

From Programmer to AI-Enabled Clinical Data Scientist:

Accelerating Clinical Development with Trusted Copilots

Daniel Christen – Advisory Solutions Engineer –
HLS Solutions at SAS

Reimagining the Clinical Data Scientist Role:

Future-Proofing Skills and Partnerships

PhUSE SDE Focus

Reimagining the clinical data scientist role means embracing a future where technical excellence, strategic thinking and cross-functional collaboration converge to create greater scientific and operational impact. As technological acceleration reshapes the clinical development landscape, our role evolves from traditional programming to orchestrating intelligent, end-to-end data ecosystems. Future-proofing our skills requires combining expertise in analytics, data standards and quality with new capabilities offered by state-of-the-art technologies.

1. How our role may evolve and future skills in demand

2. Tool transformation – practical examples of AI-assisted programming

3. Closer integration with statisticians and clinical, regulatory and technology teams in more agile and digital environments

1. How our role may evolve and future skills in demand

How much intelligence is in AI ?

1. AI is "just retrieval"

- AI does not understand information the way humans do.
- It has no consciousness, self-awareness, desires, or intentions.
- Everything it produces is ultimately derived from patterns found in training data.
- When connected to RAG* systems, it is effectively a very sophisticated search-and-summarization engine.

2. AI shows genuine intelligence

- Intelligence does not necessarily require consciousness.
- Humans are also heavily pattern learners.
- Modern AI can perform tasks it was never explicitly programmed to do.
- It can reason across multiple sources, write software, solve novel problems, and explain concepts in ways that look remarkably similar to human reasoning.

3. The middle ground

- Current AI exhibits some components of intelligence, but not human-like understanding.
- A modern LLM is **like an extremely well-read assistant that has absorbed patterns from vast amounts of text and can synthesize them into new responses**
- The validity of any AI-generated conclusion **must be reviewed and confirmed by a qualified human** subject matter expert with full knowledge of the relevant context and data.

The conversation is shifting from "Can AI find an answer?" to "Can we trust, validate, and trace how the answer was produced?"

*RAG (Retrieval-Augmented Generation) combines a large language model (LLM) with external knowledge sources, retrieving relevant information at query time to generate more accurate, contextual, and traceable responses.

Reimagining the Clinical Data Scientist Role:

Future-Proofing Skills and Partnerships

PhUSE SDE Focus

Reimagining the clinical data scientist role means embracing a future where technical excellence, strategic thinking and cross-functional collaboration converge to create greater scientific and operational impact. As technological acceleration reshapes the clinical development landscape, our role evolves from traditional programming to orchestrating intelligent, end-to-end data ecosystems. Future-proofing our skills requires combining expertise in analytics, data standards and quality with new capabilities offered by state-of-the-art technologies.

1. How our role may evolve and future skills in demand

SAS believes that the role of the data scientist will **continue to evolve from that of a primarily technical specialist to a domain-focused subject matter expert who leverages AI** and analytics to solve complex business challenges within their field of expertise.

2. Tool transformation – practical examples of AI-assisted programming

3. Closer integration with statisticians and clinical, regulatory and technology teams in more agile and digital environments

