	.	.	.
		97800	112000	Homelmp	Office	3	0	0	93.33333333
		30548	40320	Homelmp	Other	9	0	0	101.46600191
		48649	57037	Homelmp	Other	5	3	2	77.1
		28502	43034	Homelmp	Other	11	0	0	88.766029879
		32700	46740	Homelmp	Other	3	0	2	216.93333333
		.	62250	Homelmp	Sales	16	0	0	115.8
		22608	.	.	.	18	.	.	.
		20627	29800	Homelmp	Office	11	0	1	122.53333333
		45000	55000	Homelmp	Other	3	0	0	86.06666667
		64536	87400	.	Mgr	2.5	0	0	147.13333333
		71000	83850	Homelmp	Other	8	0	1	123
		24280	34687	Homelmp	Other	.	0	1	300.86666667
		90957	102600	Homelmp	Mgr	7	2	6	.
		23030	.	.	.	19	.	.	.
		28192	40150	Homelmp	Other	4.5	0	0	0
		102370	120953	Homelmp	Office	2	0	0	0
		37626	46200	Homelmp	Other	3	0	1	0
		50000	73395	Homelmp	ProfExe	5	1	0	0
		28000	40800	Homelmp	Mgr	12	0	0	0
		18000	.	Homelmp	Mgr	22	.	2	121.73333333

Check Data Quality

You

Fix data issues

SAS Viya Copilot

The following data issues are detected:

Duplicate data

Inconsistent formats

Missing values

Just fix all of the above

View issues

AI generated results. Human review recommended.

Enter the data quality task you want to perform, or type '+' for recommended data quality tasks....

Proof of concept
with mock data

Proof of concept

with mock data

Proof of concept

with mock data



Proof of concept

with mock data

Generative AI in Action with SAS



SAS Viya

Explain, Govern,
Orchestrate LLMs

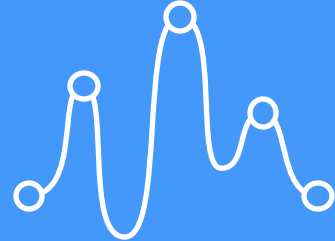
● ● ●



Viya Copilots

Accelerate AI
Lifecycle / Industry /
Business Tasks

● ● ●



**Synthetic Data
Generation**

Mitigate Data
Quality and Scarcity

hazy

 [Software](#) [Learn](#) [Support](#) [Partners](#) [About Us](#)

SAS acquires Hazy synthetic data software to
boost generative AI portfolio

Strategic move reinforces SAS leadership in data and AI, enhancing customer innovation
and data security

Copilot Use Cases – 2025 Roadmap

HLS

#	Use Case	User Story
1	Patient cohort creation	As a: Clinical Researcher I want: to use natural language processing (NLP) to generate SQL queries for creating patient cohorts So that: I can efficiently and accurately create patient cohorts without needing extensive SQL knowledge
2	Clinical record summary	As a: Healthcare Provider I want: to use natural language processing (NLP) to generate concise summaries of clinical records from structured clinical notes data So that: I can quickly understand patient history and current condition without reading through extensive clinical notes
3	SAS Code Creation in SCE	As a: Biostatistician I want: to use a SAS Viya Copilot to generate SAS code for biostatistical analysis in clinical trials So that: I can efficiently perform complex statistical analyses without manually writing extensive SAS code
4	Synthetic SDTM data creation	As a: Data Manager I want: to use Data Maker to generate synthetic Study Data Tabulation Model (SDTM) data So that: I can create realistic datasets for testing and validation without compromising patient privacy

Real-world use cases for Copilot

Unlocking new possibilities with the latest SAS Health features

BaseEpic IntegrationDashboard : Prior ResponsesNetworkQ&A

Select a Patient:

Patient Names



Ask a question about this patients chart:

Question

What is the patient's medication history?

Submit

The patient was on an unknown medication for hypertension in 1998, which was stopped after six months due to drowsiness. The patient was also on cimetidine for peptic ulcer disease in 1990, which resolved after three months.

Patient History Summary:

Benjamin Clark, a middle-aged man, presents with a complex medical history. Notable conditions include **hypertension**, and **hyperlipidemia**. Vascular abnormalities consist of a left frontal **arteriovenous malformation (AVM)**, an 8mm anterior communicating **artery aneurysm**, and a **9mm ectasia** of the left anterior cavernous carotid artery. Benjamin is currently on statin therapy for **hyperlipidemia**. His lab results

Medication History

Medication Name	Pharmacy Name	Date	Date of Prescription Pick Up
Metformin	CVS Pharmacy	2/2/2024	
Metformin	CVS Pharmacy	5/15/2024	
Metformin	CVS Pharmacy	1/15/2024	1/17/2024

Choose a language: EnglishFrenchDutchSpanishPortugueseGreekJapanese

Dear [Recipient Name],

I hope this email finds you well.

I am writing to inquire about the possibility of receiving a 10% discount on the [Model Name] [Make] that I am interested in purchasing from your dealership.

I understand that the current price of the vehicle is [Price]. I would be grateful if you could consider offering me a discounted price of [Discounted Price].

What is Trustworthy AI?



