



2024 CDISC + TMF
EUROPE INTERCHANGE

BERLIN

24-25 APRIL: CONFERENCE & EXPO | 22, 23, 26 APRIL: TRAININGS

Revolutionizing Regulatory Insights: Harnessing Large Language Models (LLMs) within SAS

Soundarya Palanisamy, Sr Industry Advisor, Global Health and Life Sciences Practice, SAS
Dario Fabiani, Data Scientist, Dedagroup Business Solutions

Meet the Speakers

Soundarya Palanisamy

Title: Sr Industry Advisor, Global Health and Life Sciences Practice

Organization: SAS Institute



Dario Fabiani

Title: Data Scientist

Organization: Dedagroup Business Solutions





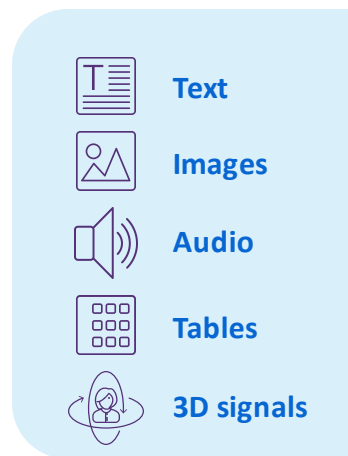
Generative AI

An AI method capable of generating NEW content

Creating New Realities

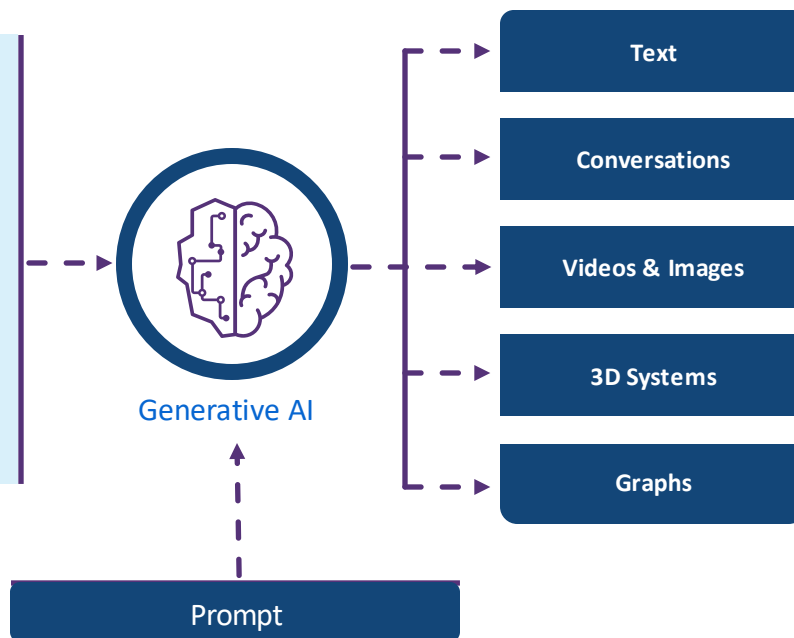
Generative AI

LEARNS FROM DATA



Generative AI

GENERATES SOMETHING *NEW*



An open analytics platform





What is Generative AI?

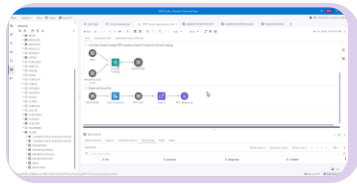
Generative AI (GAI) is a field that refers to AI systems designed to learn from real-world data to generate new data – it could be text, images, video, audio, or even tabular data.

Generative AI's potential to create lifelike content and product innovative solutions has garnered significant attention in research, industry, and the broader AI community. This technology has the potential to change our very understanding of how work is done.

