PharmaSUG 2026 - Paper AI-259

AI-Augmented SDTM Review: A Practical Framework Enhanced by a Structured Prompt Library

Vihar Patel, Thermo Fisher Scientific, PPD Clinical Research Services

ABSTRACT

SDTM review is foundational to clinical trial data quality and regulatory readiness. Historically, this has been a manual and time-intensive process involving extensive cross-checking of CRFs, mapping specifications, the Protocol, SAP, and SDTM datasets themselves. As the volume and complexity of clinical data grows, conventional review approaches struggle to scale effectively.

Recent advancements in AI, particularly Large Language Models (LLMs), offer a new paradigm: AI-Augmented SDTM Review, where AI assists by performing structured data checks, semantic reasoning, document alignment, and anomaly detection while human programmers retain responsibility for adjudication and interpretation. This aligns with emerging regulatory expectations on responsible AI use, which emphasize human-centric design, clear context of use, transparency, and governance when AI supports data used in regulatory decision-making.

A key innovation in this paper is a curated library of more than 10 specialized AI prompts designed to guide review workflows, including CRF reconciliation, protocol-window validation, safety logic checks, derivation readiness, and submission-readiness assessment.

The proposed framework is positioned as a pre-rule-based screening and decision-support tool that enhances data quality early in the review process and complements deterministic validation tools such as Pinnacle 21. Human oversight remains essential to resolve ambiguity, manage edge cases, and ensure reproducibility.

This paper presents a practical, implementable workflow supported by detailed fabricated examples, diagrams, expanded AI prompts, evaluation comparisons, and both SAS and R pseudocode illustrating how responsible AI integration can improve SDTM review efficiency, accuracy, and regulatory confidence.

INTRODUCTION

SDTM datasets are foundational to clinical trial submissions, supporting downstream activities such as ADaM derivations, TLF generation, and regulatory review. As a result, the quality and consistency of SDTM data directly impacts submission readiness and regulatory confidence.

However, SDTM review remains a labor-intensive and error-prone process, requiring extensive cross-referencing across CRFs, mapping specifications, Protocol, SAP, and SDTM datasets. This process is inherently complex, involving both structured validation and context-driven interpretation across multiple domains.

As clinical data volume and complexity continue to increase, traditional review approaches—primarily based on manual checks and deterministic validation—face limitations in scalability and coverage.

Recent advancements in artificial intelligence (AI), particularly large language models (LLMs), provide an opportunity to enhance SDTM review by enabling automated, context-aware analysis across datasets and study documents. These capabilities allow for earlier detection of inconsistencies, cross-domain discrepancies, and potential data quality issues.

This paper introduces an AI-Augmented SDTM Review framework, where AI is used to support structured data checks and reasoning, while human reviewers retain responsibility for interpretation, validation, and regulatory decision-making. The approach is designed to complement existing validation tools, such as Pinnacle 21, and align with emerging expectations for responsible AI use in clinical data workflows.

AI-AUGMENTED SDTM REVIEW (CONCEPT AND FRAMEWORK)