2. Tool transformation – practical examples of AI-assisted programming

SAS Co-Pilot for Code Generation

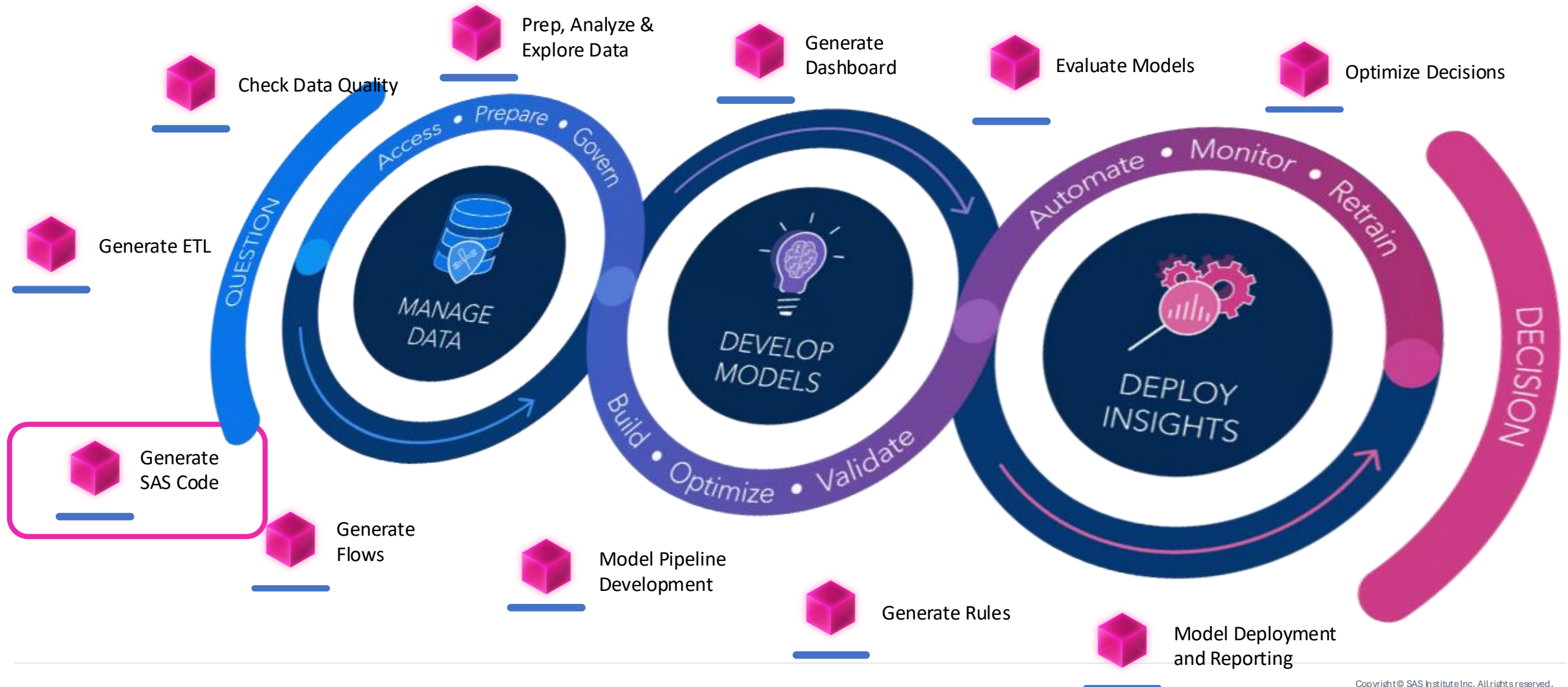
Generate

Document

Explain

2. Tool transformation – practical examples of AI-assisted programming

Your AI assistant for the Data and AI Lifecycle



Generate SAS Code

From RAW data to final TFLs

1. RAW to SDTM

- Implementation Guide
(Previous CDISC Mappings)

2. SDTM to ADaM

- Implementation Guide
- Study Design Documents

3. ADaM to TFL

- Implementation Guide
- Study Design Documents
- Mock-up TFL

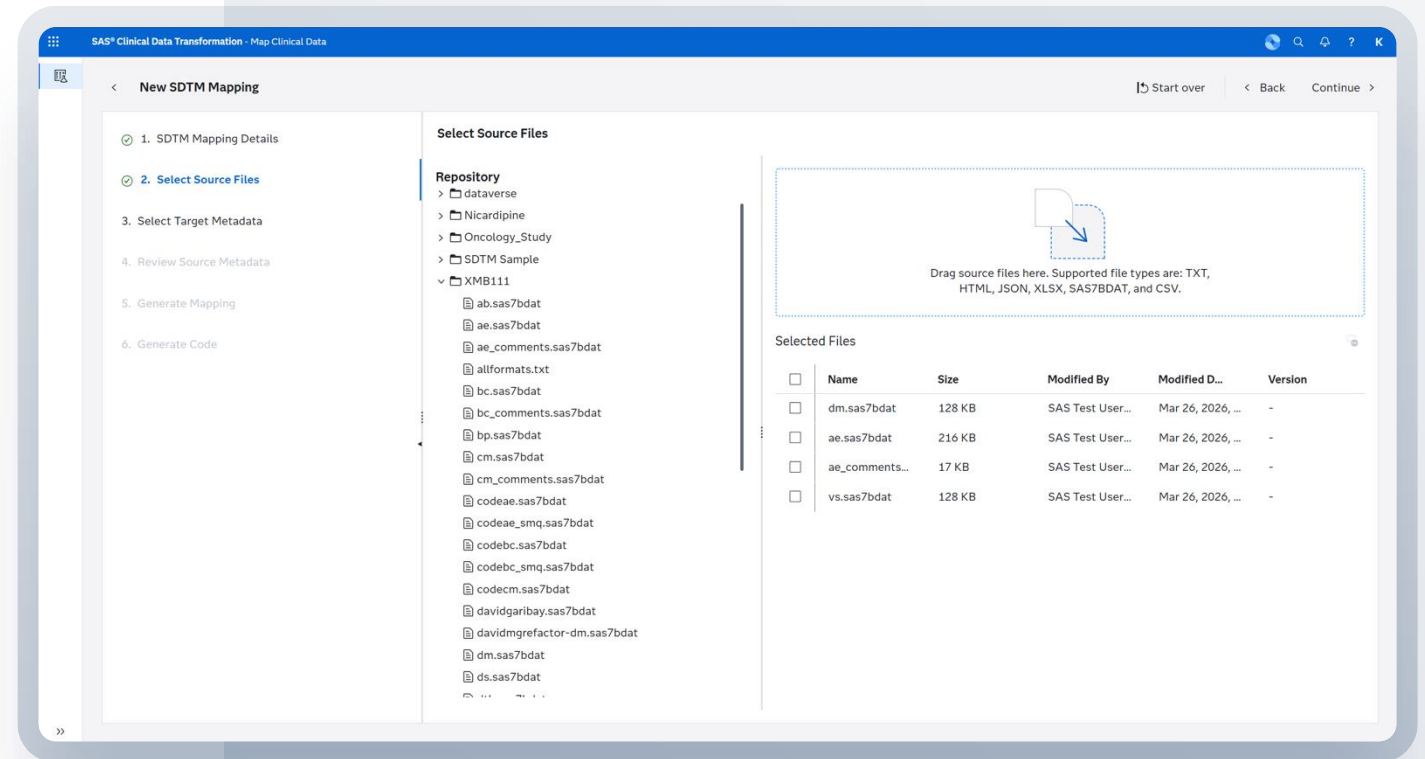
RAG (Retrieval-Augmented Generation) combines a large language model (LLM) with external knowledge sources, retrieving relevant information at query time to generate more accurate, contextual, and traceable responses.

Clinical Data Transformation

Raw to SDTM

01

Select source data
from Clinical
Acceleration
Repository select the
dataset(s) you'd like to
map



Clinical Data Transformation

Raw to SDTM

02

Select target data structure by selecting implementation guide from CDISC importing previous mapping specifications or specific target structure

The screenshot displays the 'SAS Clinical Data Transformation - Map Clinical Data' web application. The main heading is 'New SDTM Mapping'. On the left, a vertical list of steps is shown: 1. SDTM Mapping Details, 2. Select Source Files, 3. Select Target Metadata (highlighted with a blue bar), 4. Review Source Metadata, 5. Generate Mapping, and 6. Generate Code. The right-hand panel is titled 'Select Target Metadata' and contains the following fields: 'SDTMIG:' with a dropdown menu showing '3.4', 'Upload mapping specifications:' with a text input 'mappings_documen' and a 'Browse' button, and 'Domains:' with a horizontal list of tags 'AE x', 'CM x', and 'DM x', followed by a plus sign button. At the top right of the interface, there are navigation links: 'Start over', '< Back', and 'Continue >'. The browser's address bar shows the URL 'sas.com/clinical-data-transformation/map-clinical-data'.