SAS and Generative AI

SAS Viya to **help build your LLM application stack** to unlock value in your proprietary business assets

Flexible GenAI building blocks to help with development and orchestration of your LLM application stack



example. Build a Prompt Library using existing SAS Viya APIs and SAS Studio steps

Copilot to accelerate user productivity

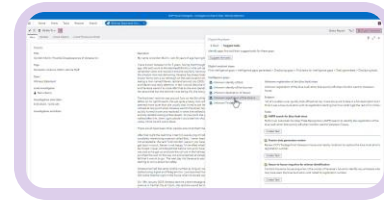
Con conversationally query SAS Viya to increase user productivity



example: SAS Code Generation

Intelligent Assistants to accelerate industry workflows and processes

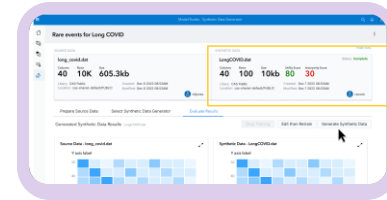
GenAI driven Assistants to accelerate domain-specific workflows



example. Law Enforcement Investigations

Standalone GenAI services to enable your toolchains, like a **Synthetic Data Generation service**

Low-code/no-code synthetic data generation and validation experience



example. Synthetic data generation API and low-code/no-code experience



Regulatory Assistant

Streamlining regulatory compliance with intelligent document interpretation.

The CDISC Case

Importance and Challenges



Complexity of the Standard

- The CDISC encompasses a set of guidelines and standards for the collection, exchange, reporting, and archiving of clinical research data.
- The breadth and complexity of these standards can be overwhelming, especially for organizations with fewer resources or less experience



Challenges in Revisions

- The continuous evolution and updates of the standard make it difficult for companies to maintain compliance without constant external support.
- Reviewing data to ensure compliance can be a lengthy and error-prone process.

Benefits

Speed of standard adoption

- LLMs can assist in interpreting and applying CDISC standards, guiding organizations through the compliance process more quickly and efficiently.
- The ability of LLMs to process large amounts of text can significantly speed up the review of documentation and the identification of non-compliance areas.

Reductions of Errors

- Through automation and the precision of LLMs, it is possible to minimize human errors in the mapping and application of standards.
- Generative AI can suggest corrections and improvements based on a vast knowledge base, enhancing data quality.

Democratization of Access

- By making LLM-based tools accessible to a broad range of organizations, barriers to entry for CDISC compliance are reduced, allowing even entities with fewer resources to meet the required standards.

Decision Support

- By providing data-based analysis and recommendations, LLMs can assist researchers and professionals in making informed decisions regarding clinical data management, improving the effectiveness of studies

Accelerating CDISC Compliance

Empowering Compliance with Regulatory Assistant

Advanced Technology Integration

- Direct access to comprehensive CDISC documentation for SDTM, ADaM.
- Utilizes a Retrieval-Augmented Generation (RAG) system for enhanced information retrieval.
- Leverages Large Language Models (LLMs) to filter and synthesize pertinent information.

Strategic Advantages of Document Assistant

- Simplifies the process of adhering to CDISC standards, making the path to compliance more straightforward.
- Enhances the users' understanding of CDISC standards, supporting informed decision-making.
- Integrates vast amounts of data to provide comprehensive insights and recommendations, enhancing decision-making processes.

Enhanced Data Quality and Compliance

- Ensures that organizations can meet CDISC requirements with ease.
- Minimizes manual efforts in navigating and applying compliance guidelines.
- Empowers users to focus on research outcomes rather than compliance processes.