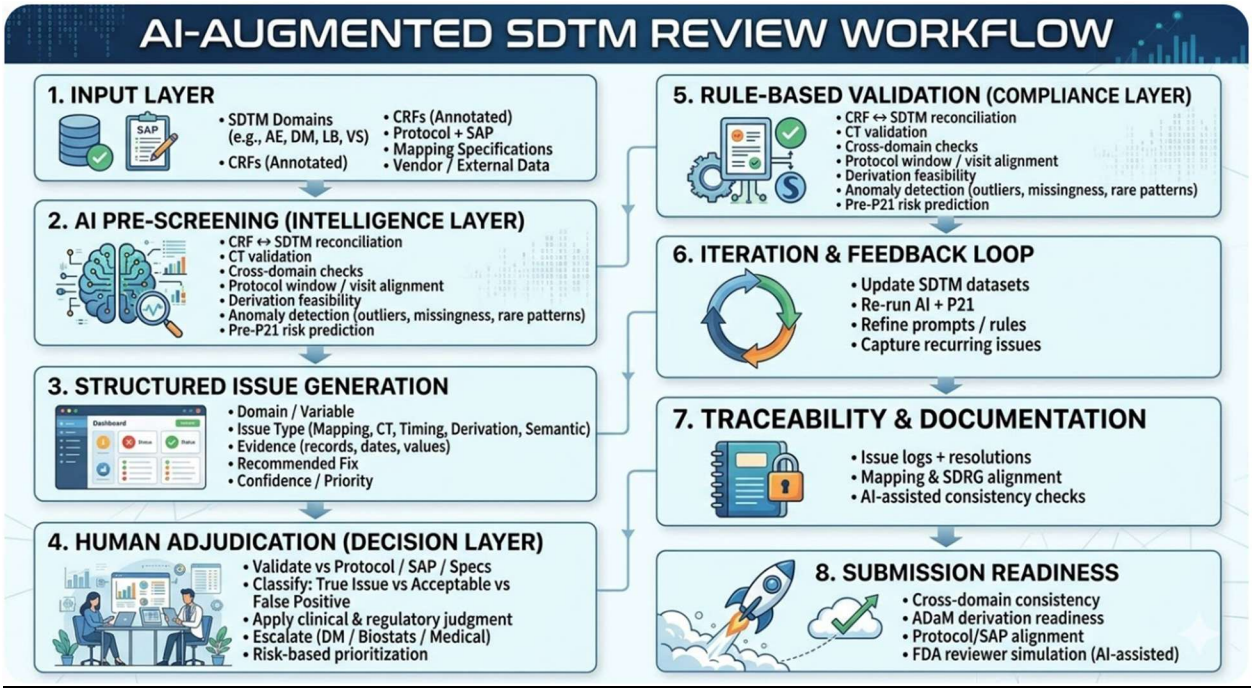
AI-Augmented SDTM Review is a layered approach where AI performs structured, automated checks and humans validate and interpret the results. As shown in the Figure below, AI is used early in the process to review SDTM data in the context of CRFs, Protocol, and SAP, while human experts make final decisions, supported by standard validation tools.

AI is well suited to rapidly analyze large datasets, identifying inconsistencies across domains, comparing SDTM data with study documents, suggesting mapping improvements, and detecting potential derivation issues early. It helps uncover patterns and gaps that may not be easily identified through rule-based checks alone.

The outputs from AI are not final decisions, they are structured findings that guide review. Human reviewers evaluate these findings using regulatory knowledge, medical context, and clinical judgment to determine what constitutes a true issue is and how it should be addressed.

Rule-based tools such as Pinnacle 21 and SAS/R checks are still essential to ensure compliance and reproducibility. In this approach, AI expands the scope of review, while traditional tools confirm adherence to standards.

This combined approach improves review coverage, efficiency, and consistency, while maintaining the quality and reliability required for regulatory submissions.



AI-DRIVEN SDTM REVIEW SCENARIOS (REALSTIC PHASE 3 EXAMPLES)

The following examples are based on a representative Phase 3 clinical study scenario spanning multiple SDTM domains, including DM, SV, EX, AE, LB, CM, QS, and SUPPQUAL. These scenarios are intentionally constructed to reflect common and high-impact data quality issues encountered during SDTM review, such as mapping inconsistencies, protocol window deviations, cross-domain discrepancies, and derivation misalignment.

For each example, AI outputs were generated using structured, task-specific prompts from the curated prompt library described later in this paper. These prompts guide the AI to perform targeted checks (e.g., CRF reconciliation, protocol window validation, AE-LB consistency), ensuring consistency and reproducibility of results.

The examples demonstrate how AI-assisted review can identify both structural issues (e.g., mapping, CT violations) and semantic inconsistencies (e.g., clinical plausibility, cross-domain conflicts), while still requiring human interpretation and validation.

