



2026 CDISC China Interchange

AI-Native SDTM Automation

Hao Chen, Chunpeng Zhao





Disclaimer

The following presentation is delivered by the speaker in his/her individual and personal capacity. All content, views, opinions, interpretations, and any statements made — whether oral or written — are those of the speaker alone and shall not be attributed to, or deemed to be endorsed, adopted, or authorized by, his/her employer or any of its affiliates, officers, directors, or shareholders.

Agenda

01

Background

Context and problem statement driving this initiative

02

Architecture

System design, key components and technical architecture

03

Metrics

Key performance indicators and how success is measured

04

Demo

Live walkthrough of the solution in action

05

Summary

Broader AI strategy alignment and SP expertise

Background

Context and problem statement driving this initiative



[Anthropic — Finance Agents](#)
[Bloomberg — Anthropic AI for financial research](#)
[GitHub — anthropic/financial-services](#)



SDTM Transformation: A Critical AI Opportunity

1 Regulatory-critical

Mandatory CDISC standard for global regulatory submissions

2 Data complexity explosion

Rapid growth in data volume, velocity, and diversity across clinical trials

3 Operational bottleneck

Traditional SDTM transformation is labor-intensive, slow, and error-prone

4 Decision lag

Rule-based and copilot automation cannot enable real-time, adaptive decision-making

Standards Matters in AI Era

pharmaverse blog

Table of contents

What an Unskilled Agent
Actually Does

The Package as
Specification

What the Skill Does

The Accountability Anchor

What the Skill Does

The [{admiral}](#) skills in pharma-skills do not replace [{admiral}](#). They connect the AI agent to it. A skill provides curated, domain-aware guidance: which functions to use for which derivations, how to structure the program for QC readability, which variables require special handling, and what assertions to include. The skill is the bridge between a capable general-purpose model and the specific, validated logic the pharmaverse community has built.

The benchmark results reflect this directly. Across ADSL, ADAE, ADVS, and ADLB:

- **With skill: 88–100%** pass rates across domains
- **Without skill: 17–59%** pass rates, with high variance

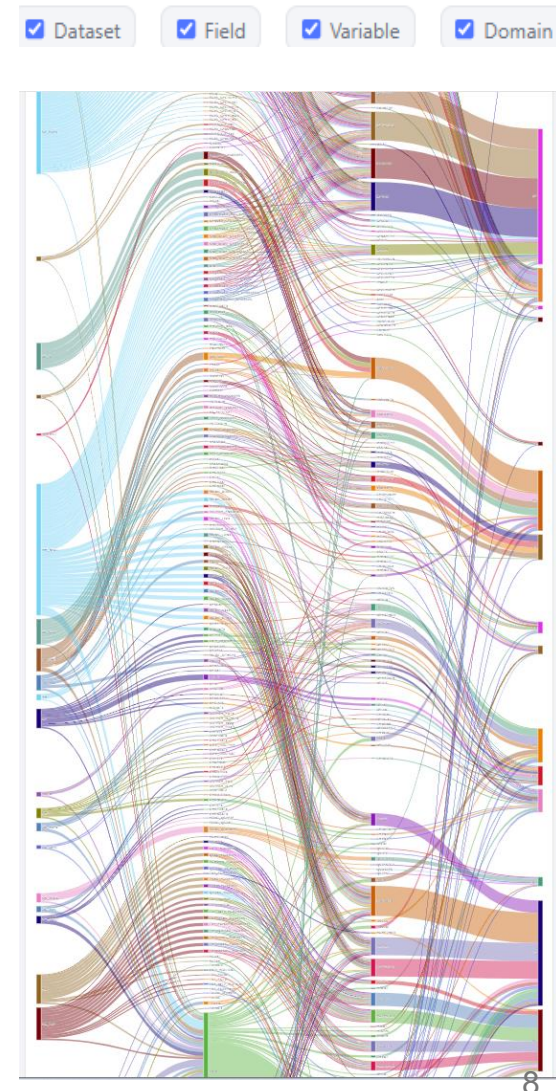
That variance in the unskilled condition matters as much as the mean. Inconsistent output is not a defensible process in a GxP context. A skill-guided agent produces consistent, traceable, [{admiral}](#)-anchored code. An unskilled agent produces something different every time.

[Why we still need {admiral} in an age of AI – pharmaverse blog](#)

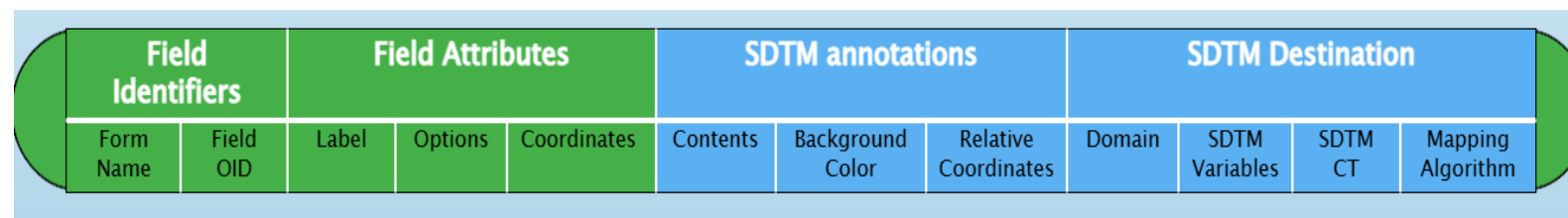
The Challenge

CDISC Mapping Complexity

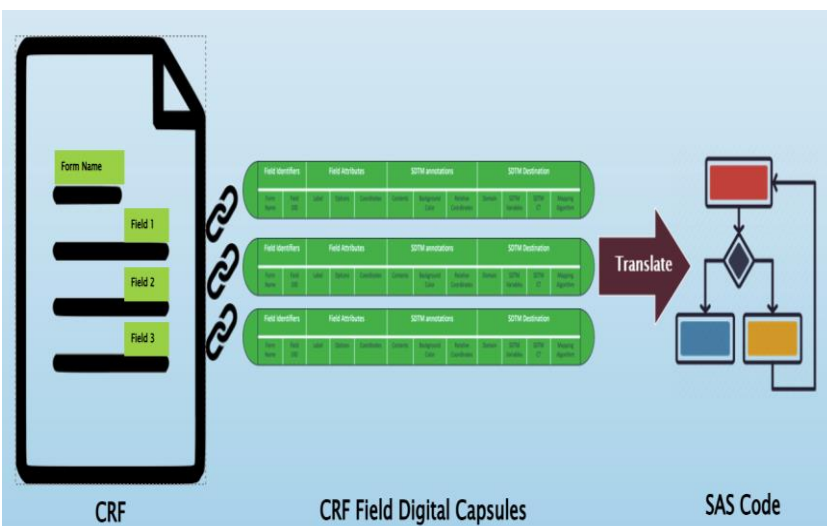
- **Structural reshaping** — many source datasets must be deconstructed and reassembled into CDISC domains + SUPPQUAL + CO.
- **Cardinality complexity** — 1:1, 1:n, n:1, transpose, and value-change transformations between EDC and SDTM.
- **Annotation fragility** — CRF page-position shifts cause inaccurate mapping.
- **Loss of linkage on transpose** — complex transposes break data relationships between fields and records.
- **Constant standard drift** — CRF library grows every version; the statement remains qualitative with counts removed.
- **Broad domain coverage** — CDISC domains span Trial Design, Special Purpose, Interventions, Events, Findings, and Relationships.
- **Heavy derivation load** — CRF copy, assigned, derived, protocol, and eDT categories all contribute materially; exact percentages removed.
- **Study-specific derivations** — derived variables change per study, requiring flexible reusable macros.
- **Weak variable-role context** — ID, Topic, Qualifier, and Timing linkage must be preserved for accuracy.
- **Process pain points** — accuracy, protocol reading, data quality, and long-term maintenance across the full pipeline.



