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.

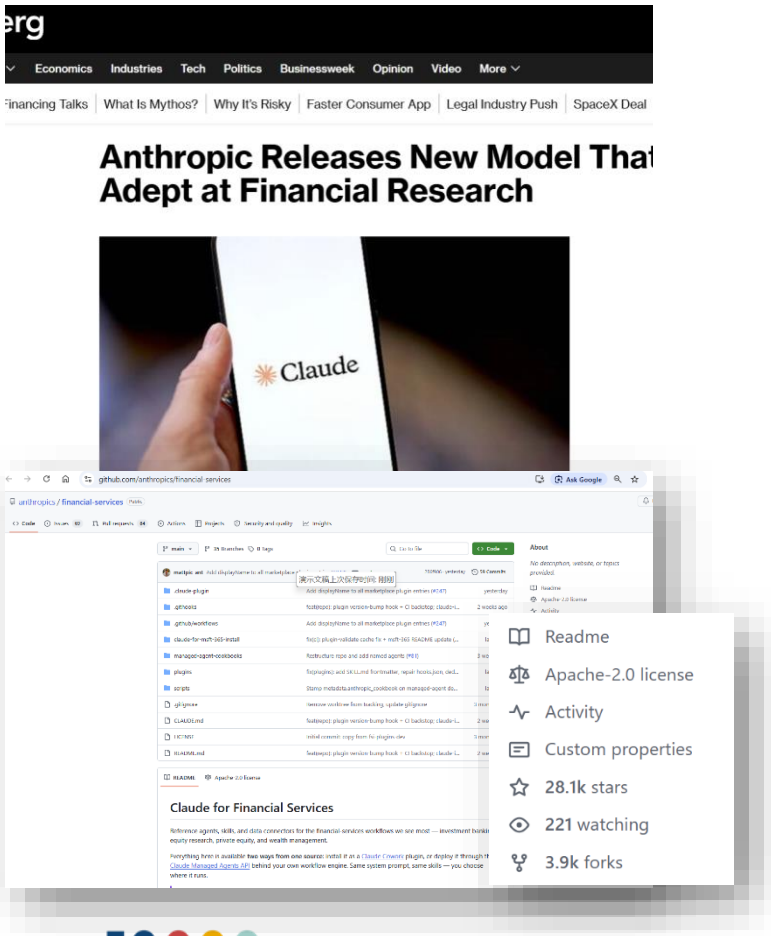
Agenda

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

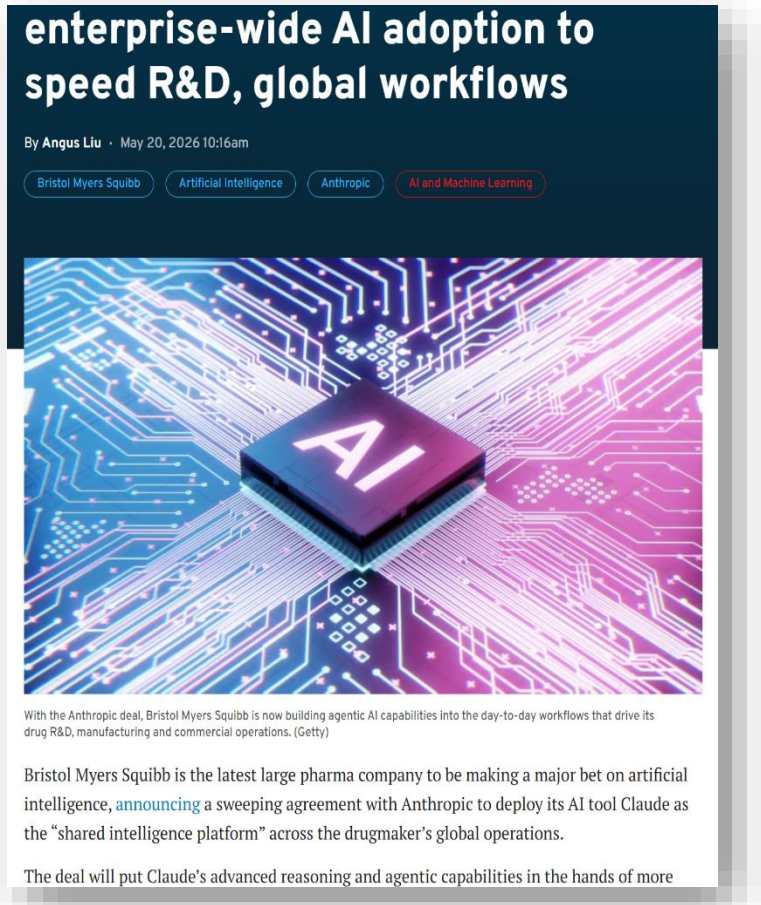
Background

Context and problem statement driving this initiative

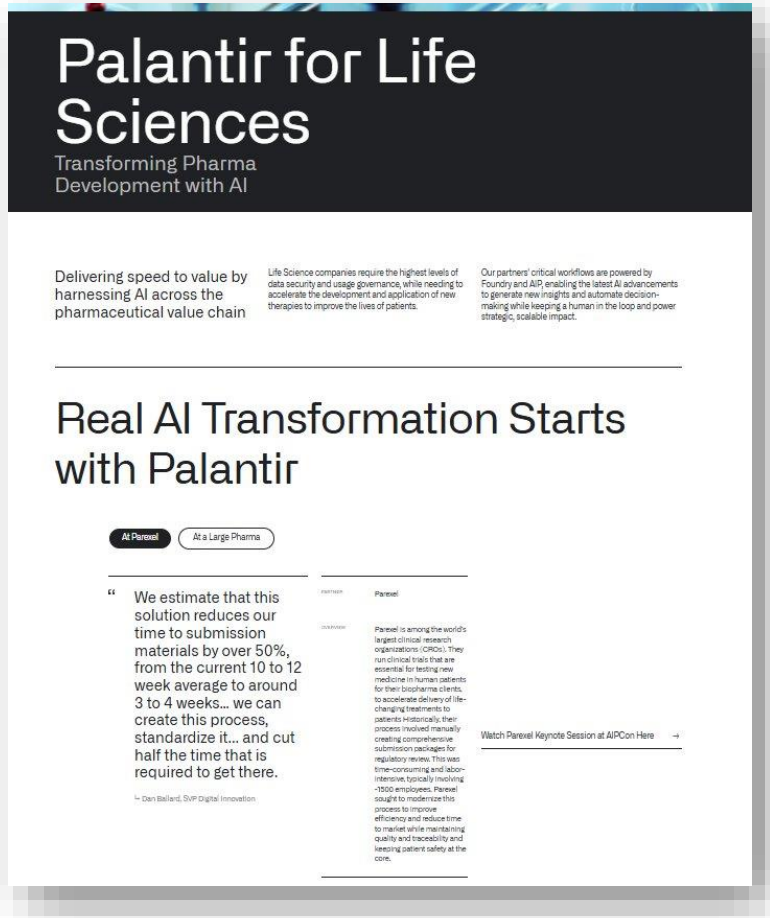
AI Adoption is Accelerating — Life Sciences Included