Example 1: CRF Mapping Errors - Missing CRF Field in SDTM

AI Output	Human Review Consideration
🔎 CRF field "Date of first angioedema episode" missing in SDTM 🔎 Suggest SUPPMH variable: QNAM=FSTANGDT, QVAL=<date> 🔎 Potential data loss during mapping	✓ Confirm field exists in source/CRF ✓ Add SUPPMH mapping using appropriate QNAM value ✓ Update mapping specifications and traceability

Example 2: Timing & Visit Structure Issues - SV and EX Mismatch

SV dataset:

VISIT	VISITNUM	SVSTDTC	SVSTDY
WEEK 8	8	4/20/2024	96

EX dataset:

VISIT	EXSTDTC	EXSTDY
WEEK 8	2/26/2024	42

Protocol Window: Week 8: Day 57 ± 7

AI Output	Human Review Consideration
🔎 WEEK 8 visit date (SVSTDTC) in SV out-of-window by +39 days (Day 96 vs expected Day 57±7) 🔎 EXSTDTC inconsistent with VISIT assignment 🔎 Visit timing mismatch across domains	✓ Validate actual visit date vs recorded VISITNUM/VISIT ✓ Correct VISIT or date mapping ✓ Confirm whether late/early visit needs documentation

Example 3: AE–LB Inconsistency (Cytopenia Example) - AE Severity Not Supported by Labs

AE dataset:

AETERM	AESEV	AESTDTC
Neutropenia	Grade 3	3/1/2024

LB dataset:

LBTESTCD	LBORRES	LBSTNRLO	LBSTNRHI
NEUT	3.5	1.5	7
NEUT	3.2	1.5	7

CTCAE v 4.03: Grade 3 Neutropenia requires ANC < 1.0.

AI Output	Human Review Consideration
🔎 AE "Neutropenia Grade 3" reported with Lab values within normal range 🔎 Severity not supported by data; likely miscoded or requires query	✓ Verify AE grading criteria used ✓ Check for missing lab values ✓ Query site or correct AE severity

Example 4: Prohibited CM Use (Protocol Violation) - Prednisone Rescue Limited to 3 Days

CM Dataset

CMTRT	CMSTDTC	CMENDTC	CMINDC
Prednisone	2024-04-01	2024-04-20	Urticaria

AI Output	Human Review Consideration
🔎 Prednisone used for 19 days 🔎 Exceeds allowed rescue duration as per protocol (≤3 days) 🔎 Potential protocol deviation	✓ Confirm clinical justification ✓ Add DV record if needed ✓ Notify data management/medical monitor

Example 5: Incorrect UAS7 Derivation (SAP Misalignment) - Missing Daily Diary Entries

QS Dataset for USUBJID= ABC-001 where VISIT=Week 6

QSCAT	QSSCAT	QSTESTCD	QSTEST	QSDTC	QSORRES
DAILY DIARY	HIVE SCORE	HSSD1	Hive Severity Score Day 1	2/1/2024	1
DAILY DIARY	HIVE SCORE	HSSD2	Hive Severity Score Day 2	2/2/2024	.
DAILY DIARY	HIVE SCORE	HSSD3	Hive Severity Score Day 3	2/3/2024	.
DAILY DIARY	HIVE SCORE	HSSD4	Hive Severity Score Day 4	2/4/2024	1
DAILY DIARY	HIVE SCORE	HSSD5	Hive Severity Score Day 5	2/5/2024	.
DAILY DIARY	HIVE SCORE	HSSD6	Hive Severity Score Day 6	2/6/2024	2
DAILY DIARY	HIVE SCORE	HSSD7	Hive Severity Score Day 7	2/7/2024	.
DAILY DIARY	ITCH SCORE	ISSD1	Itch Severity Score Day 1	2/1/2024	1
DAILY DIARY	ITCH SCORE	ISSD2	Itch Severity Score Day 2	2/2/2024	1
DAILY DIARY	ITCH SCORE	ISSD3	Itch Severity Score Day 3	2/3/2024	0
DAILY DIARY	ITCH SCORE	ISSD4	Itch Severity Score Day 4	2/4/2024	1
DAILY DIARY	ITCH SCORE	ISSD5	Itch Severity Score Day 5	2/5/2024	0
DAILY DIARY	ITCH SCORE	ISSD6	Itch Severity Score Day 6	2/6/2024	2
DAILY DIARY	ITCH SCORE	ISSD7	Itch Severity Score Day 7	2/7/2024	.
WEEKLY SCORE	DERIVED	HSS7	Weekly Hive Severity Score Over 7 Days	2/7/2024	4
WEEKLY SCORE	DERIVED	ISS7	Weekly Itch Severity Score Over 7 Days	2/7/2024	5
WEEKLY SCORE	DERIVED	UAS7	Urticaria Activity Score Over 7 Days	2/7/2024	9

SAP: ≥4 non-missing entries per subscore (ISS & HSS) required.