Parsing SDTM annotations



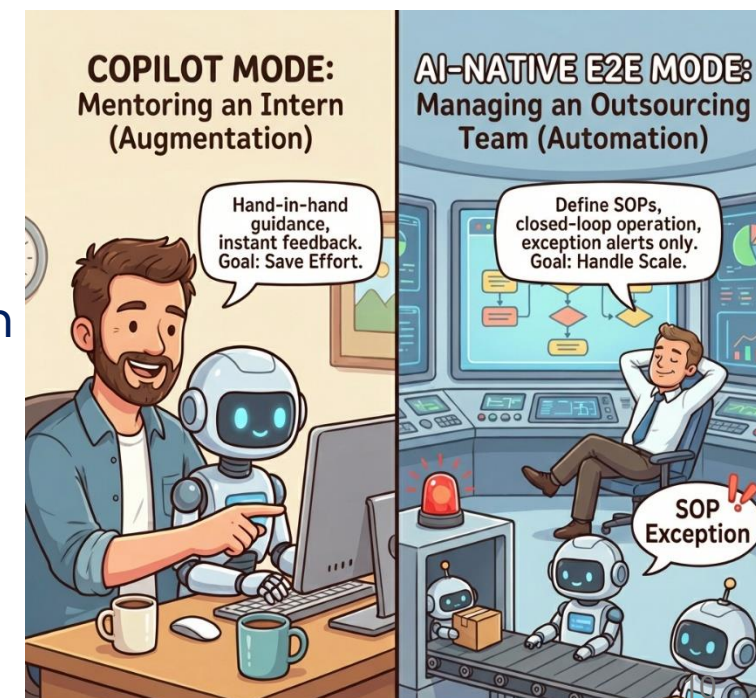
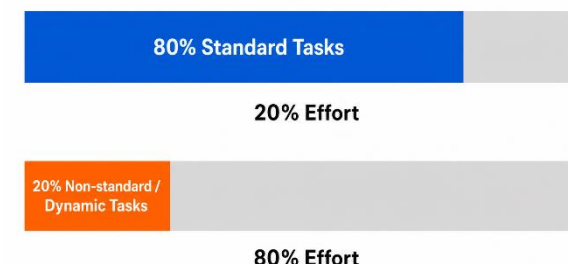
Source Dataset	FieldName	SDTM Annotation	FieldValue	Domain	Dataset	SDTM RecordID	Variable Name	VariableValue	Algorithm
DM	Form	DM = DemograpShics		DM	DM		DOMAIN	DM	Assigned
DM	BRTHDAT	BRTHDTC		DM	DM		BRTHDTC		DT2DTC
DM	AGE	AGEU = YEARS		DM	DM		AGEU	YEARS	Assigned
DM	AGE	AGE		DM	DM		AGE		CopyUcase
DM	SEX	SEX = F	Female	DM	DM		SEX	F	IfAssigned
DM	SEX	SEX = M	Male	DM	DM		SEX	M	IfAssigned
DM	DMCBP	RPTEST = Childbearing Potential		RP	RP	DMCBP	RPTEST	Childbearing Potential	Assigned
DM	DMCBP	RPORRES = N	No	RP	RP	DMCBP	RPORRES	N	IfAssigned
DM	DMCBP	RPTESTCD = CHILDPOT		RP	RP	DMCBP	RPTESTCD	CHILDPOT	Assigned
DM	DMCBP	RPORRES = Y	Yes	RP	RP	DMCBP	RPORRES	Y	IfAssigned
DM	DMCBP	RP = Reproductive System Findings		RP	RP		DOMAIN	RP	Assigned
DM	RACEOTH	SUPPDM.QVAL where QNAM = RACEOTH		DM	SUPPDM		RACEOTH		CopyUcase



Source: Chunpeng Zhao, Mijun Hu, etc.
SDTM Annotated CRF Digitalization, PharmaSUG China 2021
Real-time SDTM, PharmaSUG China 2022

Evolution of SDTM Mapping

- 1 Manual Programmer-Centric**
Study-specific programs; limited reuse; long cycle times
- 2 Metadata-Driven**
Central repositories store mapping rules; specs auto-generated(copied)
- 3 AI/ML-Assisted Mapping (NLP, Classification)**
ML models suggest mappings
- 4 Generative-AI Copilots**
LLM assistants draft code and explain standards with human review
- 5 AI-Native End-to-End**
Orchestrated workflows/Agent from Raw to SDTM under human governance



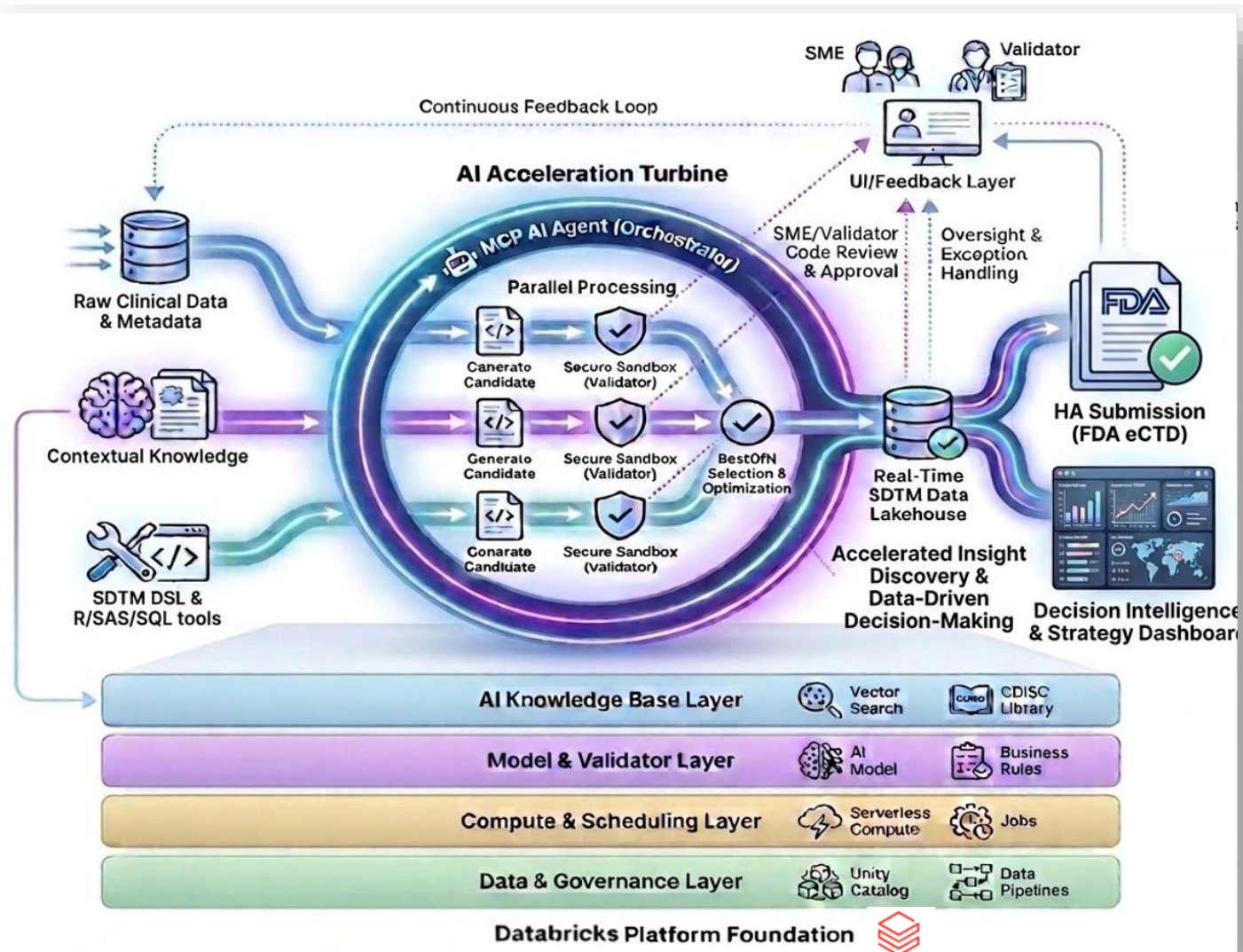
■ General-Purpose LLMs ■ Embedded or Task-Specific GenAI