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



[BMS taps Anthropic's Claude for enterprise-wide AI adoption](#)



[Palantir Life Sciences Solutions](#)

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

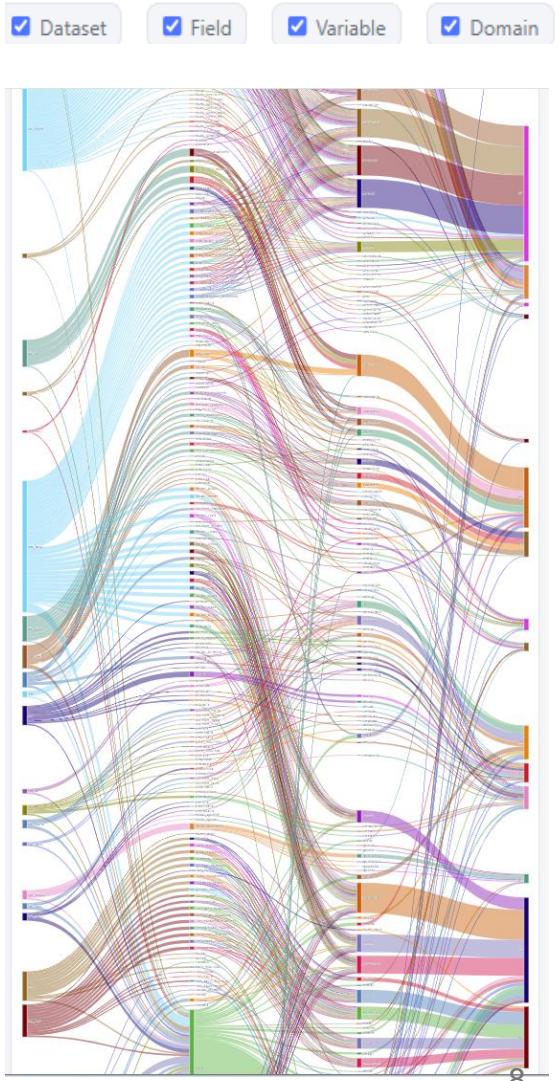
That variance in the unskilled condition matters as much as the mean. Inconsistent output is not a defensible process in a GxP context. A skill-guided agent produces consistent, traceable, {admiral} -anchored code. An unskilled agent produces something different every time.

Why we still need {admiral} in an age of AI – pharmaverse blog

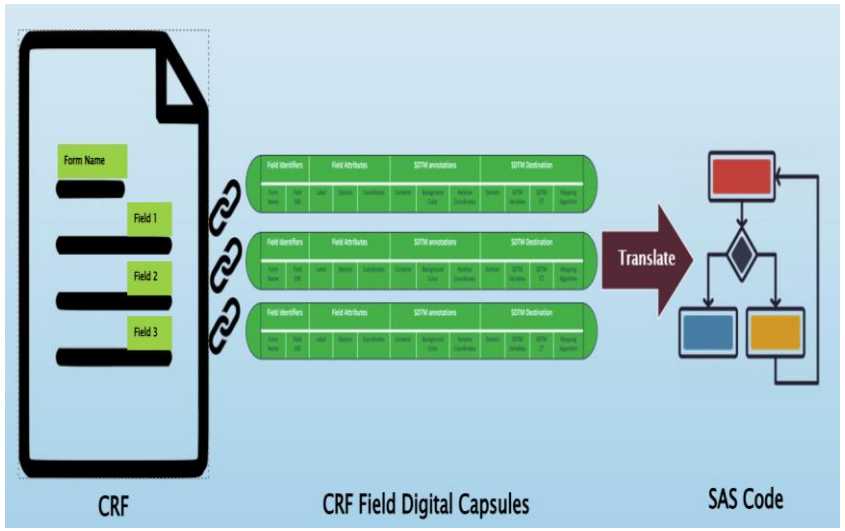
The Challenge

CDISC Mapping Complexity

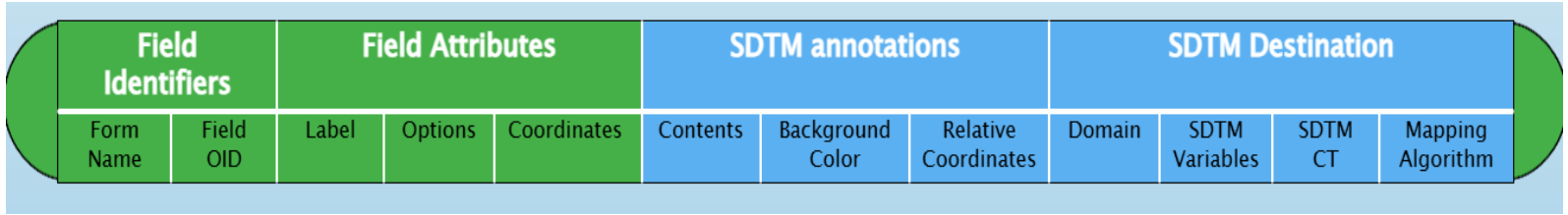
- **Structural reshaping** — many source datasets must be deconstructed and reassembled into CDISC domains + SUPPQUAL + CO.
- **Cardinality complexity** — 1:1, 1:n, n:1, transpose, and value-change transformations between EDC and SDTM.
- **Annotation fragility** — CRF page-position shifts cause inaccurate 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