Developing and using AI technologies in an **ethical** manner



Ensuring AI does not harm people

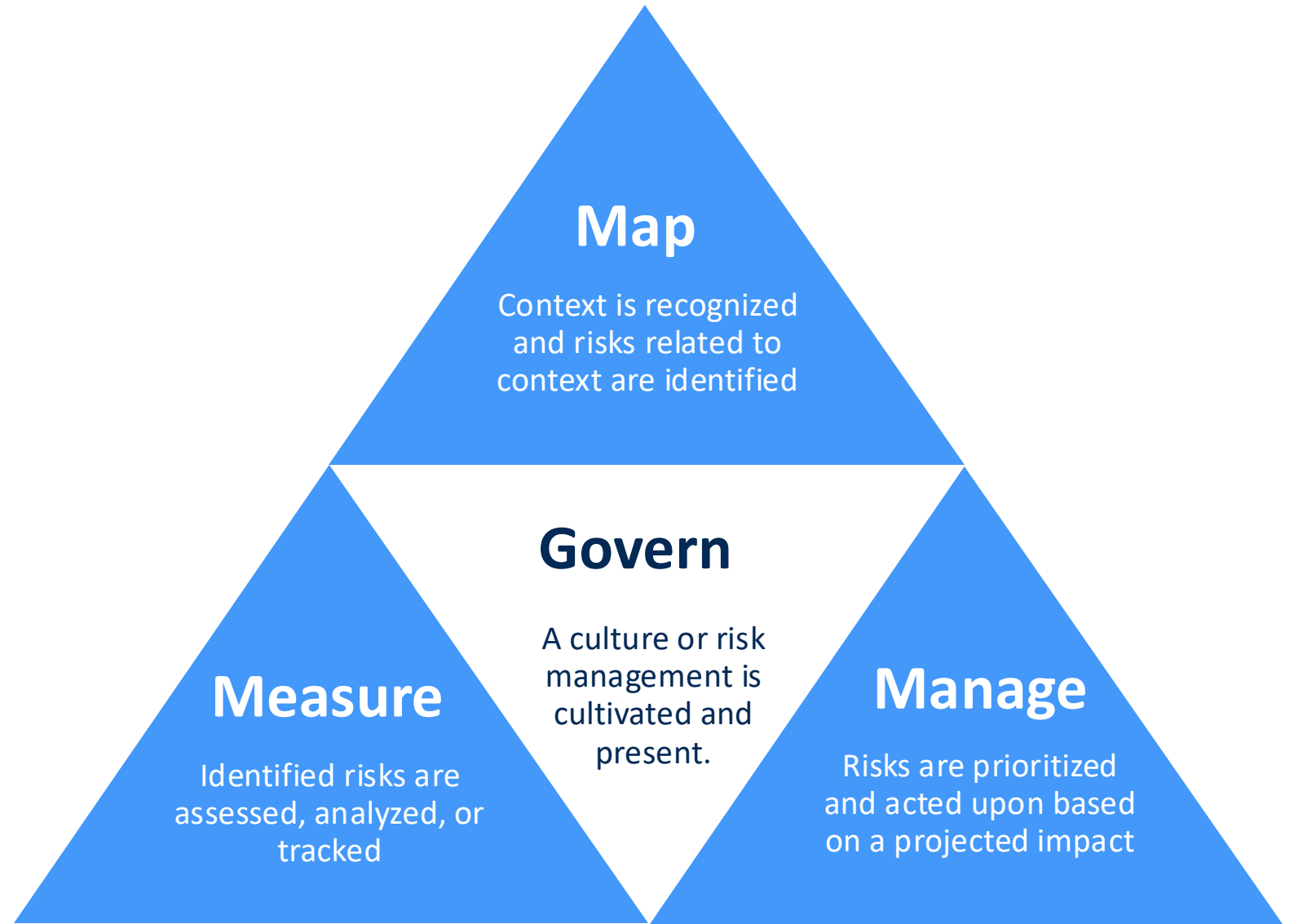


Asking not just, “Could we?”, but also “Should we?”



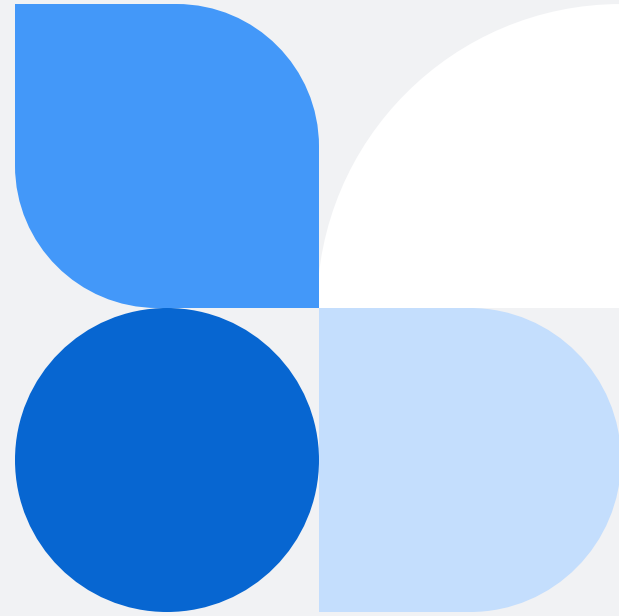
Building AI that reflects **our values** as a society

AI Risk Management Framework



SAS Viya Trustworthy AI capabilities





Benefits for Generative AI

■ Accelerated Innovation

Quickly transform LLMs into actionable insights by seamlessly integrating GenAI models into decisioning workflows, AI/ML applications and existing business processes

■ Data Protection

Uphold user privacy and security with robust data protection measures, including data minimization, anonymization and encryption, ensuring that sensitive information remains safeguarded

■ Trustworthy Results

Apply SAS natural processing techniques to preprocess data, ensuring only high-quality data are fed to LLMs, minimizing computational waste, reducing costs and producing reliable outcomes

■ Enhanced Governance

Use built-in workflows that validate the entire life cycle of LLMs, from regulatory compliance to model risk management. Additionally, SAS offers responsible ideation, experimentation and operationalization support

■ Precise Decisioning

Quantitative decisioning capabilities, critical for successful GenAI reasoning, are built into the Viya platform.

Thank you