Design and Architecture

System design, key components and technical architecture

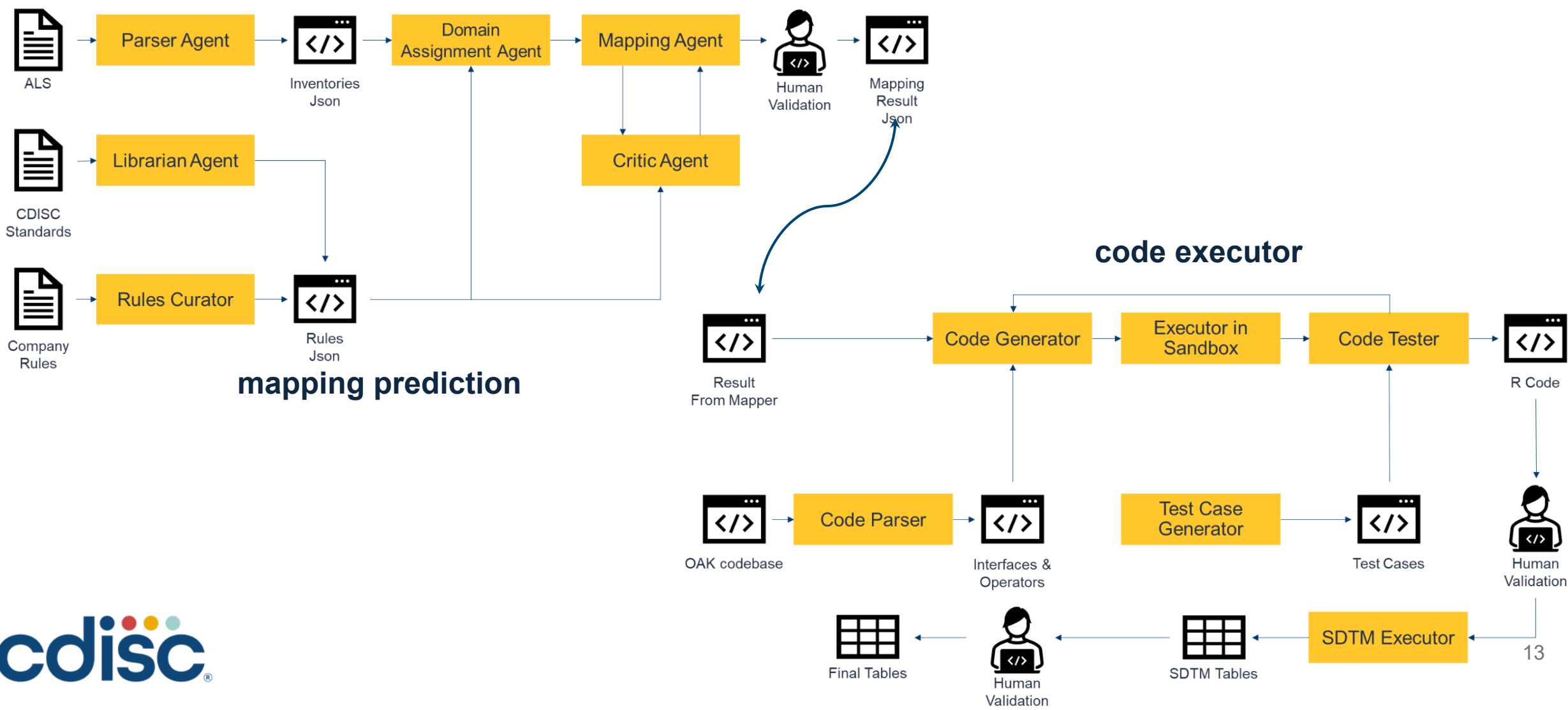


Real-time SDTM Data Mapping and Analysis



- ✓ **AI Spec Prediction**
- ✓ **SAS/R Code Generation**
- ✓ **Execution in Sandbox**
- ✓ **Automatic Testing**
- ✓ **Human-in-the-loop Validation**