Raw to SDTM

Clinical Data Transformation

03

Run Agent

to map from source to target with access to edit mapping and transformation directly via the UI

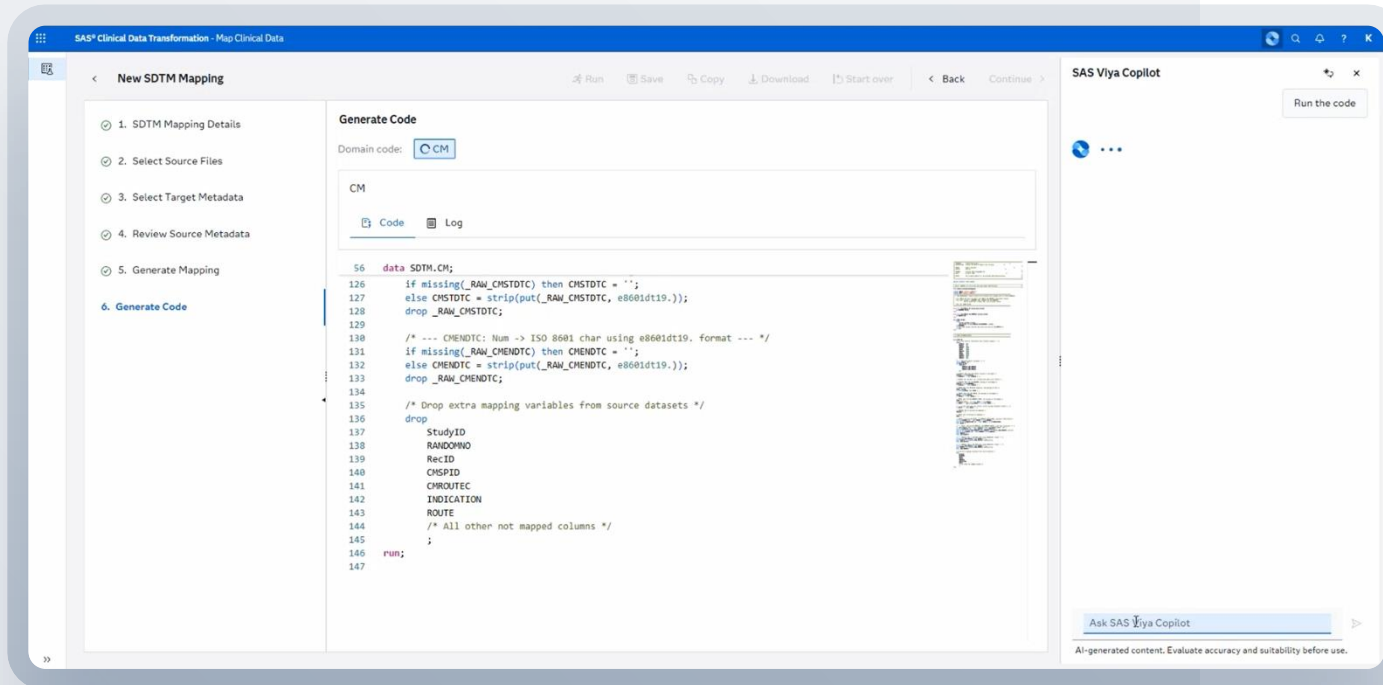
The screenshot displays the 'SAS Clinical Data Transformation - Map Clinical Data' application. The 'New SDTM Mapping' window is open, showing a progress bar on the left with steps: 1. SDTM Mapping Details, 2. Select Source Files, 3. Select Target Metadata, 4. Review Source Metadata, 5. Generate Mapping (selected), and 6. Generate Code. The main area is titled 'Generate Mapping' and shows a 'Domain code' of 'CM'. Below this is a table with columns: Target Column, Source Data, and Transformation.

Target Column	Source Data	Transformation
STUDYID	CM	Direct mapping (rename) from CM.StudyID to CM.STUDYID
DOMAIN	Constant	None
USUBJID	CM	Direct mapping (rename) from CM.RANDOMNO to CM.USUBJID
CMSEQ	CM	Direct mapping (rename) from CM.RecID to CM.CMSEQ
CMSPID	CM	Direct mapping (rename) from CM.CMSPID to CM.CMSPID
CMTRT	CM_COMMENTS	Direct mapping (rename) from CM_COMMENTS.CMTRT to CM.CMTRT
CMDECOD	Unmapped	None
CMCAT	Unmapped	None
CMINDC	CM	Coalesce CM.CMINDC and CM_COMMENTS.INDICATION (union or preference order if both present)
CMROUTE	CM	Coalesce CM.CMROUTE and CM_COMMENTS.ROUTE (union or preference order if both present)
CMSTDTC	CM	Direct mapping (rename) from CM.CMSTDTC to CM.CMSTDTC; ensure ISO 8601 formatting if needed
CMENDTC	CM	Direct mapping (rename) from CM.CMENDTC to CM.CMENDTC; ensure ISO 8601 formatting if needed

Raw to SDTM

Clinical Data Transformation

04



Produce your code
and run it directly within Clinical
Data Transformation, engage
with Code Assist or Co-Pilot or
save to your workspace to begin
work

SDTM to ADaM

Clinical Data Transformation

05

The screenshot displays the SAS Clinical Data Transformation - Map Clinical Data interface. The left sidebar shows a progress list with steps: 1. ADaM Mapping Details, 2. Select Files, 3. Review Source Metadata, 4. Select Target Metadata (highlighted), 5. Suggested Output, 6. Generate Mapping, and 7. Generate Code. The main panel is titled 'Study One' and contains the 'Select Target Metadata' section. It prompts the user to 'Required target metadata (select at least one): *'. The 'ADaMIG' option is selected with a checkbox, and a dropdown menu shows '3.4'. Below this, there are three optional target metadata sections: 'Mapping specifications', 'SAP', and 'Protocol', each with an 'Upload' button and a 'Selected file: none' status. At the bottom, there is a 'Mock shells' section with an 'Upload' button and a 'Selected file: none' status. A small disclaimer at the bottom right reads: 'AI-generated content. Evaluate accuracy and suitability before use.'