AI Output	Human Review Consideration
🔎 HSS has <4 non-missing days 🔎 UAS7 is incorrectly derived, it should be missing 🔎 SAP rule violated	✓ Recalculate UAS7 per SAP ✓ Validate diary completeness ✓ Ensure derivation logic is corrected

Example 6: SUPPQUAL Naming Issues - Excessive QNAM Length

SUPPEX dataset

QNAM	QVAL
DOSDELAYREAS	"AE"

AI Output	Human Review Consideration
🔎 QNAM exceeds CDISC naming best practices 🔎 Inconsistent abbreviation 🔎 Suggested: DLYREASON	✓ Rename to standard format (e.g., DLYREASON) ✓ Update metadata/specs ✓ Ensure consistency across SUPPQUAL

Example 7: Visit Occurrence Inconsistency (SV vs Vendor Data)

SV Dataset

USUBJID	VISIT	VISITNUM	SVOCCUR
ABC-010	EOT	24	N

AI Output	Human Review Consideration
🔎 Visit exists in vendor data but SVOCCUR = N 🔎 Inconsistent visit occurrence flag 🔎 Cross-source discrepancy	✓ Reconcile EDC vs vendor data ✓ Update SVOCCUR if appropriate ✓ Document discrepancy if needed

Example 8: DS.DSSEQ Derivation and Sorting Inconsistency

Expected sorting order per SDTM specs: STUDYID, USUBJID, **DSSTDTC**, DSCAT, DSSCAT, DSDECOD, DSTERM

AI Output	Human Review Consideration
🔎 DSSEQ not aligned with DSSTDTC sort order 🔎 Records out of sequence 🔎 Sorting logic inconsistent	✓ Check SDTM specs and re-sort dataset using defined keys ✓ Re-derive DSSEQ ✓ Ensure consistent derivation logic

AI PROMPTS FOR SDTM REVIEW

The following prompts are structured, actionable, and designed to guide large language models (e.g., GPT-5.3, Claude Opus, Gemini) in performing targeted SDTM review tasks. Each prompt focuses on a

specific review objective, such as CRF reconciliation, protocol window validation, cross-domain consistency, or derivation checks.
These prompts were used to generate the AI outputs presented in the examples above and are intended to ensure consistency, reproducibility, and transparency in AI-assisted review. They can be adapted and reused across studies to support standardized SDTM review workflows.

Prompt 1: Full CRF-to-SDTM Reconciliation

- Using the CRF pages, SDTM datasets, and SDTM mapping specifications provided, identify:
  - CRF fields missing entirely from SDTM
  - CRF fields mapped incorrectly or partially
  - Values in SDTM that do not match CRF response options
  - Cases where SUPPQUAL should have been used but was not
- For each issue, recommend the correct SDTM variable, domain, and mapping rule.
- Return results as a structured table: CRF Item, SDTM Issue, Recommended Fix.

Prompt 2: CT Validation Audit

- Validate all SDTM variables that use controlled terminology.
- Check for:
  - Values not found in CDISC CT
  - Inconsistent casing or spelling
  - Values that match older or deprecated CT versions
  - Incorrect terminology domains (e.g., AESEV mapped using wrong severity scale)
- List all violations and propose corrected terminology.

Prompt 3: Detect Semantic CT Errors

- Identify SDTM variables where values are technically valid CDISC terms but semantically incorrect for the context. Examples:
  - CM class codes used for AEs
  - Race terms misapplied
  - AE relationship values reversed or mis mapped
- Provide corrected CT usage and rationale.