Mapping Prediction and Code Executor Agents

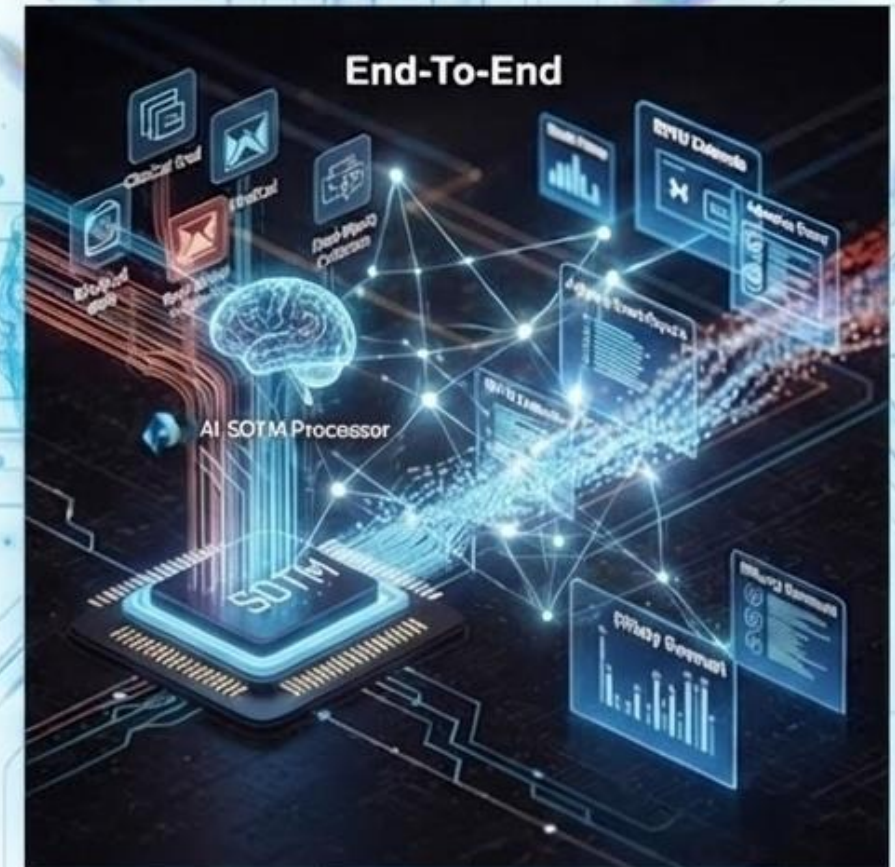
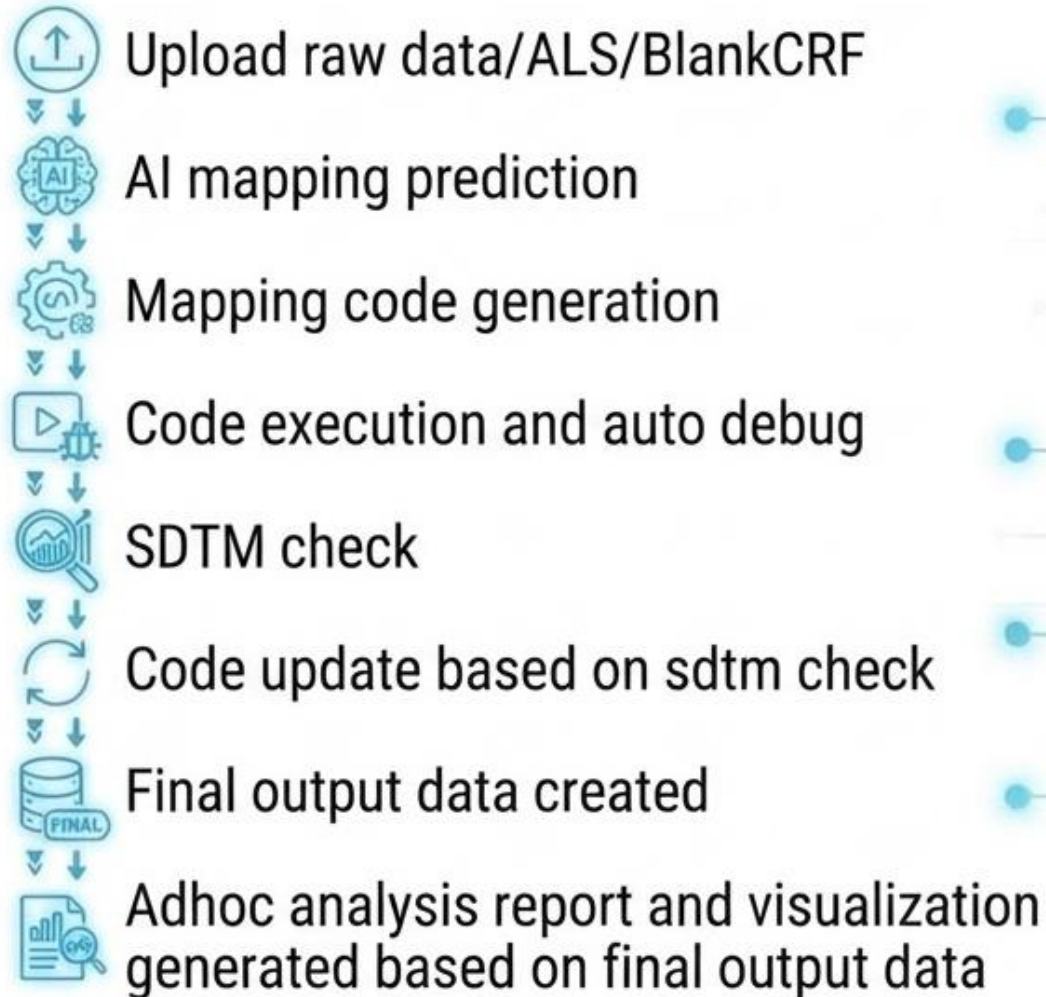


Core Capabilities Powered by Three Agent Engines



- All-in-One integrated web App
- Dramatic efficiency gain – Reduce SDTM-minus mapping time from Days to hours and even to minutes
- Broad data support - Supports Both corporate and industry standard raw data
- No mapping spec needed - Only metadata from Raw data / ALS / BlankCRF /Data Description
- Real Time analysis and Full traceability
- Seamless AI+HI integration

End-to-End Workflow



Metrics

Key performance indicators and how success is measured



Prediction Metrics Sample

VSTEST = Vital Signs if No

Were vital signs taken?

VSTESTCD = VSALL if No

VSSTAT = NOT DONE if No

Yes

No

1

VSREASND

Reason for assessment not done

COVID-19

Other

2

VSTEST = Weight

Weight at pre-dose

VSORRES where VSTESTCD = WEIGHT

VSORRESU

kg

lb

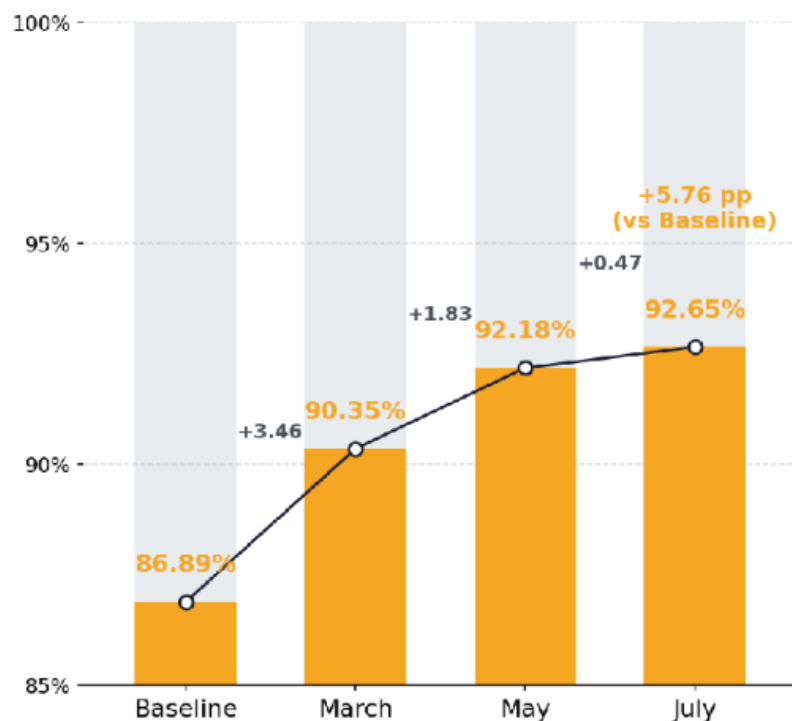
3

AI Prediction	Ground Truth	Is this field an exact match?			
SVSTDTC	SVSTDTC	1	1	1	✓
SVOCCUR = Y; SVSTDTC	SVOCCUR = Y; SVSTDTC	2	2	2	✓
SV = Subject Visits;	SV = Subject Visits; SVPRESP = Y; SVOCCUR = Y	2	3	2	✗
DC = Demographics as Collected; DS = Disposition; DSCAT = PROTOCOL MILESTONE; EPOCH = SCREENING	DC = Demographics as Collected; DS = Disposition; DSCAT = PROTOCOL MILESTONE	4	3	3	✗
Total:		9	9	8	
		Field accuracy: 2/4=50% (A field is counted as a match only when all annotations within it are matched.)		Annotation Rule accuracy: 8/9=89% (Each correctly matched annotation is counted individually.)	