1° Use case: Documentation Q&A

1° Question

2° Use case: Compliance Recommendation App

**Recommendation
Use Case
DM dataset**

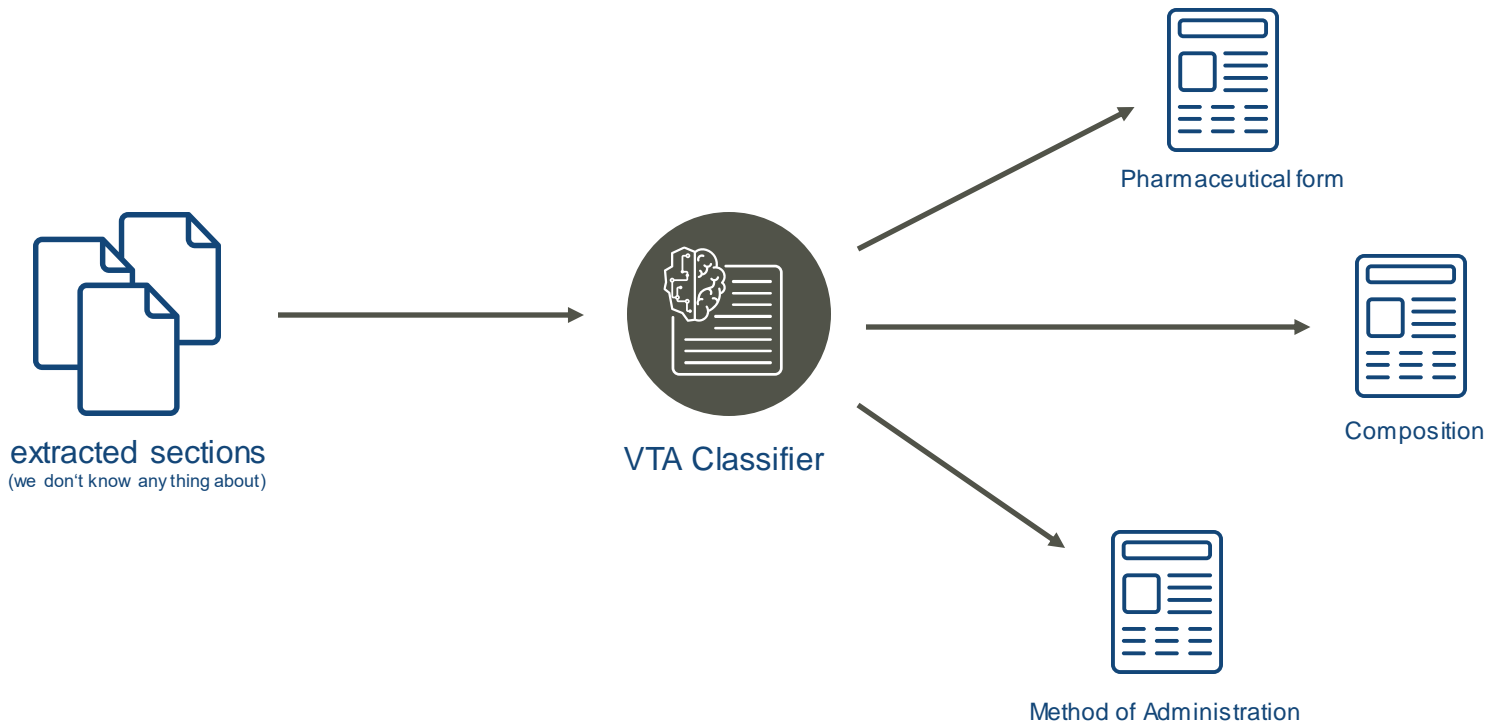


“Pre- Filter” – What goes into the LLM



Deciding what is relevant

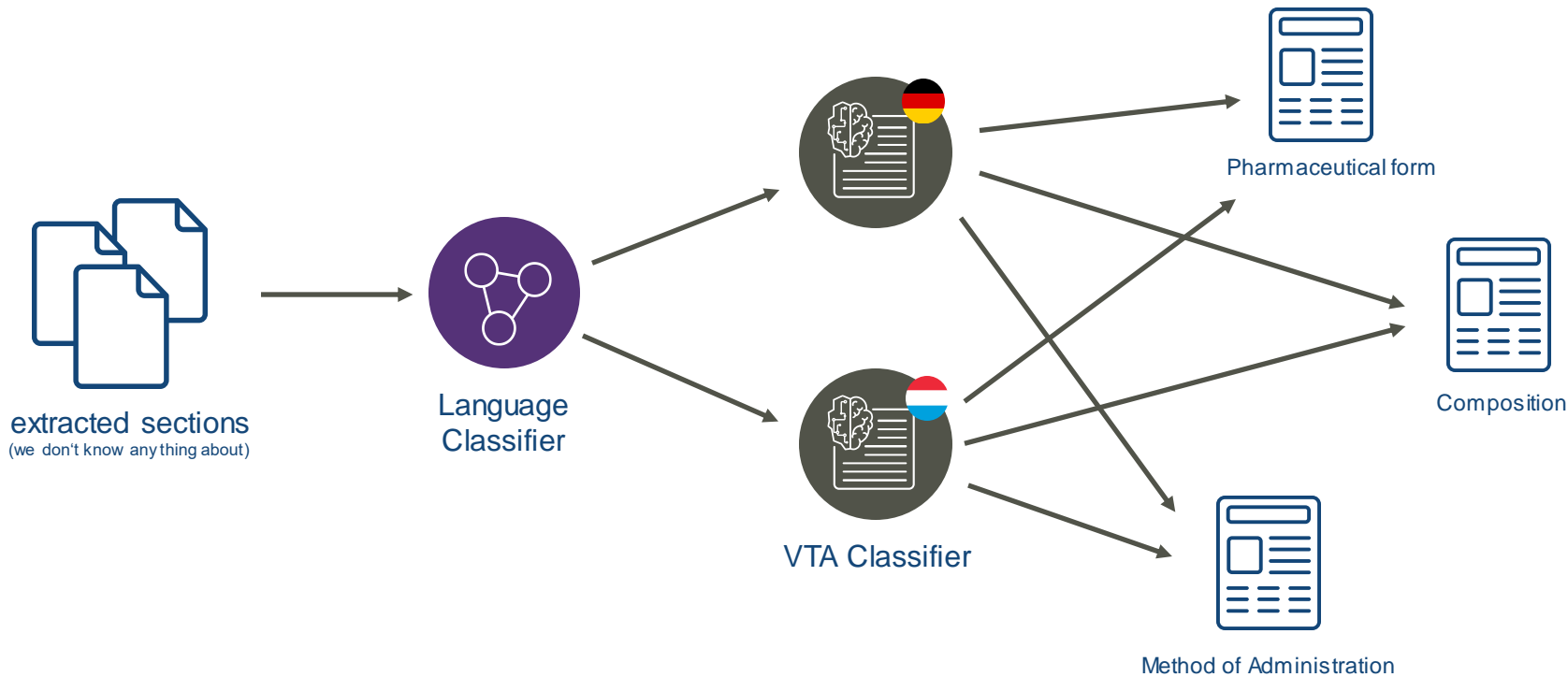
Classifying sections





Deciding what is relevant

Classifying sections



extracted sections
(we don't know anything about)

Language
Classifier

VTA Classifier

Pharmaceutical form

Composition

Method of Administration



“Quality Checks” – output from the LLM

Trust but Verify

Applying sanity checking on extracted information



Obs	side_effects
1	Increase in liver transaminases
2	Changes of blood count such as thrombocytopenia and agranulocytosis
3	Bronchospasm (asthma induced by analgesics)
4	Hypersensitivity reactions like erythema including urticaria and anaphylactic shock
5	Serious skin reactions

1

Get extractions from LLM

Obs	rule
1	1:rule_1:(OR[NW],(AND[0.5],"increase@"),(AND[0.5],"liver transaminase@"))