Select target data structure by selecting implementation Guide from CDISC and importing study design documents.

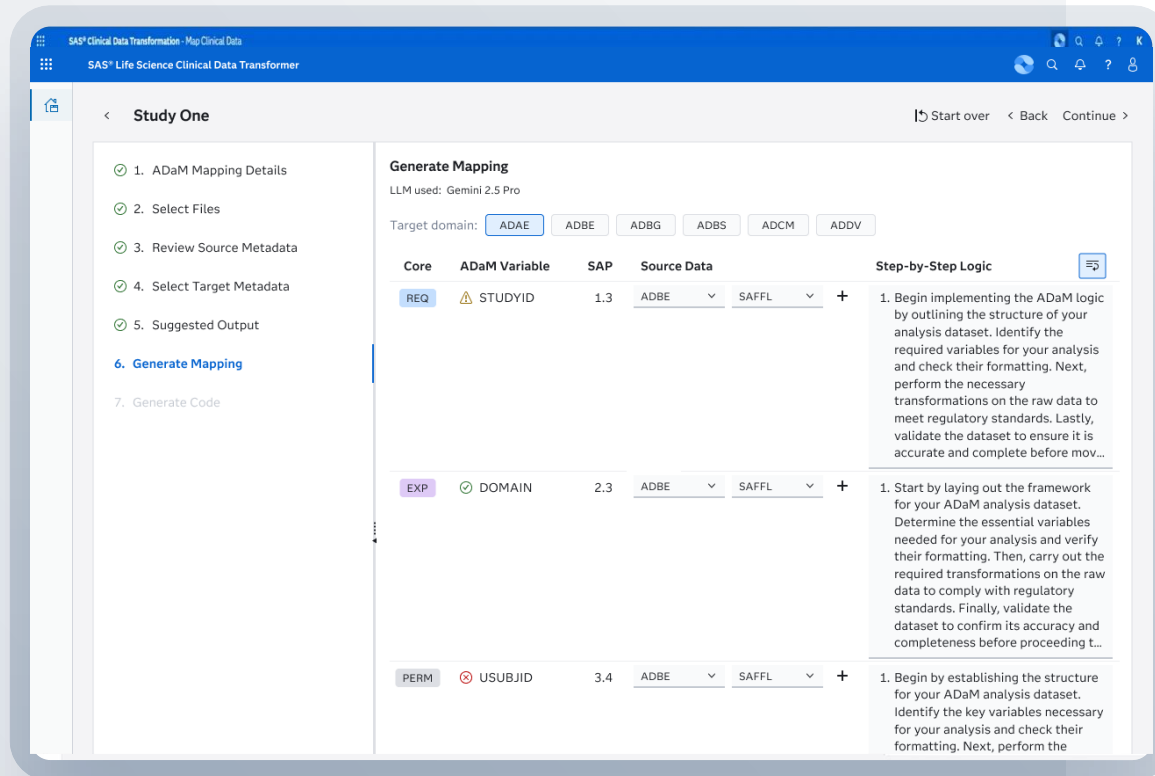
SDTM to ADaM

Clinical Data Transformation

06

Run Agent

to map from source to target with access to edit mapping and transformation directly via the UI





Software Learn Support Partners About Us SAS Innovate



Welcome, sinpan

Clinical Data Flow AI Assistant



Project Name: DependencyChanges_Small

Estimated Cost: \$13.52

Clinical Data Transformation Steps

Convert Raw Data to SDTM standards

Start here as the first step. You can choose to upload a Study Design document. Once done press Start next to 'Identify.' to get SDTM datasets, variables, and mapping logic. Once satisfied, click Start next to 'Generate.' to produce SDTM Datasets & SAS code.

Study Design: [Choose File](#) No file chosen Uploaded: [STUDY_Design.docx](#)

Instructions: Optional instructions for the study protocol...

Identify SDTM Datasets: [Start](#)

Macros File: [Choose File](#) No file chosen Uploaded: [All_macros.txt](#)

Macros Instructions: Optional instructions for the SDTM macros file...

Generate SDTM Datasets: [Start](#)

State: [Approved](#)

Extract TLF shells logic

Upload the TLF shells document, and press Start to extract TLF shell details. When content is satisfactory, select "Approved" in status to proceed to the next step.

TLF Shells: [Choose File](#) No file chosen Uploaded: [CDISCPLOT_TLF_Shells.pdf](#)

Instructions: Optional instructions for TLF shells...

