

SDTM Specification Automation Using an ML Approach to Provide Mappings

Benzi Alex Mathew, Jaya Simha Inampudi
(Saama Technologies)



US Connect 2025

Agenda



1. Introduction to the Challenge
2. Solution Overview
3. System Architecture
4. Key Features
5. Validation Results
6. Benefits & Future Direction



The Challenge

Traditional SDTM specification creation is

- Labor-intensive and time-consuming
- Prone to inconsistencies
- Heavily reliant on manual mapping
- Risk-prone due to human error
- Difficult to scale with increasing data volume

AI-powered automation can streamline this process while ensuring consistency.



AI-Powered Solution

An AI-powered SDTM mapping tool that:

- Automates SDTM specification creation
- Leverages historical mappings
- Incorporates CDISC standards
- Enables human oversight
- Provides editable specifications

Key Outcomes of the pilot

89.8% accuracy in variable prediction

50-70% time reduction in specification drafting

System Workflow Overview



Input Phase

- Raw clinical trial data & metadata.
- SDTM Implementation Guide, historical specifications.
- Global function definitions ensuring CDISC compliance.

Agentic Flow Phase

- HyDE Query Transformation
- Vector Database Processing
- Two-Stage Context Retrieval
- RAG Prompt Integration
- LLM-Based Generation and Validation

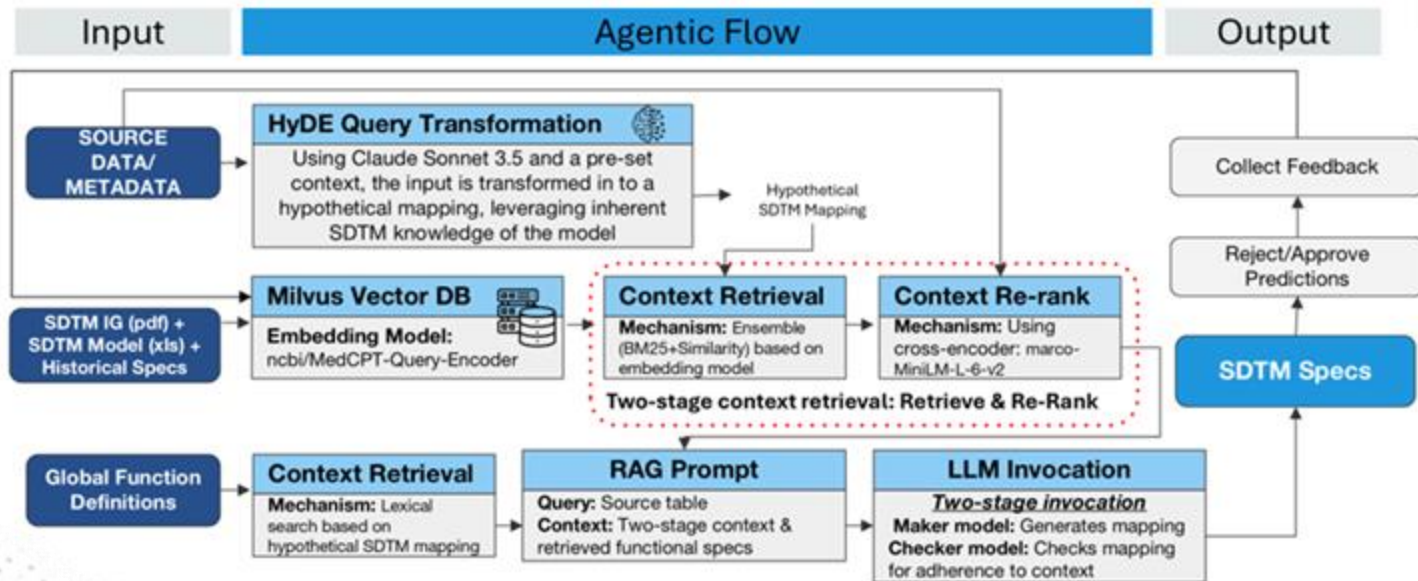
Output Phase

- SDTM Specification Generation
- Validation and Feedback Cycle



RAG Approach to SDTM Specs Generation

Solution Flow





Key Features of the Solution

1. Human-Centric Design
2. Regulatory Compliance by Default
3. Hybrid AI Architecture
4. Contextual Layering
5. Two stage LLM Validation
6. Human in the Loop Feedback

Evaluation

- SME review of recommended vs. historical mappings.
- Analysis of error patterns, edge cases, and complex scenarios.
- Tool fine-tuned based on observations.



Results

Tested on a Phase 2 study across multiple data types

Source Form	# Correct Variables Predicted	# Total number of Variables	% Accuracy
aesae	17	20	85.0%
cmeds	11	12	91.7%
demog	8	8	100.0%
drugdisp	5	5	100.0%
ecg	12	13	92.3%
eq5d3l	18	18	100.0%
exposure	14	14	100.0%
phexam	10	12	83.3%
vits	14	17	82.4%
Grand Total	83	118	92.7%



Key Benefits

- Significant reduction in specification creation time
- Improved consistency across studies
- Reduced manual effort
- Automated validation against SDTM IG
- Better utilization of historical mapping decisions
- Human oversight maintained for quality assurance

Future Enhancements

- Integration of latest LLM advancements
- Incorporation of Therapeutic Area User Guides (TAUGs)
- Enhanced pattern recognition for rare variables
- Continuous refinement based on user feedback

Summary



- AI-driven SDTM specification automation significantly reduces effort while ensuring compliance.
- Human oversight remains crucial for refining and validating AI-generated outputs.
- Continuous learning ensures system improvement over time.



Benzi Alex Mathew

Head, Clinical AI & Analytics Solutions and Services
(CAAS)

VP & Partner

Saama

benzi.mathew@saama.com

Jaya Simha Inampudi

Sr Director, Clinical Solutions

Saama

jaya.inampudi@saama.com