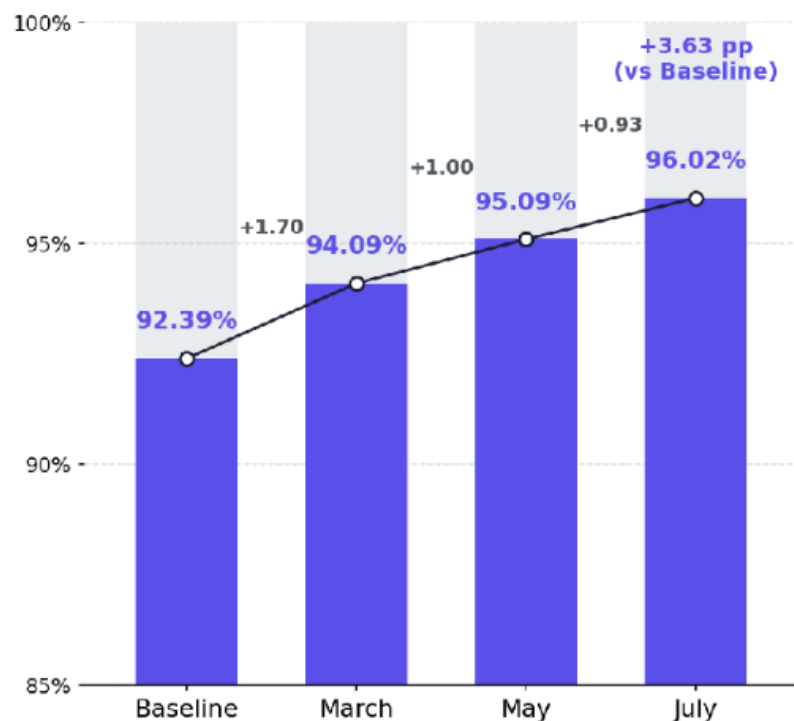
Metrics Summary of Mapping Prediction

Overall Accuracy Progress — Baseline → March → May → July

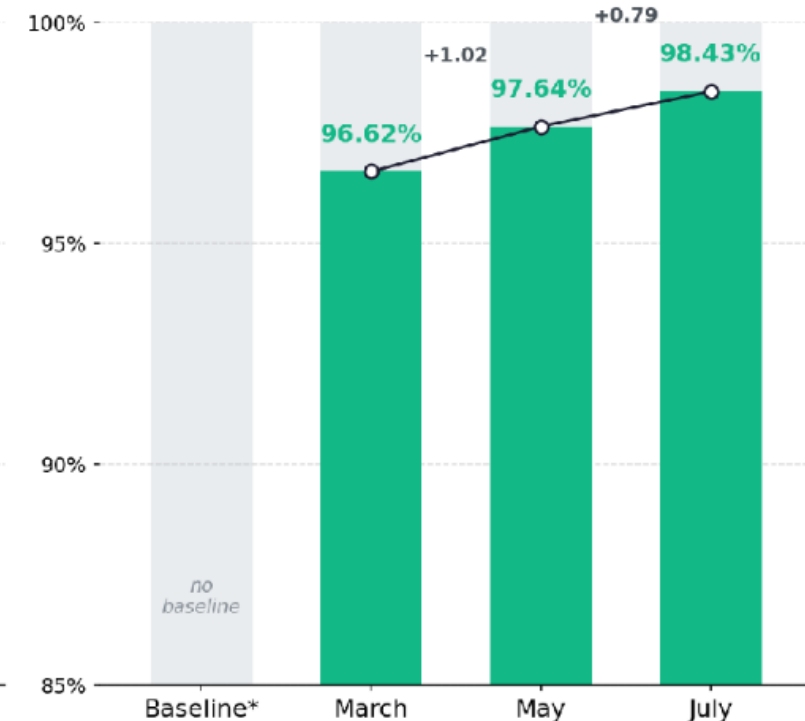
Field Accuracy (Exact)



Annotation Rule Accuracy



Variable (Name) Accuracy

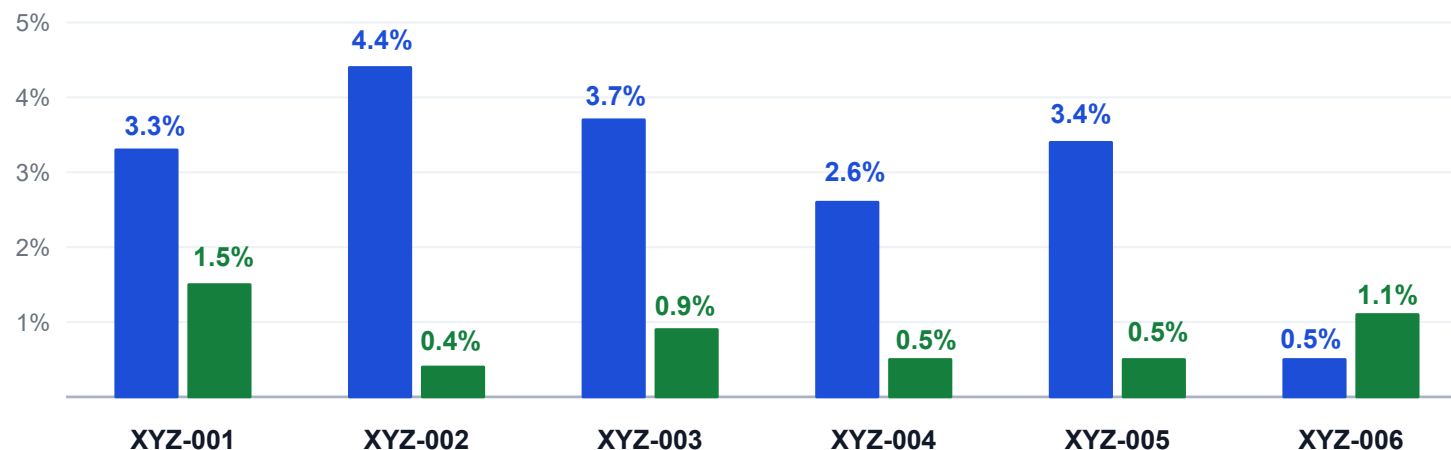


Colored bars = achieved accuracy · Gray = remaining gap to 100% | *Variable Accuracy has no Baseline-method score (new metric)