Agent Outputs

SDTM Datasets	SDTM Code	ADaM Datasets	ADaM Code	TLF Logic	TLF Code
TA	TV	TI	TS	TE	DM
DS	AE	EX	CM	LB	SC
VS	QS	MH	SE	SUPPDS	RELREC
SUPPAE	SUPPLB				

Code Last Run Log Last Run Output

[Run Code](#) [Save Code](#) Last Run Status: [Success](#)

```
/* PROGRAM: create_sdtm_ta.sas */
/* DESCRIPTION: Creates the SDTM Trial Arms (TA) domain from raw data. */
/* INPUT: raw.PATIENT_INFORMATION, raw.VISIT_SCHEDULE, raw.PROTOCOL_DEFINITIONS */
/* OUTPUT: sdtm.TA */
/* AUTHOR: Clinical Data Programmer AI */
/* DATE: 29 March 2026 */
/* NOTES: This program adheres to the provided SDTM specifications. */

options nocenter ls=132 ps=64;

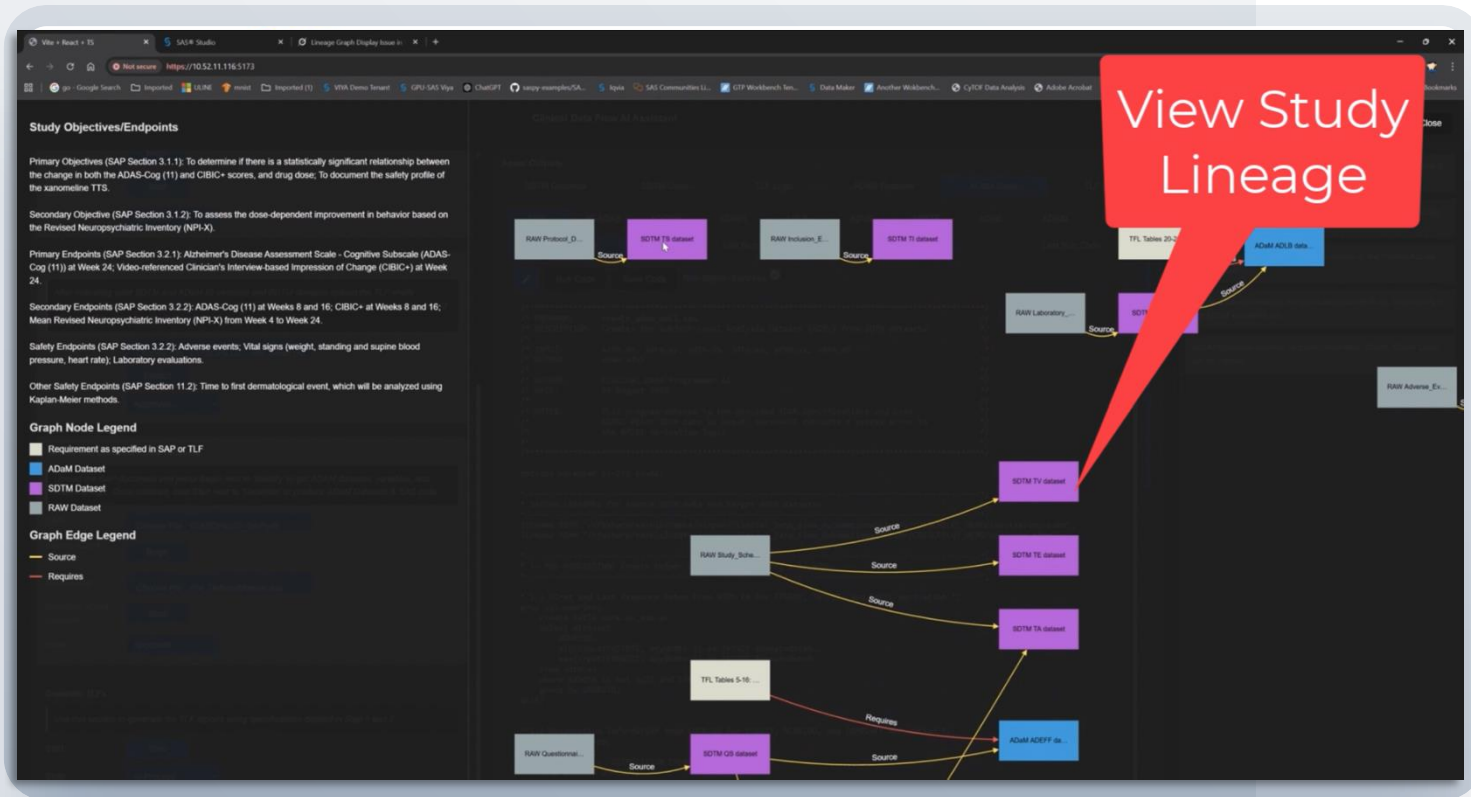
/* Define LIBNAMEs for source Raw data and target SDTM datasets. */
libname RAW "/nfeshare/sashts2/data/sinpan/Clinical_Data_Flow_AutomationwithLLMs/CDISCPLOT_DEMO/raw-data";
libname SDTM "/nfeshare/sashts2/data/sinpan/Clinical_Data_Flow_AutomationwithLLMs/CDISCPLOT_DEMO/tabulations-sdtm";

/* 1) Retrieve STUDYID from PROTOCOL_DEFINITIONS (constant across the study). */
/* - STUDYID: Set to the study identifier from PROTOCOL_DEFINITIONS.STUDY_ID. */

proc sql noprint;
  select distinct STUDY_ID into :STUDYID trimmed
  from RAW.PROTOCOL_DEFINITIONS
  having STUDY_ID=MIN(STUDY_ID);
```