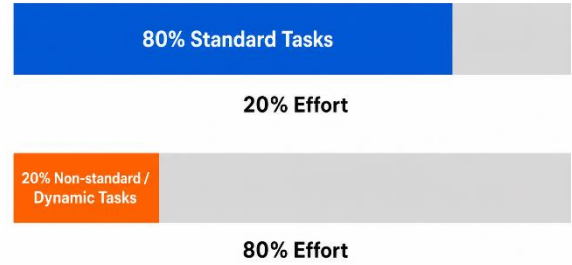
Source: Chunpeng Zhao, Mijun Hu, etc.
SDTM Annotated CRF Digitalization, PharmaSUG China 2021
Real-time SDTM, PharmaSUG China 2022

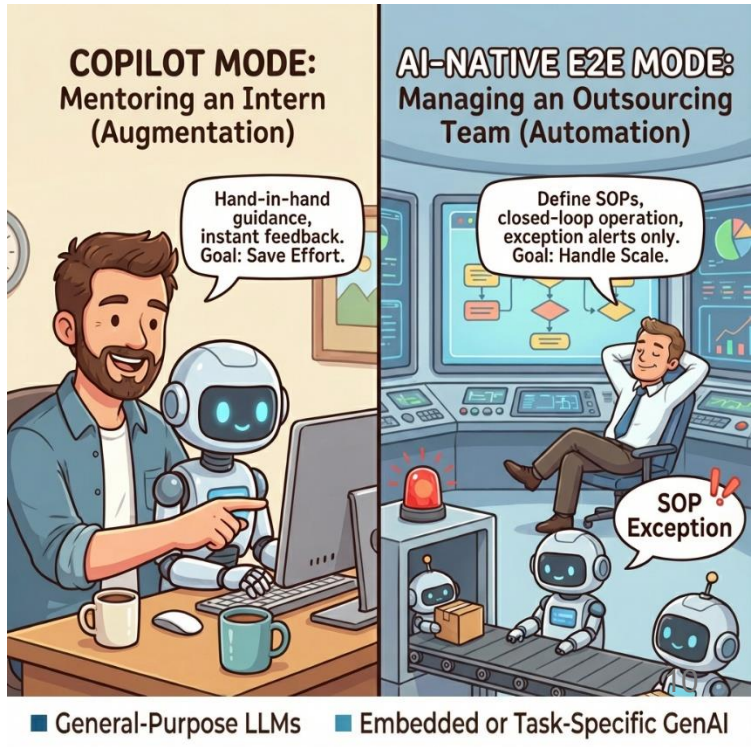


Source Dataset	FieldName	SDTM Annotation	FieldValue	Domain	Dataset	SDTM RecordID	Variable Name	VariableValue	Algorithm
DM	Form	DM = DemograpShics		DM	DM		DOMAIN	DM	Assigned
DM	BRTHDAT	BRTHDTC		DM	DM		BRTHDTC		DT2DTC
DM	AGE	AGEU = YEARS		DM	DM		AGEU	YEARS	Assigned
DM	AGE	AGE		DM	DM		AGE		CopyUpcase
DM	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		CopyUpcase