Metrics for Output Data Comparison

Error rate by study — Variables vs. record grouping

■ Variable ■ Record grouping



WEIGHTED AVG — VARIABLES

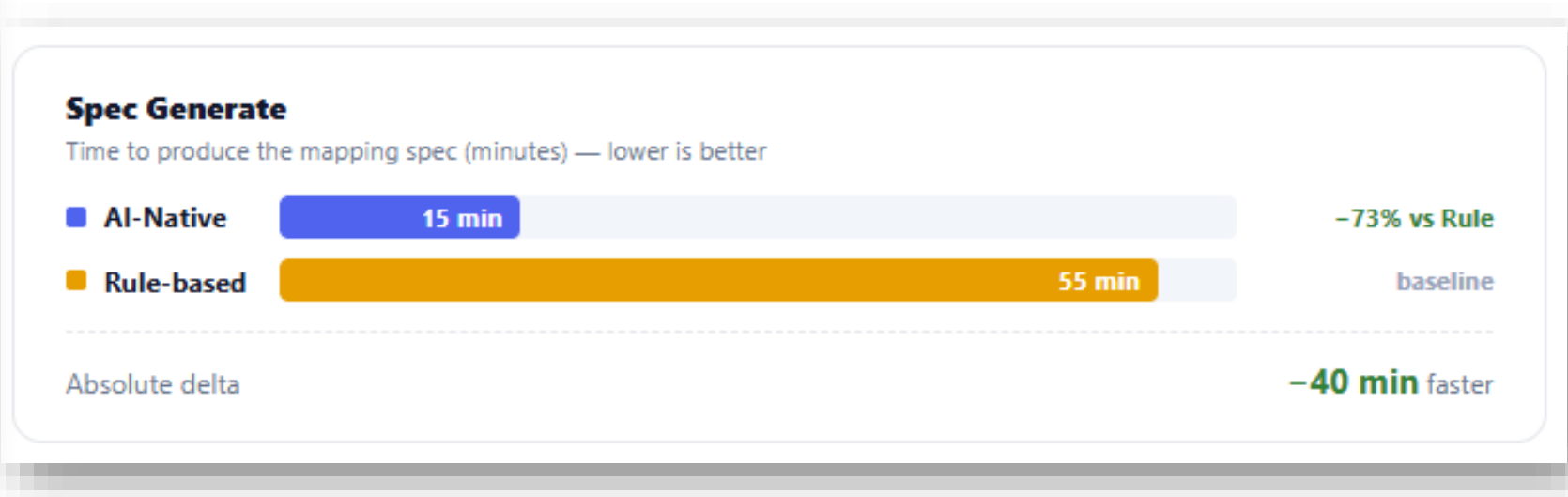
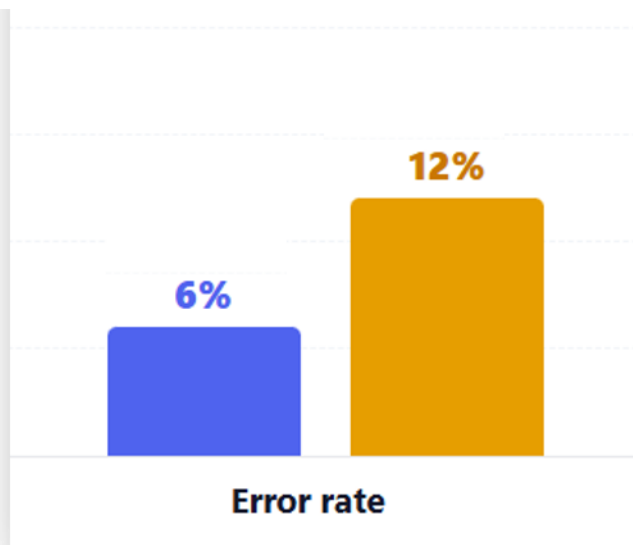
2.9%

WEIGHTED AVG — RECORD GROUPING

0.8%

Weighted avg = Σ numerators / Σ denominators

Metrics from apples to apples compare



The Task: Translate a Whole SAS Macro System into R

Not one macro — the entire engine that generates SDTM, rebuilt as an R system

OVERALL PERFECT RATE

>99%

The remaining diffs are mostly reference-side UTF-8 artifacts

FROM · the SAS macro system

- A top-level SDTM macro + a whole family of sub-macros
- Deeply nested, shared state, call-order dependent
- Metadata-driven: spec → generator → domain code → data
- The production engine — our authoritative gold standard



TO · a clean R system

- An R code generator producing the same SDTM data
- Modular, domain-organized, human- & AI-readable
- Must match SAS cell-for-cell — variable-level exact
- Reproducible & cloud-native (Databricks / renv)

► Not a literal port — a system-level rebuild that must still match SAS cell-for-cell.

Hybrid Inference Strategy: Local 14B + Prompt Fallback

RECOMMENDED thr ≥ 0.999 · LOCAL COVERAGE

63.4%

5,746 / 9,070 need no external model

FIELD LOSS (VS PROMPT-ONLY)