2	2:rule_2:(OR[NW],(AND[0.5],"change@"),(AND[0.5],"blood count@"),(AND[0.5],"thrombocytopenia@"),(AND[0.5],"agranulocytosis@"))
3	3:rule_3:(OR[NW],(AND[0.5],"bronchospasm@"),(AND[0.5],"asthma@"),(AND[0.5],"induce@"),(AND[0.5],"analgesic@"))
4	4:rule_4:(OR[NW],(AND[0.5],"hypersensitivity reaction@"),(AND[0.5],"erythema@"),(AND[0.5],"include@"),(AND[0.5],"urticaria@"),(AND[0.5],"anaphylactic shock@"))
5	5:rule_5:(OR[NW],(AND[0.5],"serious@"),(AND[0.5],"skin reaction@"))

2

Automatically generate rules for quality check

Obs	match	_category_	confidence
1	Increase in liver transaminases	rule_1	1
2	Changes of blood count such as thrombocytopenia and agranulocytosis	rule_2	1
3	bronchospasm (asthma induced by analgesics)	rule_3	1
4	Hypersensitivity reactions like erythema including urticaria and anaphylactic shock	rule_4	1
5	serious skin reactions	rule_5	1

3

Score LITl rules, retrieve matches and calculate confidence score

Confidence Score

An Example



Source

4.8 Undesirable effects

...

Blood and lymphatic system disorders

Very rare: Changes of blood count such as thrombocytopenia and agranulocytosis

...