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

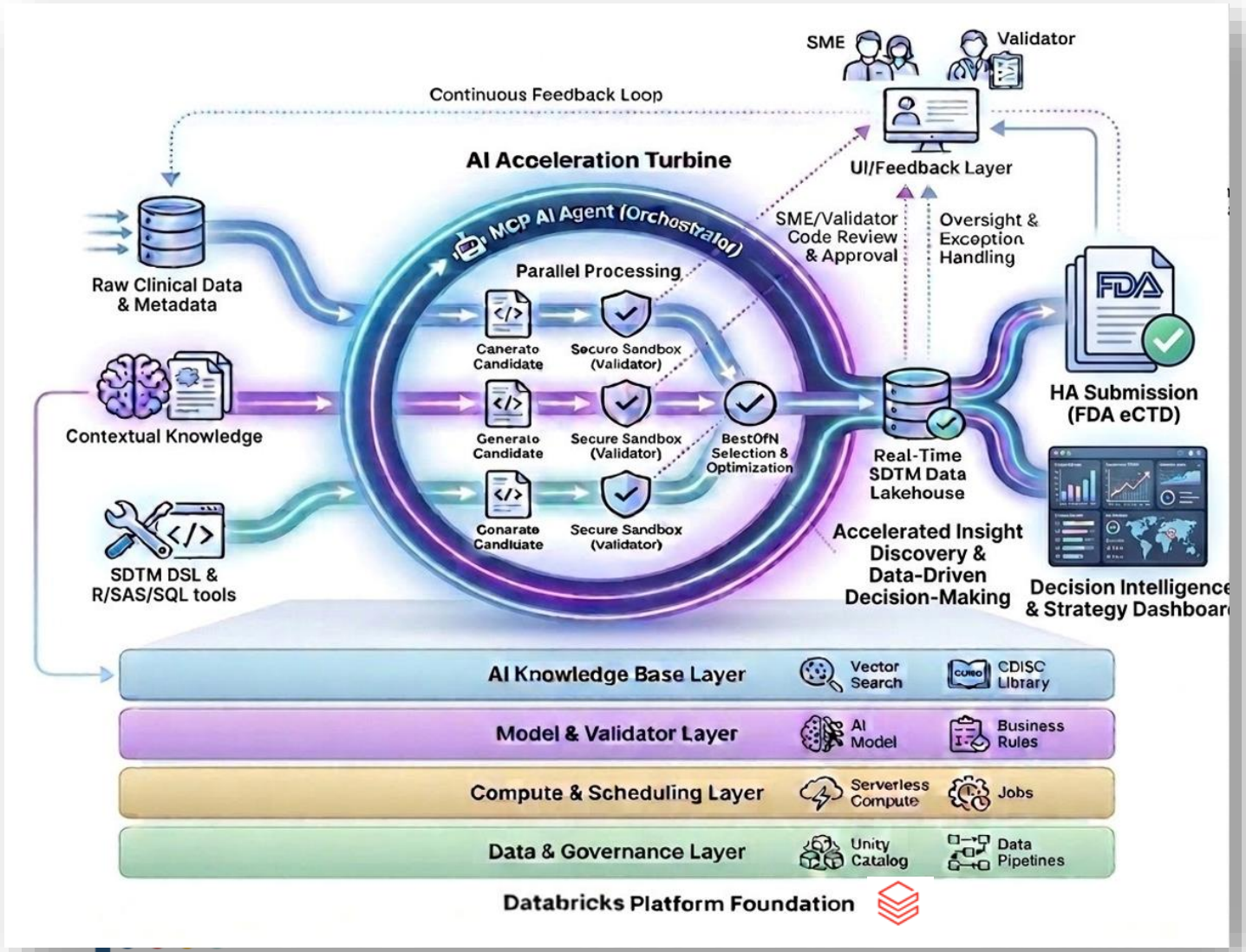




Design and Architecture

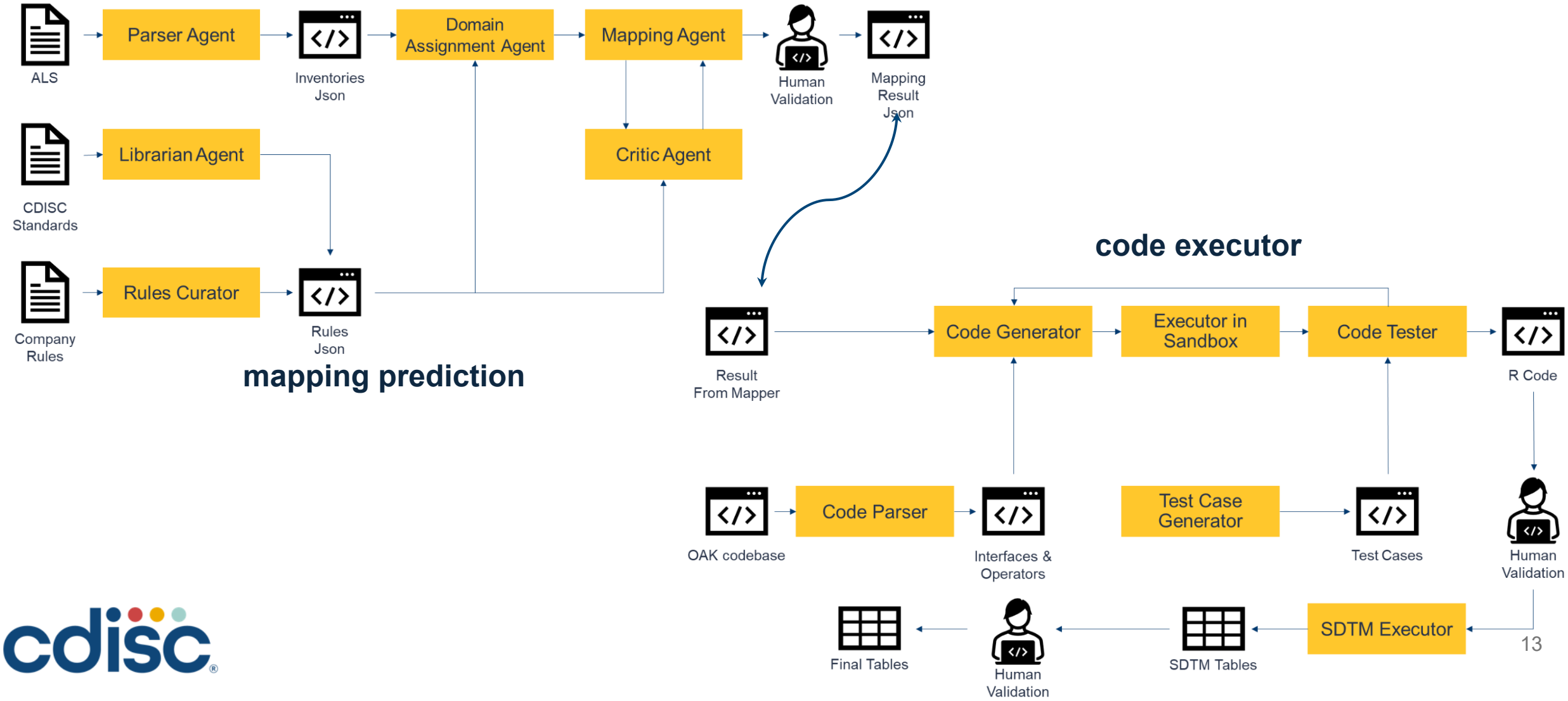
System design, key components and technical architecture

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



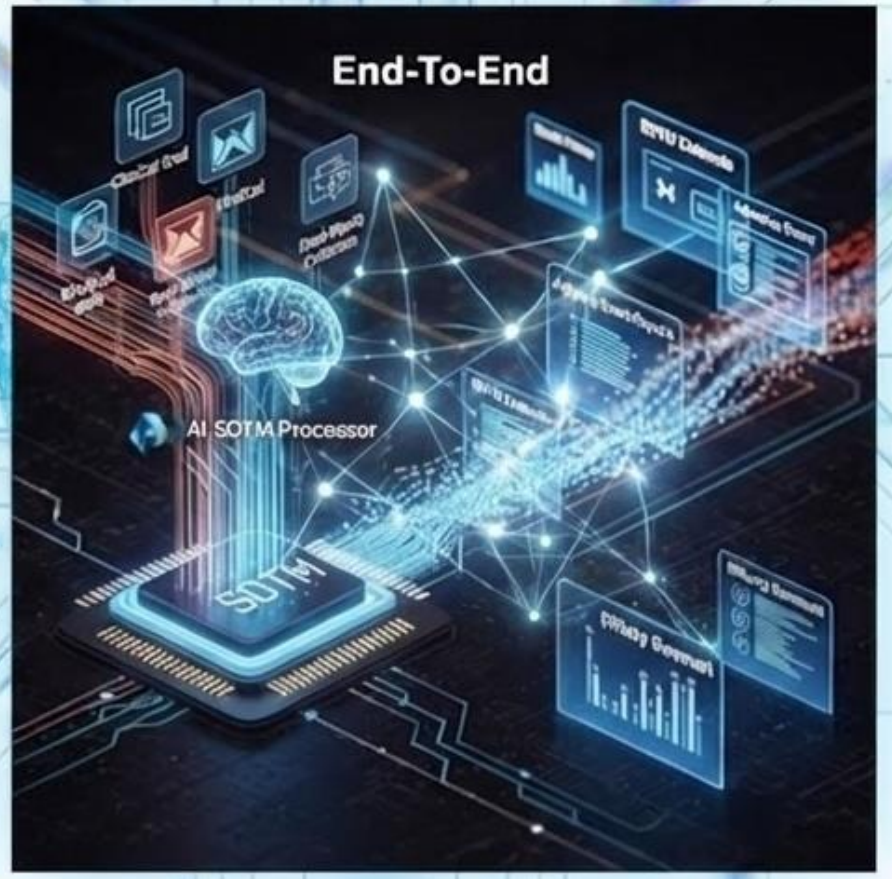
Core Capabilities Powered by Three Agent Engines