-1.95pp

91.53% $\rightarrow$ 89.58%, keeps 98% of quality

EXTERNAL API TOKENS

↓ 63%

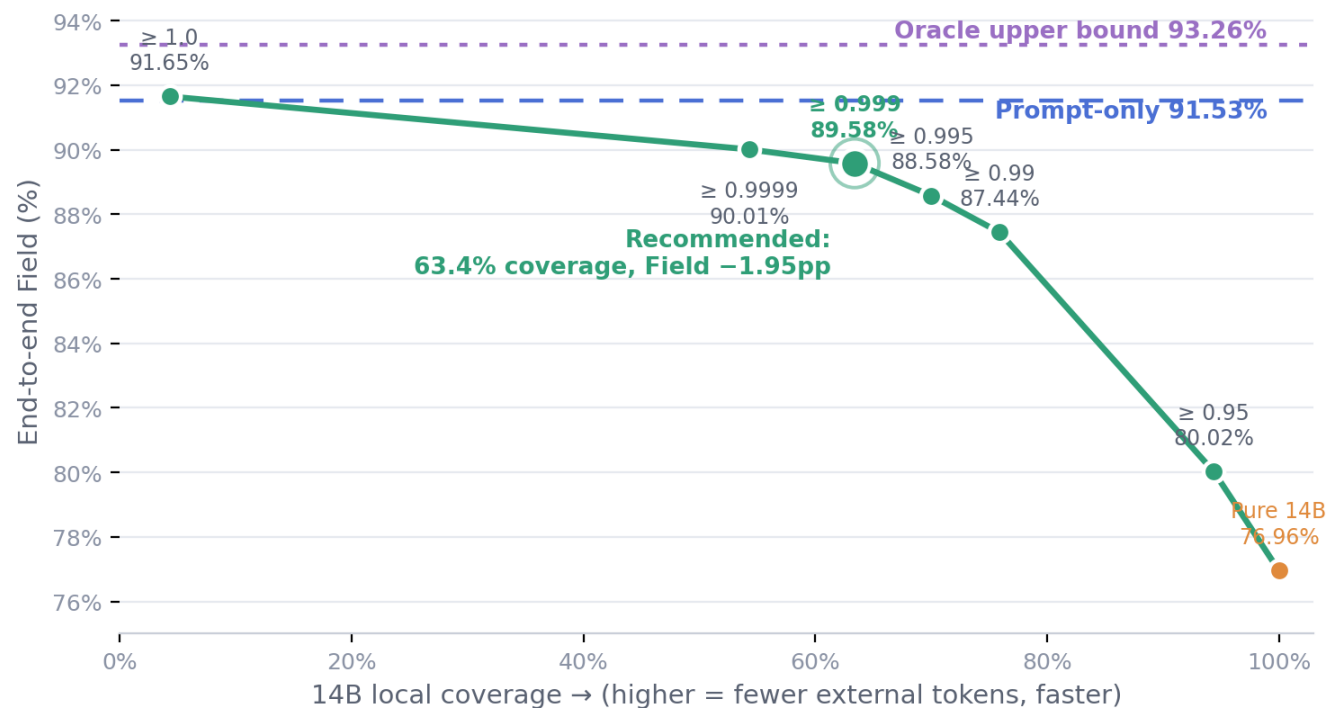
Only 36.6% of samples call Claude

14B ACCURACY ON CONFIDENT SUBSET

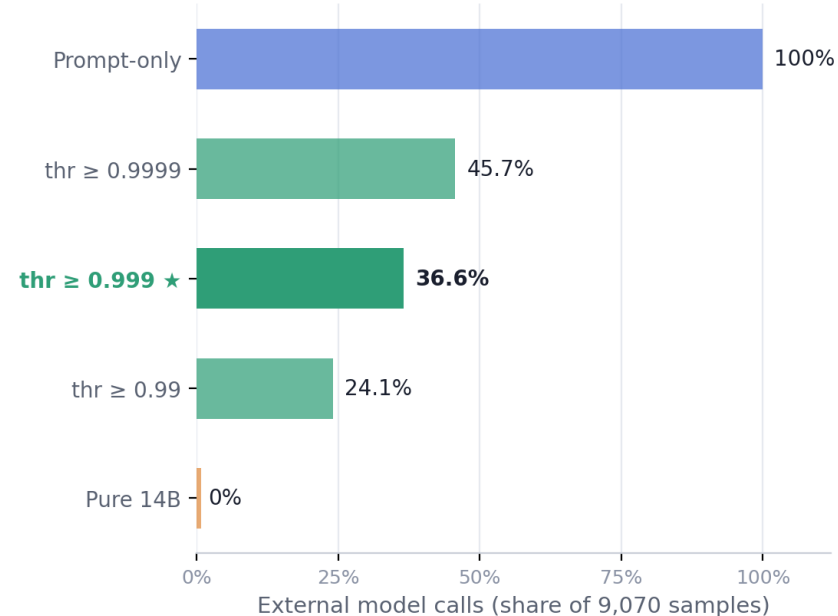
89.58%

Field%, vs 76.96% overall

Trade-off: local coverage vs end-to-end Field accuracy



External API call volume vs Prompt-only



Recommended thr ≥ 0.999 : 63.4% of samples run fully local (faster inference, zero external token cost) at only -1.95pp Field accuracy.

Demo

Live walkthrough of the solution in action



1 Track 1 — SAS Platform

- **Solution on SAS foundation.**

2 Track 2 — E2E SDTM R Solution

- **A) UI + AI Agent : Auto Mode / Step-by-Step Mode**
 - **B) AI Agent only: Chat / Skill / API Mode**
-

Substance Use 1 page



A1	A	B	C	D	E	F
1	SDTMIGVersion	CRFID	SourceDataset	UForm	FieldSec	FieldName
2	SDTMIG v3.3	ISPP_DEMO_V8	SUBJECT	Subject		0 Form
3	SDTMIG v3.3	ISPP_DEMO_V8	SUBJECT	Subject		1 SUBJID
4	SDTMIG v3.3	ISPP_DEMO_V8	SUBJECT	Subject		2 COUNTRY
5	SDTMIG v3.3	ISPP_DEMO_V8	SV	Date of Visit		0 Form
6	SDTMIG v3.3	ISPP_DEMO_V8	SV	Date of Visit		1 SVOCCUR
7	SDTMIG v3.3	ISPP_DEMO_V8	SV	Date of Visit		1 SVOCCUR
8	SDTMIG v3.3	ISPP_DEMO_V8	SV	Date of Visit		1 SVOCCUR
9	SDTMIG v3.3	ISPP_DEMO_V8	SV	Date of Visit		2 SVSDAT
10	SDTMIG v3.3	ISPP_DEMO_V8	DS_IC	Informed Consent		0 Form
11	SDTMIG v3.3	ISPP_DEMO_V8	DS_IC	Informed Consent		0 Form
12	SDTMIG v3.3	ISPP_DEMO_V8	DS_IC	Informed Consent		1 DSSTAT
13	SDTMIG v3.3	ISPP_DEMO_V8	DS_IC	Informed Consent		1 DSSTAT
14	SDTMIG v3.3	ISPP_DEMO_V8	DS_IC	Informed Consent		1 DSSTAT
15	SDTMIG v3.3	ISPP_DEMO_V8	DS_IC	Informed Consent		2 ICPV
16	SDTMIG v3.3	ISPP_DEMO_V8	DS_OIFC	Optional informed consent		0 Form
17	SDTMIG v3.3	ISPP_DEMO_V8	DS_OIFC	Optional informed consent		0 Form
18	SDTMIG v3.3	ISPP_DEMO_V8	DS_OIFC	Optional informed consent		1 OIFC_DSCTYP
19	SDTMIG v3.3	ISPP_DEMO_V8	DS_OIFC	Optional informed consent		1 OIFC_DSCTYP
20	SDTMIG v3.3	ISPP_DEMO_V8	DS_OIFC	Optional informed consent		3 OIFC_DSSTAT

25

Track 2a — E2E SDTM Mapping and R Analysis - UI + AI Agent

Track 2b — E2E SDTM Mapping and Analysis - AI Agent only

收到! 参数齐全, 直接开始全自动流程:

现在获取 output_path:

crf2sdmspec_count:

output_path: annotat

现在转换 JSON → xlsx 并触发 R Codegen Job: ☒ S

MB, MH, MI, OE, PR, RE, RP, RS, SS, SU, SV, TR, TU, VS

触发 R Codegen Job: ☒ R Codegen Job 已触发! rs

帮我做study的SDTM, 这是ALS, CRFID, REFCRFID和rawdata path如下

帮我用ses数据生成TFL

Jul 10, 2026, 12:03 AM

根据刚生成的 SDTM 数据, 已有以下 Domains 可用: AE, CM, DC, DS, EG, LB, MH, VS 等。

请选择需要生成的 TFL (可多选):

Demography

DMT01 — Demographics Summary Table

Disposition

DST01 — Patient Disposition Table

Adverse Events

AET01 — TEAE Summary Table (Standard S

AET02 — TEAE by SOC/PT

AET03 — Serious TEAE

AET05 — Grade ≥ 3 TEAE (CTCAE)