Prompt 4: AE-Lab Consistency

- Compare AE records with corresponding LB records, identify:
  - AEs whose severity is not supported by lab results
  - Clinical lab abnormalities missing an AE term
  - Discrepancies in timing between AE onset and abnormal labs
  - AEs requiring lab confirmation (e.g., cytopenias) without supporting labs
- Produce a table: AE Term, Expected Lab Pattern, Actual Lab, Issue, Suggested Correction.

Prompt 5: EX-SV Visit Alignment

- Using protocol visit definitions, identify:
  - Exposure records with wrong VISIT or VISITNUM
  - Visits occurring outside protocol windows
  - SV records missing for visits where dosing occurred
  - Early/late dosing patterns not documented in DS or EX
- Return the table with: USUBJID, Visit, Issue, Correction.

Prompt 6: Protocol Window Validation

- Reconstruct expected visit windows from the protocol text.
- Compare SVSTDTC, EXSTDTC, and LB/Vital Signs dates, Flag:
  - Visits outside allowed windows
  - Wrong VISITNUM assignments
  - Out-of-sequence visits
  - Missing visits

- Return a structured list with protocol citation, observed issue, and fix.

Prompt 7: SAP-Driven Variable Checks

- Using SAP definitions (e.g., baseline, endpoint, imputation rules), identify variables in SDTM that do not satisfy ADaM derivation feasibility. Examples:
    - Missing baseline values
    - Incorrect baseline timing
    - Missing required flags (e.g., ASEV)
    - Required endpoints not derivable from SDTM
- Return findings with SAP rule references.

Prompt 8: Multidomain Safety Signal Detector

- Scan AE, CM, LB, VS, and EX domains.
- Identify clusters indicating potential safety signals:
    - Hypertension + drug exposure timing
    - Liver enzyme elevations + hepatotoxic drugs
    - Neutropenia + immunosuppressants
    - QTc prolongation patterns
- Provide a summary of possible signals with evidence paths.

Prompt 9: Pattern Detection for Unexpected Data

- Using unsupervised pattern analysis, highlight:
    - Outliers
    - Rare AEs
    - Unusual medication changes
    - Sudden lab spikes/drops
- Explain potential causes and recommend clinical queries.

Prompt 10: Variable-Level SDTMIG Compliance

- Check each SDTM variable for:
    - Correct TYPE (char/num)
    - Proper LENGTH
    - Allowed values
    - Required variables present
    - Permissible variables used correctly
- Identify all SDTMIG v3.2 violations.

Prompt 11: FDA Reviewer Simulation

- Simulate an FDA reviewer assessing the SDTM package, identify:
    - Data inconsistencies are likely to trigger reviewer queries
    - Missing traceability
    - Implausible medical patterns
    - Poorly documented derivations
    - Discrepancies between stated and implemented mappings
- Provide a list of predicted FDA questions.

Prompt 12: SDRG Consistency Check

- Check SDTM domain contents against SDRG text, identify:
    - Missing justifications
    - Outdated descriptions
    - Incomplete data issues
    - Wrong algorithm descriptions
- Return recommended SDRG updates.

Prompt 13: Pinnacle 21 Error Prediction

- Predict Pinnacle 21 findings BEFORE running P21, identify:
  - Missing values causing Rejects
  - Invalid CT values
  - Incorrect formats
  - Duplicate records
  - Future dates
- Provide a P21-style error table.

Together, this prompt library serves as a practical foundation for integrating AI into SDTM review in a controlled and scalable manner.

DISCUSSION

Complementary Roles in AI-Augmented SDTM Review

As described in the preceding sections, AI, deterministic validation tools, and human reviewers each contribute distinct strengths to the SDTM review process. Rather than overlapping, these roles are complementary and operate together to provide comprehensive coverage.

AI primarily enhances the breadth of review by identifying potential issues across domains, documents, and derivations. Rule-based tools ensure adherence to established standards and provide reproducible validation results. Human reviewers remain responsible for interpreting findings, resolving ambiguity, and ensuring alignment with study-specific and regulatory expectations.

The table below summarizes how these responsibilities are distributed across key review capabilities.