Agentic Agentic Agentic

- 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

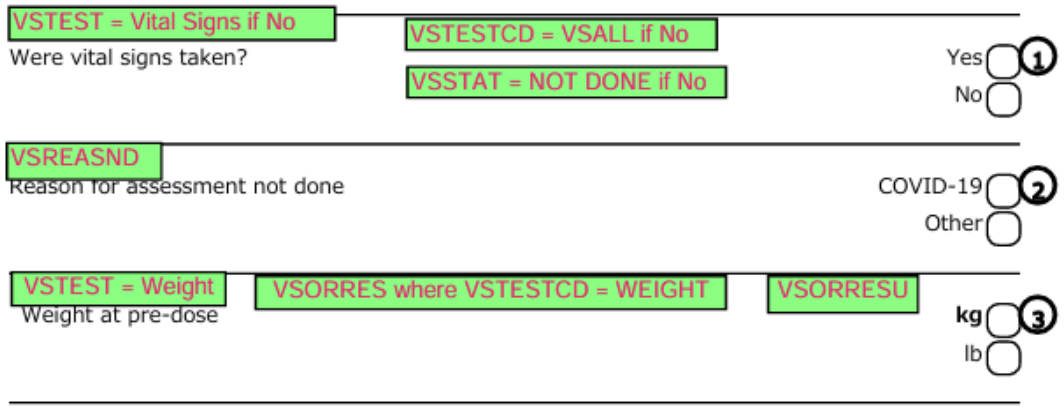
- Upload raw data/ALS/BlankCRF
- AI mapping prediction
- Mapping code generation
- Code execution and auto debug
- SDTM check
- Code update based on sdtm check
- Final output data created
- Adhoc analysis report and visualization generated based on final output data



Metrics

Key performance indicators and how success is measured

Prediction Metrics Sample

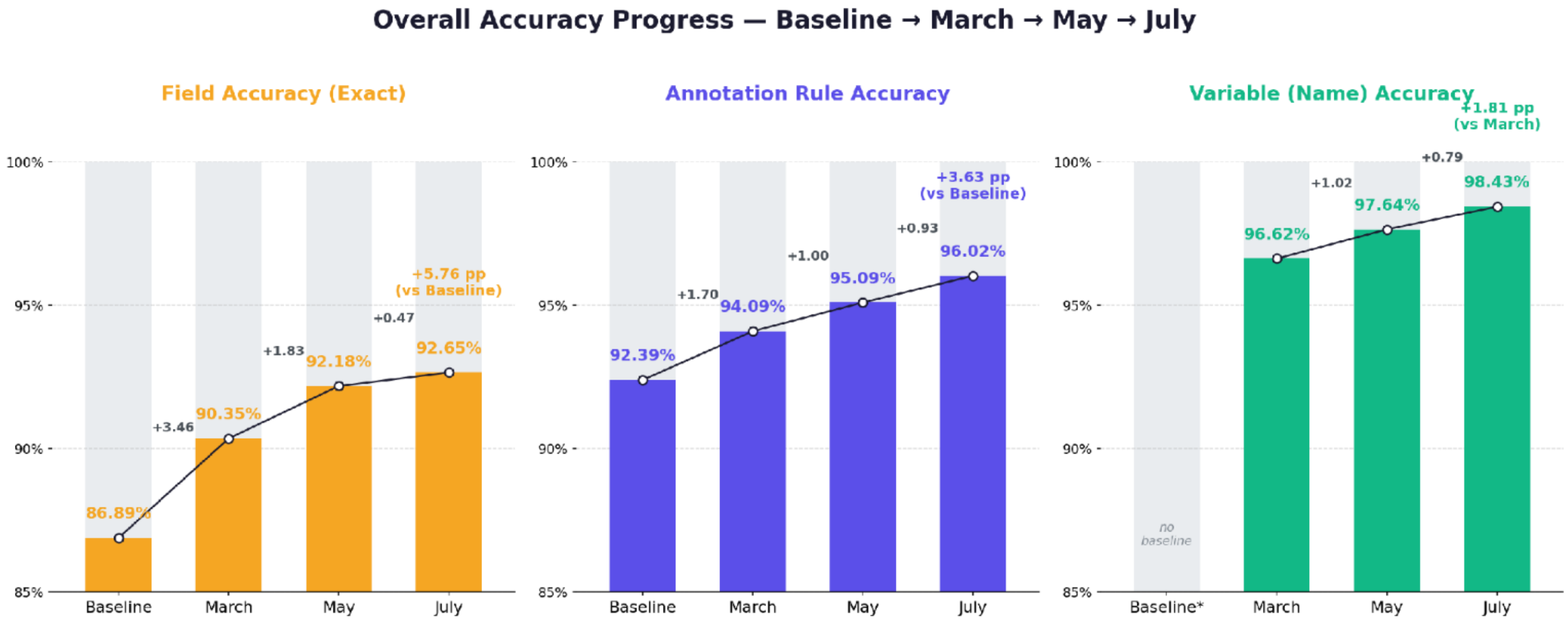


AI Prediction	Ground Truth				Is this field an exact match?
SVSTDTC	SVSTDTC	1	1	1	✓
SVOCCUR = Y; SVSTDTC	SVOCCUR = Y; SVSTDTC	2	2	2	✓
SV = Subject Visits;	SV = Subject Visits; SVPRESP = Y; SVOCCUR = Y	2	3	2	✗
DC = Demographics as Collected; DS = Disposition; DSCAT = PROTOCOL MILESTONE; EPOCH = SCREENING	DC = Demographics as Collected; DS = Disposition; DSCAT = PROTOCOL MILESTONE	4	3	3	✗
	Total:	9	9	8	