AET06 — Related TEAE

AEL01 — AE Listing

Lab Results

LBT01 — Lab Parameters Mean/SD Summar

LBT04 — Lab Parameters Shift Table

LBL01 — Lab Listing

Vital Signs

VST01 — Vital Signs Mean/SD

VSL01 — Vital Signs Listing

ECG

EGT01 — ECG Parameters Mean/SD

EGL01 — ECG Listing

Concomitant Medications

CMT01 — Concomitant Medications Summ

LTG01 是按治疗组的实验室指标纵向轨迹图 (Lattice Plot)。先看一下 lb.c everything I need. render_comparison_html is text-only, so for the graph images directly. Let me write the script.LTG01 生成完成!

ltg01.html HTML

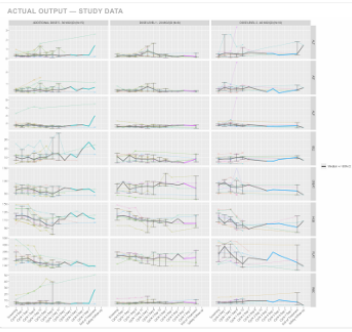
LTG01

Figure 1. Lattice Plot of Laboratory Tests by Treatment Group Over Time

Safety Population

Show Reference

ACTUAL OUTPUT — STUDY DATA



1. Source: LB domain

2. Safety Population: subjects with DSBCOD = 'ENTERED INTO TRIAL' (N=33)

3. Lab parameters shown: ALT, AST, ALP, BIL, CREAT (renal); HGB, PLAT, WBC (hematology)

4. Unscheduled visits excluded. X-axis ordered by scheduled visit sequence

5. Black line: Median per arm per visit. Error bars: 80% confidence interval of median (term:star_median_ci)

6. LBSTRESN used as AYAL. Individual subject trajectories shown in grey (alpha=0.3)

再画一张LTG01

将LTG01 生成html dashboard

Jul 10, 2026, 12:36 AM

Agent trace: 2 steps

Read

1. 2

Using a fresh execution context to avoid any cross-session s placeholder-based gsub replacement insteadExecution suc (Safety Population N=33, ADLB 记录 1515 行, 32 位受试者)。

ltg01_dashboard.html HTML

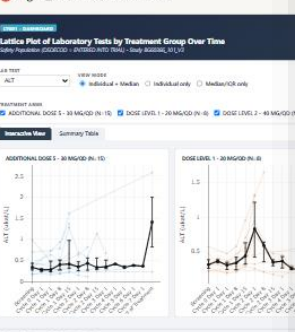
ltg01_dashboard.html

Figure 1. Lattice Plot of Laboratory Tests by Treatment Group Over Time

Safety Population (DSBCOD = 'ENTERED INTO TRIAL', N=33)

Show Reference

ACTUAL OUTPUT — STUDY DATA



1. Source: LB domain

2. Safety Population: subjects with DSBCOD = 'ENTERED INTO TRIAL' (N=33)

3. Lab parameters shown: ALT, AST, ALP, BIL, CREAT (renal); HGB, PLAT, WBC (hematology)

4. Unscheduled visits excluded. X-axis ordered by scheduled visit sequence

5. Black line: Median per arm per visit. Error bars: 80% confidence interval of median (term:star_median_ci)

6. LBSTRESN used as AYAL. Individual subject trajectories shown in grey (alpha=0.3)

再生成AET01和AEL01

Jul 10, 2026, 12:12 AM

先读取 AET01 和 AEL01 的参考模板: 检查 AE 域的编码: 数据结构清晰。开始写两个脚本:

FATAL: AESDTH == "Y" (本研究无死亡AE)

AEREL: "RELATED"/"NOT RELATED" (二值编码)

WD: AEACN == "DRUG WITHDRAWN" (6条)

DSM: AEACN == "DRUG INTERRUPTED" (48条)

AESEV: 不可用 → SEV = FALSE

DCSREAS: 2例因AE退出, 从DS派生

AESOC: 不可用 → 用 AEBODSYS 替代Now execute AET01:AET01 done. Now execute AEL01:Both executed successfully. Now downloading all 4 output files:AET01 and AEL01 均已完成。共生成 6 个输出文件:

AET01 — TEAE Overview Table

aet01.html HTML

AET01

Table 2. Overview of Deaths and Adverse Events

Safety Population

Show Reference

ACTUAL OUTPUT — STUDY DATA

	AET01 TOTAL: 33
Total number of patients with at least one AE	13 (40%)
Total number of AEs	15
Total number of deaths	2 (6%)
Total number of patients withdrawn from study due to an AE	2 (6%)
Total number of patients with at least one TEAE	13 (40%)
TEAE	6 (14)
DSM	4 (12)
DCS	2 (6)
DSM	6 (14)
DCS	10 (30)
DCS	2 (6)
DCS	2 (6)

odisco

TLG Catalog - Stable

功能:

化验指标选择器:ALT/AST/ALP/BIL/CREAT/HGB/PLAT/WBC

视图模式切换:个体轨迹 / 中位数+80% CI / 两者叠加

治疗组筛选: Arm 勾选显示/隐藏

实时 KPI 卡片:受试者数、观测数、访视数随筛选联动更新

Interactive View 标签页:每个 Arm 一张 Plotly 图,个体受试者

27

Service Spans Deliverables and Platform Access

A Data & Deliverables

- ✓ SDTM Data
- ✓ SDTM Data Analysis Report
- ✓ SDTM mapping spec
- ✓ SDTM mapping code

B Platform & Access

- ✓ SDTM mapping platform
- ✓ SDTM mapping API
- ✓ SDTM mapping chatbot
- ✓ SDTM mapping as Context for Other AI solution (better understand the Raw data via SDTM mapping)

Summary: Success = AI Strategy + SP Expertise

Right AI strategy and Good SP expertise



AI Strategy

- ✓ **Standard Flywheel:** Use **AI** to standardize the data and process; use those **standards** to make AI more accurate, stable, and reliable — a self-reinforcing loop.
- ✓ **Accuracy matters:** Metrics as the Goal. Ignore precision, and everything looks doable with AI. However, without production-grade accuracy, we remain far from moving PoCs into production and evolving chatbots into agents.
- ✓ **Technique hierarchy:** In-context learning → ontology → vector search. Each step down the ladder can drop AI accuracy by one to several orders of magnitude.
- ✓ **Engineering discipline:** Avoid over-reliance on complex LLM frameworks. Write custom code for fine-grained LLM control

Concept	Definition
Prompt Engineering	Telling the model what to do
Context Engineering	Determining what the model can see
Loop Engineering	Designing how the agent iterates toward goals
Harness Engineering	Ensuring agents operate reliably, safely, and at scale
Graph Engineering	Structuring relationships among knowledge, tools, and agents so they can reason and navigate effectively

SP Expertise



Business

- ✓ **SP understand the document/data/code from protocol to submission**
- ✓ **SP is data/code/Spec/detail-oriented**
- ✓ **SP understand the industry standards**



Technical

- ✓ **Write custom code for fine-grained LLM control**
- ✓ **SAS to R auto translation**
- ✓ **Model Fine-tuning**
- ✓ **Web App Development**
- ✓ **Databricks Platform**



Thank You



Name: Hao(Harry) Chen
E-mail: Mathhao@gmail.com

Name: Chunpeng Zhao
E-mail: Chunpeng.zhao@icloud.com



WeChat



赵春鹏 (Chunpeng)
上海 浦东新区



扫二维码，添加我为朋友。



LinkedIn