Continue with your standard operating process to validate code and data





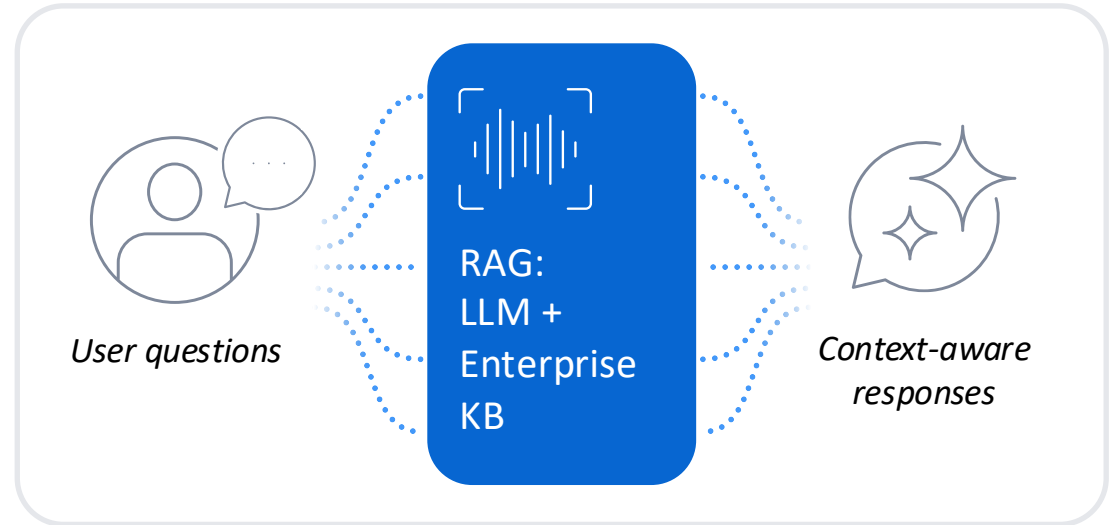
View complete
Study Lineage

RAG:

The Enterprise Game-Changer

Retrieval-Augmented Generation = search + generate

so outputs are **grounded** in your
enterprise knowledge base



Why RAG matters to users:

Factuality & Trust

Model cites the source;
reduces hallucinations

Freshness

Pulls latest policies/specs
at answer time—no
retraining

Governance

Keeps enterprise access
controls & auditing in the
loop

Flexible

Not locked to a specific AI
model

Reimagining the Clinical Data Scientist Role:

Future-Proofing Skills and Partnerships

PhUSE SDE Focus

Reimagining the clinical data scientist role means embracing a future where technical excellence, strategic thinking and cross-functional collaboration converge to create greater scientific and operational impact. As technological acceleration reshapes the clinical development landscape, our role evolves from traditional programming to orchestrating intelligent, end-to-end data ecosystems. Future-proofing our skills requires combining expertise in analytics, data standards and quality with new capabilities offered by state-of-the-art technologies.

1. How our role may evolve and future skills in demand

SAS believes that the role of the data scientist will continue to evolve from that of a primarily technical specialist to a domain-focused subject matter expert who leverages AI and analytics to solve complex business challenges within their field of expertise.

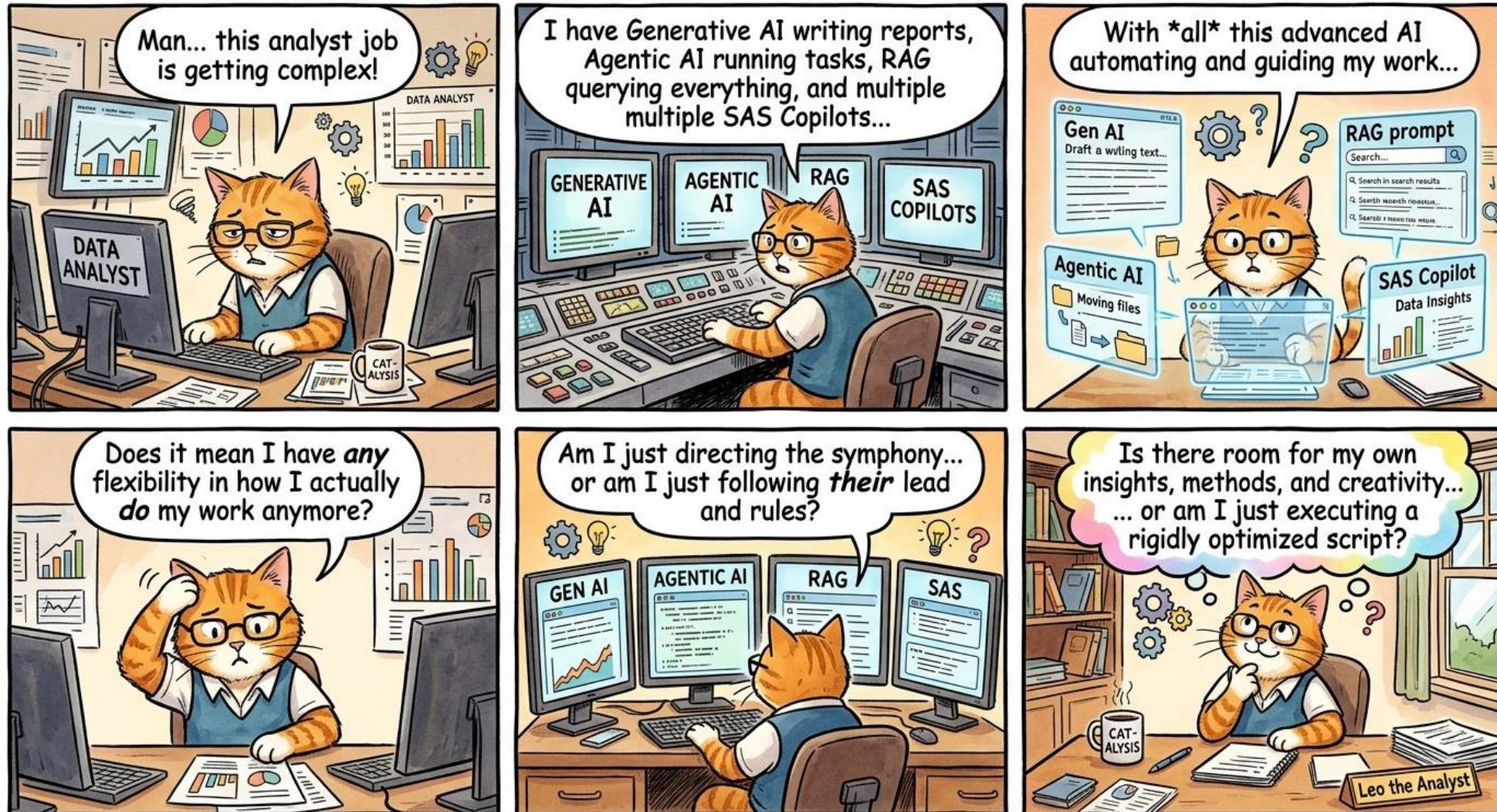
2. Tool transformation – practical examples of AI-assisted programming

SAS approaches tool transformation by **embedding Generative AI and Retrieval-Augmented Generation (RAG) directly into the analytics lifecycle**, enabling users to generate code, access contextual knowledge, automate complex programming tasks, and accelerate the path from data to insight while maintaining transparency and governance.