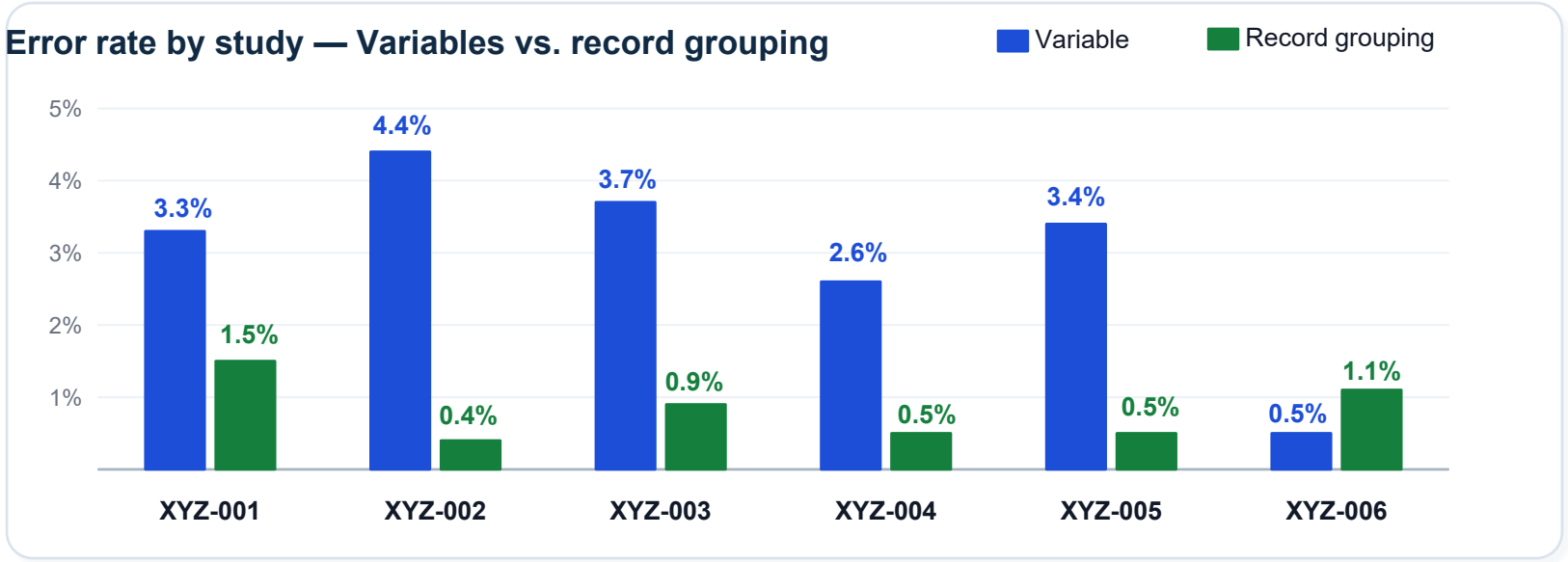
Field accuracy:
2/4=50%
(A field is counted as a match only when all annotations within it are matched.)

Annotation Rule accuracy:
8/9=89%
(Each correctly matched annotation is counted individually.)

Metrics Summary of Mapping Prediction



Metrics for Output Data Comparison



WEIGHTED AVG — VARIABLES

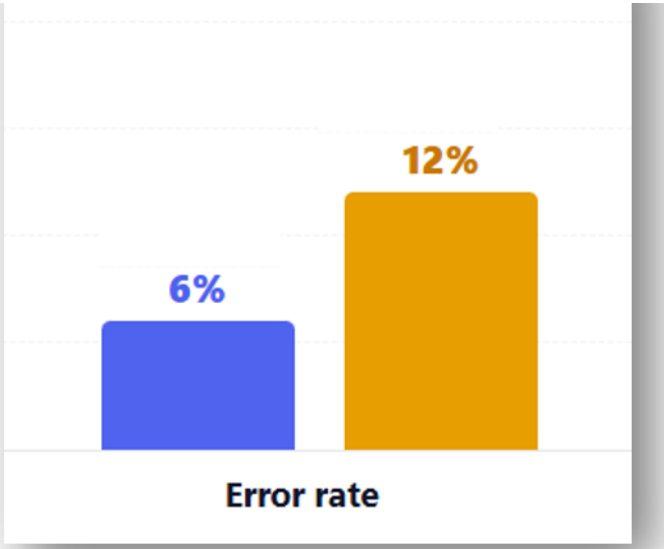
2.9%

WEIGHTED AVG — RECORD GROUPING

0.8%

Weighted avg = Σ numerators / Σ denominators

Metrics from apples to apples compare



Spec Generate

Time to produce the mapping spec (minutes) — lower is better

AI-Native	15 min	−73% vs Rule
Rule-based	55 min	baseline

Absolute delta: **−40 min** faster

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% → 89.58%, keeps 98% of quality