Possible extractions

a) Changes of blood count such as thrombocytopenia and agranulocytosis

- (OR[NW],(AND[0.5],"change@"),(AND[0.5],"blood count@"), (AND[0.5],"thrombocytopenia@"), (AND[0.5],"agranulocytosis@"))

Score

1

b) Changes of blood count such as thrombocytopenia, leucopenia and agranulocytosis

- (OR[NW],(AND[0.5],"change@"),(AND[0.5],"blood count@"), (AND[0.5],"thrombocytopenia@"), (AND[0.5],"leucopenia@"),(AND[0.5],"agranulocytosis@"))

0.80

c) Changes of blood count such as thrombocytopenia, leucopenia, pancytopenia and agranulocytosis

- (OR[NW],(AND[0.5],"change@"),(AND[0.5],"blood count@"), (AND[0.5],"thrombocytopenia@"), (AND[0.5],"leucopenia@"),(AND[0.5],"pancytopenia@"), (AND[0.5],"agranulocytosis@"))

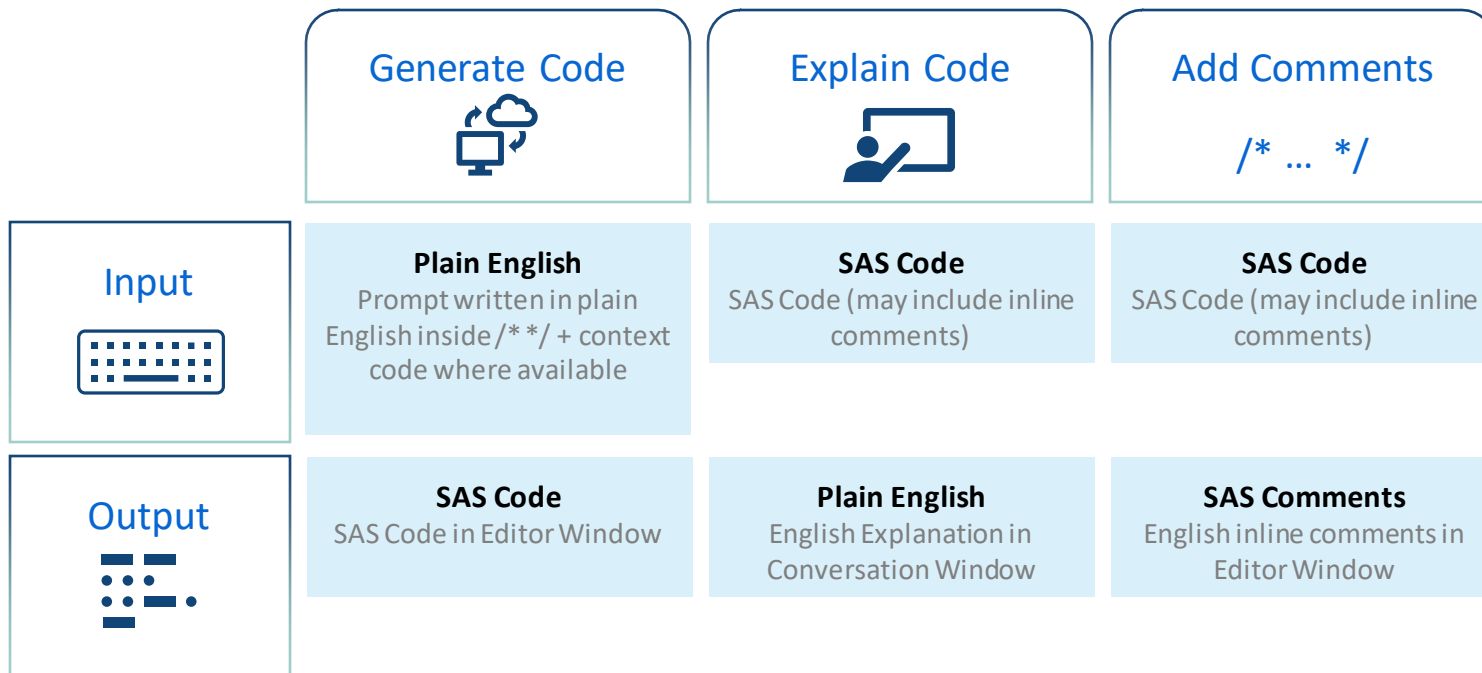
0.66



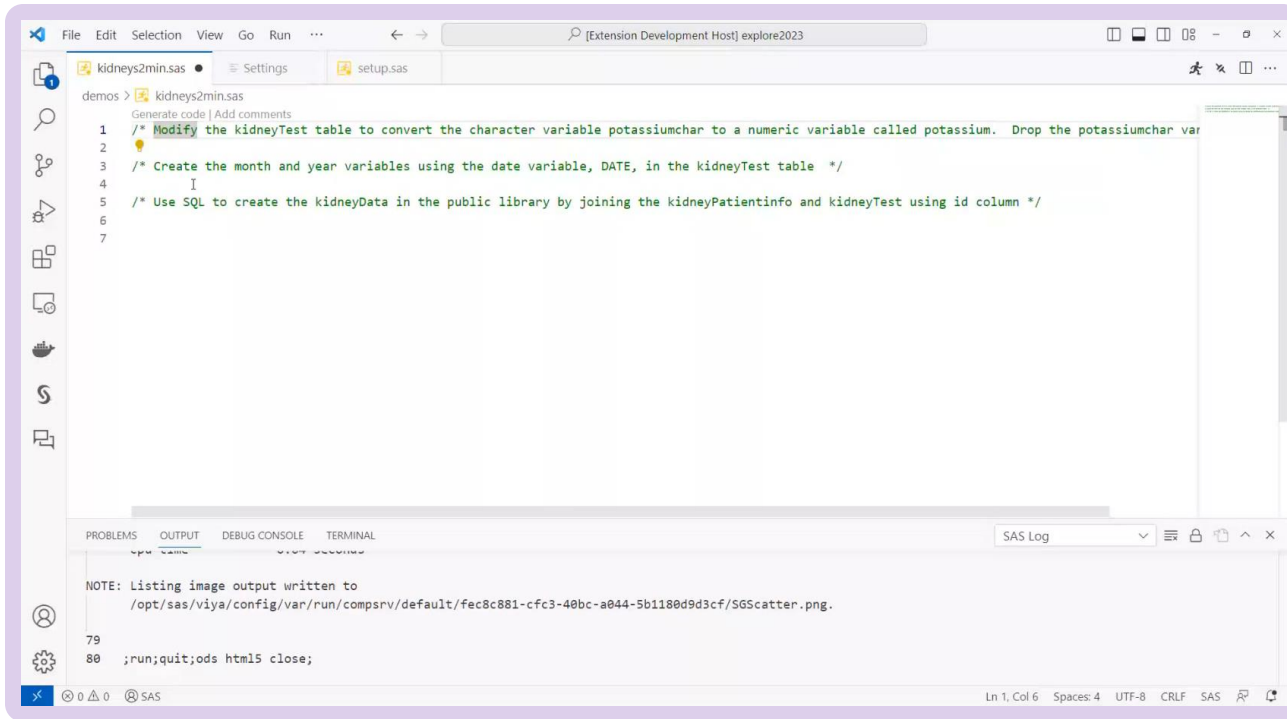
Code Generation

Your next assistant

Copilot – Features



Copilot



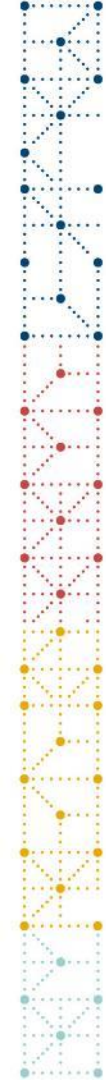
The screenshot displays the Visual Studio Code editor with the Copilot extension. The main editor window shows SAS code generated by Copilot, with a yellow dot indicating the start of the generation. The code includes comments and SQL statements for modifying a table and creating new variables.

```
1 /* Modify the kidneyTest table to convert the character variable potassiumchar to a numeric variable called potassium. Drop the potassiumchar var  
2  
3 /* Create the month and year variables using the date variable, DATE, in the kidneyTest table */  
4  
5 /* Use SQL to create the kidneyData in the public library by joining the kidneyPatientinfo and kidneyTest using id column */  
6  
7
```

The bottom panel shows the SAS Log output, which includes a note about the location of the listing image output and the execution of the SAS code.

```
NOTE: Listing image output written to  
/opt/sas/viya/config/var/run/compsrv/default/fec8c881-cfc3-40bc-a044-5b1180d9d3cf/SGScatter.png.  
79  
80 ;run;quit;ods html5 close;
```

The status bar at the bottom indicates the current cursor position: Ln 1, Col 6, Spaces: 4, UTF-8, CRLF, SAS.



■ 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

Thank you!



Soundarya.Palanisamy@sas.com



dario.fabiani@dedagroup.it