3. Closer integration with statisticians and clinical, regulatory and technology teams in more agile and digital environments

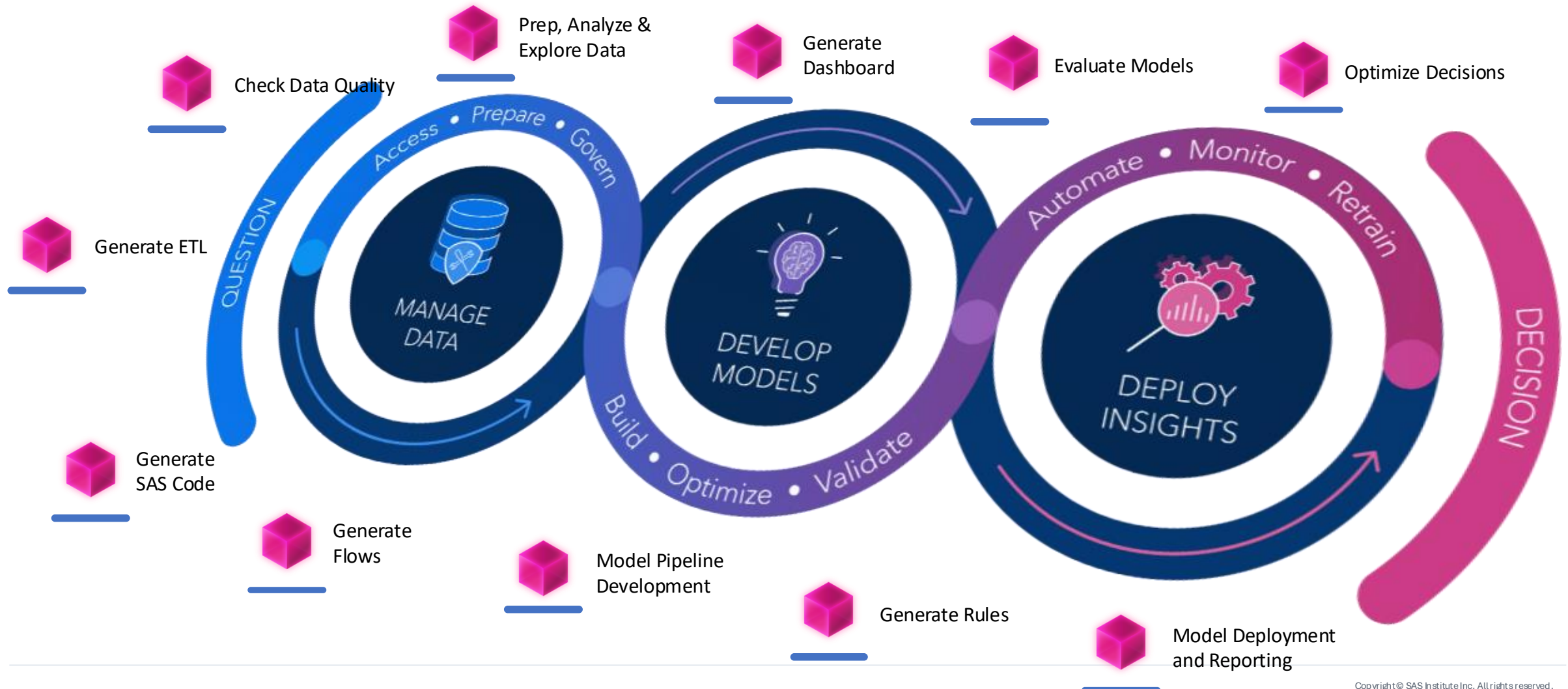
3. Closer integration with statisticians and clinical, regulatory and technology teams in more agile and digital environments:

Future-Proofing Skills and Partnerships



3. Closer integration with statisticians and clinical, regulatory and technology teams in more agile and digital environments:

Future-Proofing Skills and Partnerships



Agents:

The Enterprise Game-Changer

AI Agents = reason + act

so systems don't just answer – they can decide and act



Why Agents matter:

Autonomy & Action

Move from answering to acting — trigger workflows, updates, or decisions automatically