EXTERNAL API TOKENS

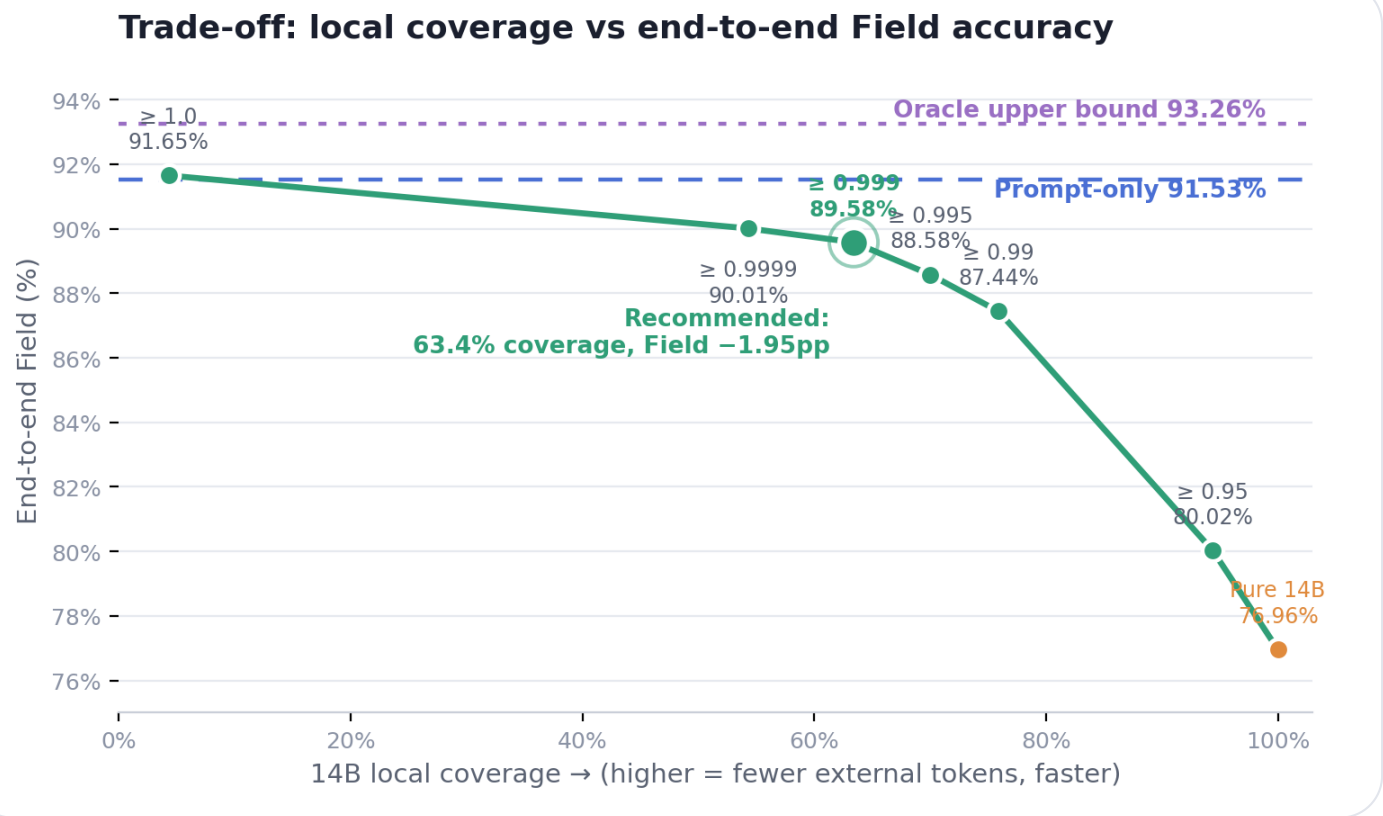
↓ 63%

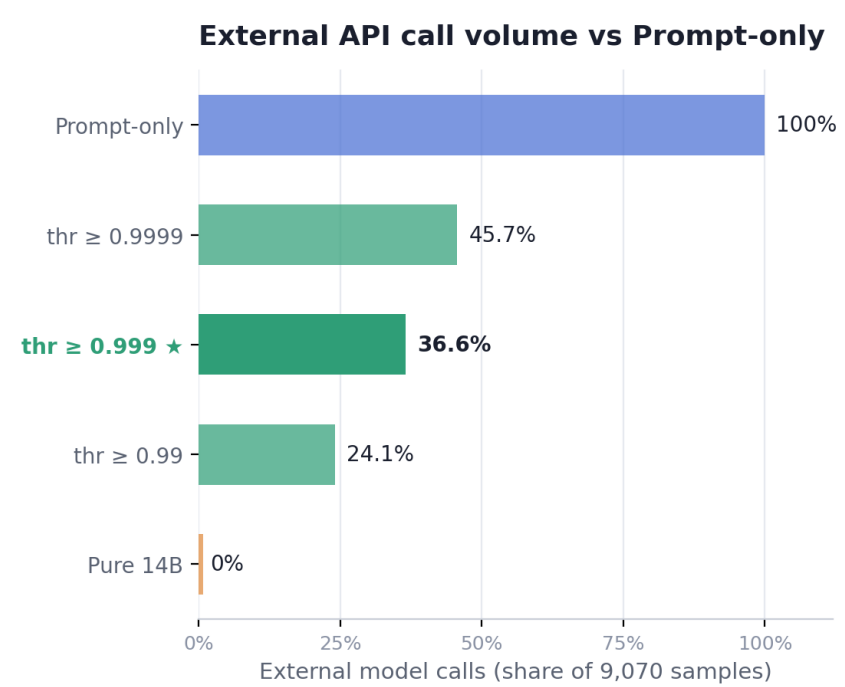
Only 36.6% of samples call Claude

14B ACCURACY ON CONFIDENT SUBSET

89.58%

Field%, vs 76.96% overall





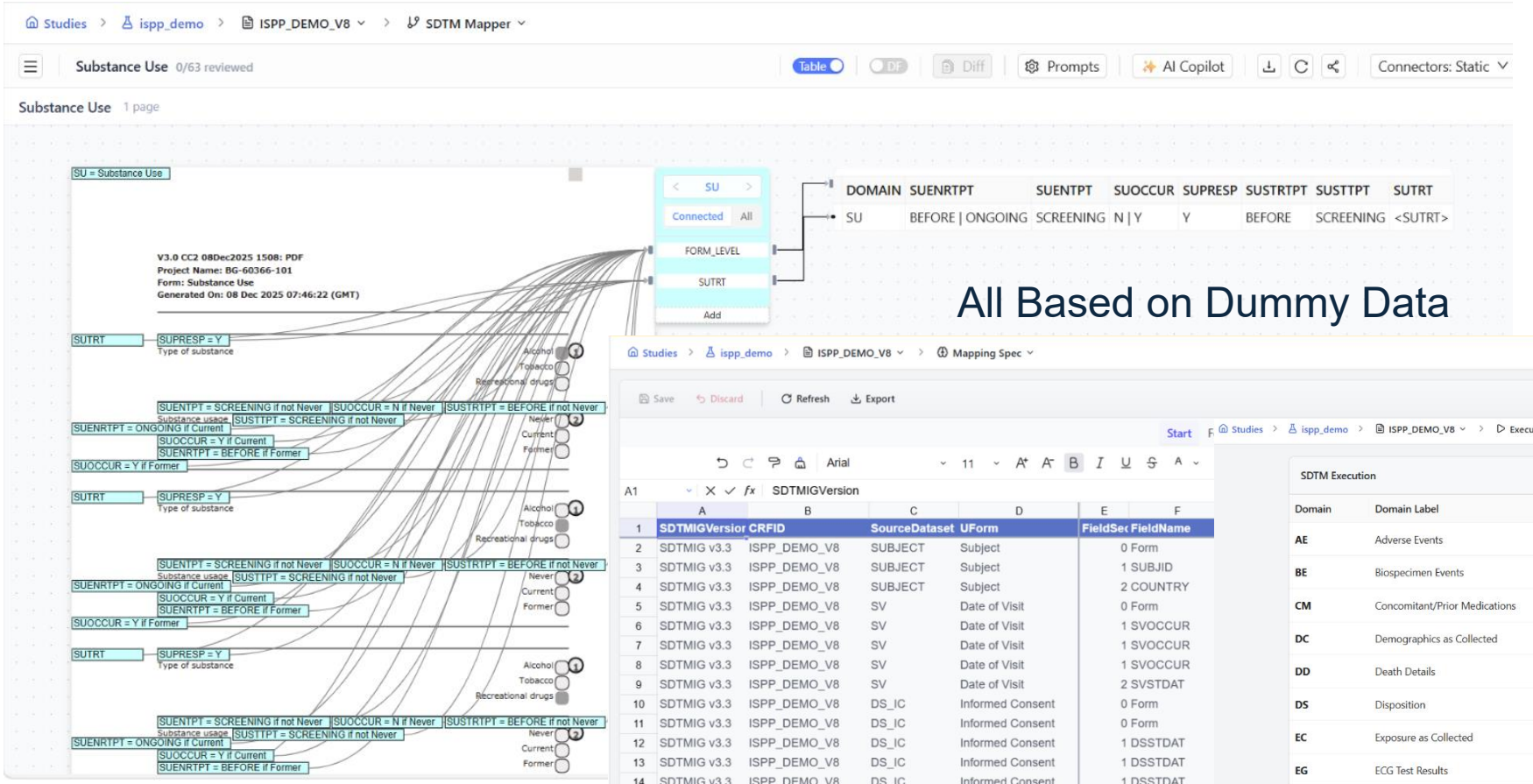
Recommended thr ≥ 0.999: 63.4% of samples run fully local (faster inference, zero external token cost) at only –1.95pp Field accuracy.