Capability	AI (Intelligence Layer)	Rule-Based Checks (P21/SAS/R)	Human (Decision Layer)
CRF-to-SDTM mapping gaps	✓ Identifies gaps, inconsistencies, and missing mappings	✓ Detects predefined mapping checks (if programmed)	✓ Confirms and resolves
Cross-domain reasoning	✓ Detects complex and non-obvious relationships	✓ Detects rule-based cross-domain checks (limited scope)	✓ Interprets clinical relevance
Protocol window validation	✓ Flags potential deviations using context	✓ Applies programmed window checks	✓ Confirms acceptability
Controlled terminology checks	✓ Detects syntactic and semantic issues	✓ Enforces standard terminology rules	✓ Validates context and appropriateness
Deterministic rule validation	✕	✓ Core function	✓ Reviews outputs
Reproducibility	Requires governance	High	High
Interpretation of ambiguity	✕	✕	✓ Core responsibility

Impact of AI-Augmented Review on Issue Detection

The results demonstrate that AI significantly improves issue detection coverage, particularly for cross-domain inconsistencies, mapping gaps, and derivation-related issues that are often missed in manual review.

However, not all AI-identified findings represent true errors. Some require interpretation, and others may reflect acceptable study-specific conventions or data collection nuances. This reinforces the need for human adjudication and deterministic validation to ensure accuracy and reproducibility.

Together, these results support the use of AI as a pre-validation and detection layer, complementing traditional validation tools and expert review.

Category	Manual Review (Before AI)	AI Detected	Human Confirmed
Mapping Gaps	3	7	6
CT Violations	1	4	4
Timing Issues	2	6	6
AE–LB Conflicts	1	5	4
Derivation Errors	0	3	3
Total Issues	**7**	**23**	**20**

The preceding discussion establishes both the distribution of responsibilities and the measurable impact of AI-assisted review. The broader implications of this approach are discussed below.

Interpretation and Implications

AI-Augmented SDTM review represents a shift from traditional, rule-driven validation to a more integrated and proactive review model. By combining AI-driven detection with deterministic validation and human expertise, the process moves beyond isolated checks toward a more holistic assessment of data quality.

This approach enables earlier identification of complex, cross-domain issues while improving overall review efficiency without compromising regulatory expectations. At the same time, it highlights the need for clear governance, as AI-generated findings must be consistently interpreted, validated, and documented.

Importantly, the value of this framework lies not in replacing existing methods, but in orchestrating AI, deterministic validation, and human judgment into a cohesive workflow. This creates a more scalable and adaptable SDTM review process aligned with increasing data complexity and evolving regulatory expectations.

CONCLUSION

AI-Augmented SDTM Review provides a practical and scalable advancement in clinical data quality review. By integrating AI-driven detection, deterministic validation, and human expertise, this approach significantly improves issue detection coverage while maintaining regulatory rigor and reproducibility.

This paper presented a structured framework supported by realistic examples and a curated prompt library, demonstrating how AI can be systematically applied to SDTM review. The results show that AI can uncover substantially more issues, particularly complex, cross-domain, and derivation-related, than manual review alone.

AI does not replace clinical programmers, it amplifies their ability to detect, interpret, and resolve data issues earlier and more effectively. When implemented with appropriate governance and human oversight, this hybrid model enables a shift from reactive validation to **proactive, intelligence-driven data review**.

AI-Augmented SDTM Review is not a future concept, it is **ready for adoption today**, providing a clear path toward more efficient, consistent, and submission-ready clinical data processes.

ACKNOWLEDGMENT

I would like to express my heartfelt gratitude to my colleagues for their invaluable support and collaboration. Special thanks to Jhelum Naik, Ryan Wilkins, Ivy Ni and Sajeet Pavate for their insightful feedback and unwavering encouragement.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Vihar Patel
Thermo Fisher Scientific, PPD Clinical Research Services
201-736-2760
viharpatel64@gmail.com

DISCLAIMER

The opinions expressed in this paper and related slides for the presentation are solely those of the presenter and not necessarily those of Thermo Fisher Scientific. Thermo Fisher Scientific does not guarantee the accuracy or reliability of the information provided herein.
Brand and product names are trademarks of their respective companies.