Personalization

Tailor responses as per Organizational context

Adaptability

Evaluate results, adjust strategies and improve in dynamic environments

Governance

Keeps enterprise access controls & auditing in the loop

SAS Viya MCP* Server

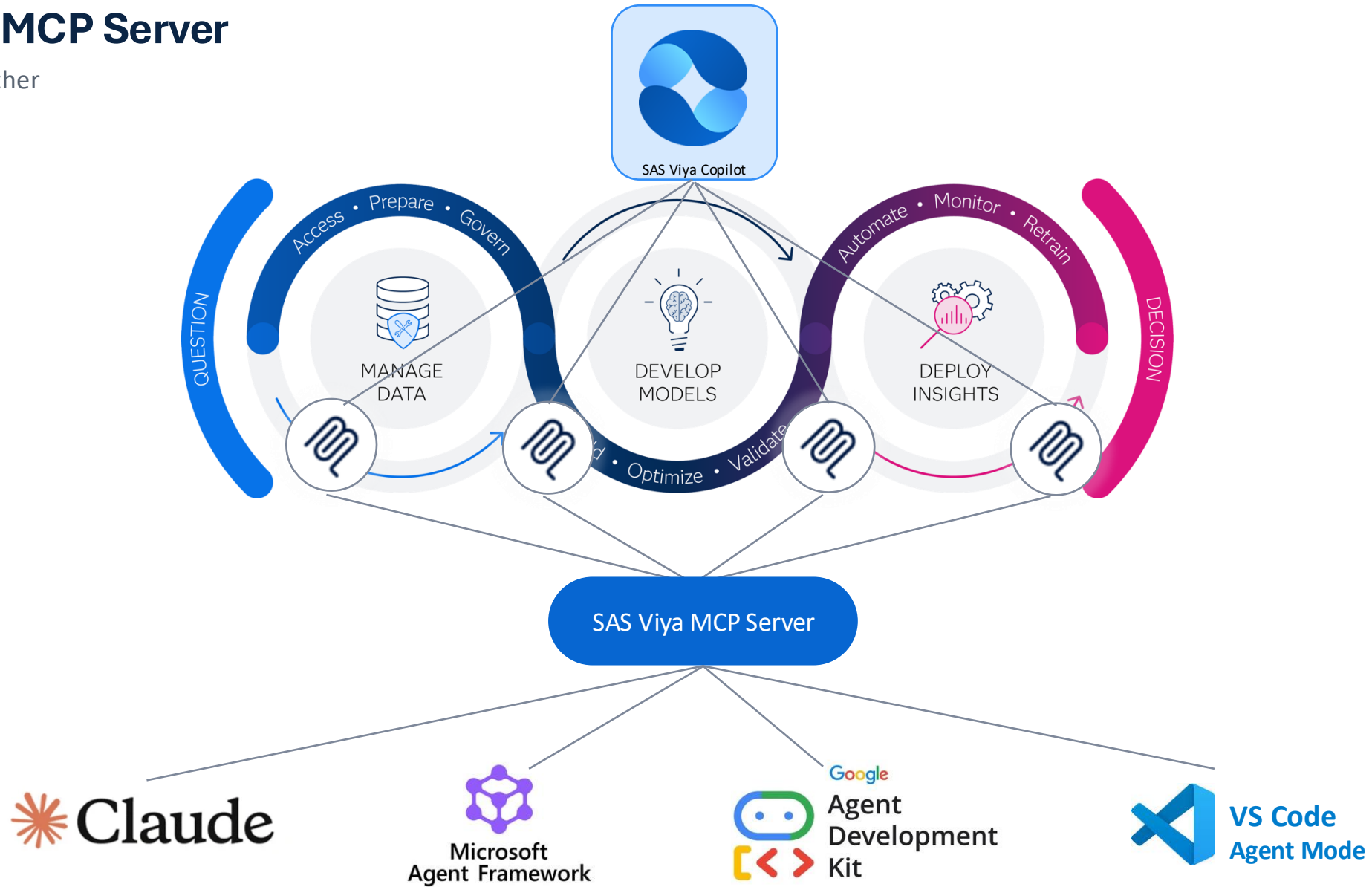
Tying it all together

- **Connect any AI assistant to SAS** by enabling tools such as Microsoft Copilot, Claude, and custom agents to securely access SAS analytics and decisioning capabilities.
- **Expose SAS as a trusted set of AI tools**, allowing agents to prepare data, execute SAS code, run models, score data, and trigger business decisions within governed workflows.
- **Bridge external LLMs with enterprise analytics**, combining conversational AI with SAS' trusted, explainable, and validated analytical engines.
- **Maintain governance and control through security**, auditability, transparency, and human oversight while AI agents interact with enterprise data and processes.
- **Enable interoperable agentic AI architectures, allowing organizations to use their preferred AI models** while leveraging SAS for analytics, decisioning, and operational execution.
- Accelerate the path from insight to action by allowing AI agents to move beyond conversation and **directly interact with SAS data, reports, models, and decision services**.

*MCP stands for Model Context Protocol. This is the open standard introduced by Anthropic for connecting AI assistants and agents to external tools, data sources, and systems.

SAS Viya MCP Server

Tying it all together



Reimagining the Clinical Data Scientist Role:

Future-Proofing Skills and Partnerships

PhUSE SDE Focus

Reimagining the clinical data scientist role means embracing a future where technical excellence, strategic thinking and cross-functional collaboration converge to create greater scientific and operational impact. As technological acceleration reshapes the clinical development landscape, our role evolves from traditional programming to orchestrating intelligent, end-to-end data ecosystems. Future-proofing our skills requires combining expertise in analytics, data standards and quality with new capabilities offered by state-of-the-art technologies.

1. How our role may evolve and future skills in demand

SAS believes that the role of the data scientist will continue to evolve from that of a primarily technical specialist to a domain-focused subject matter expert who leverages AI and analytics to solve complex business challenges within their field of expertise.

2. Tool transformation – practical examples of AI-assisted programming

SAS approaches tool transformation by embedding Generative AI and Retrieval-Augmented Generation (RAG) directly into the analytics lifecycle, enabling users to generate code, access contextual knowledge, automate complex programming tasks, and accelerate the path from data to insight while maintaining transparency and governance.

3. Closer integration with statisticians and clinical, regulatory and technology teams in more agile and digital environments

SAS views Agentic AI in Life Sciences as a trusted, governed digital collaborator that augments scientific and clinical experts **by automating complex analytical workflows while ensuring transparency, traceability, and regulatory compliance.**

From Programmer to AI-Enabled Clinical Data Scientist:

Accelerating Clinical Development with Trusted Copilots

Daniel Christen – Advisory Solutions Engineer –
HLS Solutions at SAS

Thank You