Demo

Live walkthrough of the solution in action

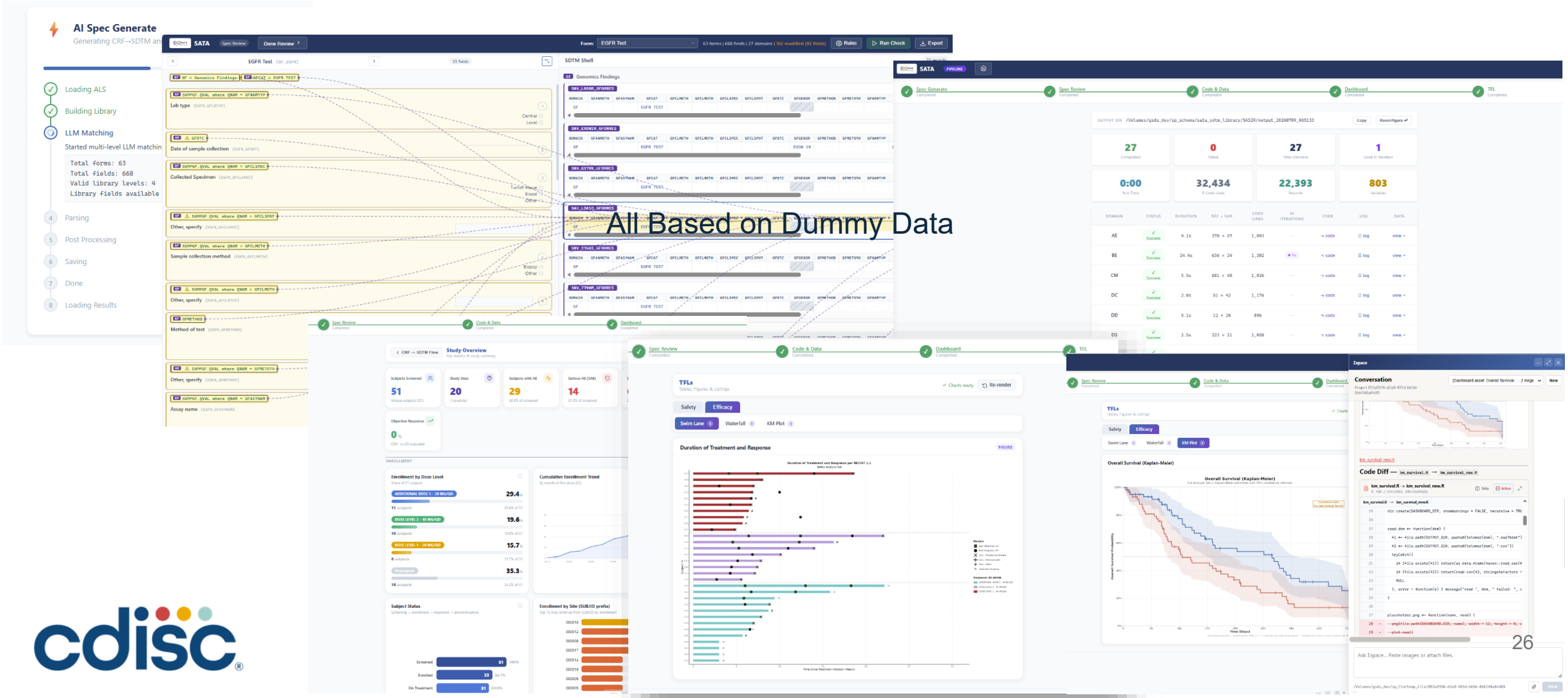
1	**Track 1 — SAS Platform**	• Solution on SAS foundation.
2	**Track 2 — E2E SDTM R Solution**	• A) UI + AI Agent : Auto Mode / Step-by-Step Mode • B) AI Agent only: Chat / Skill / API Mode

Track 1 — SAS Platform

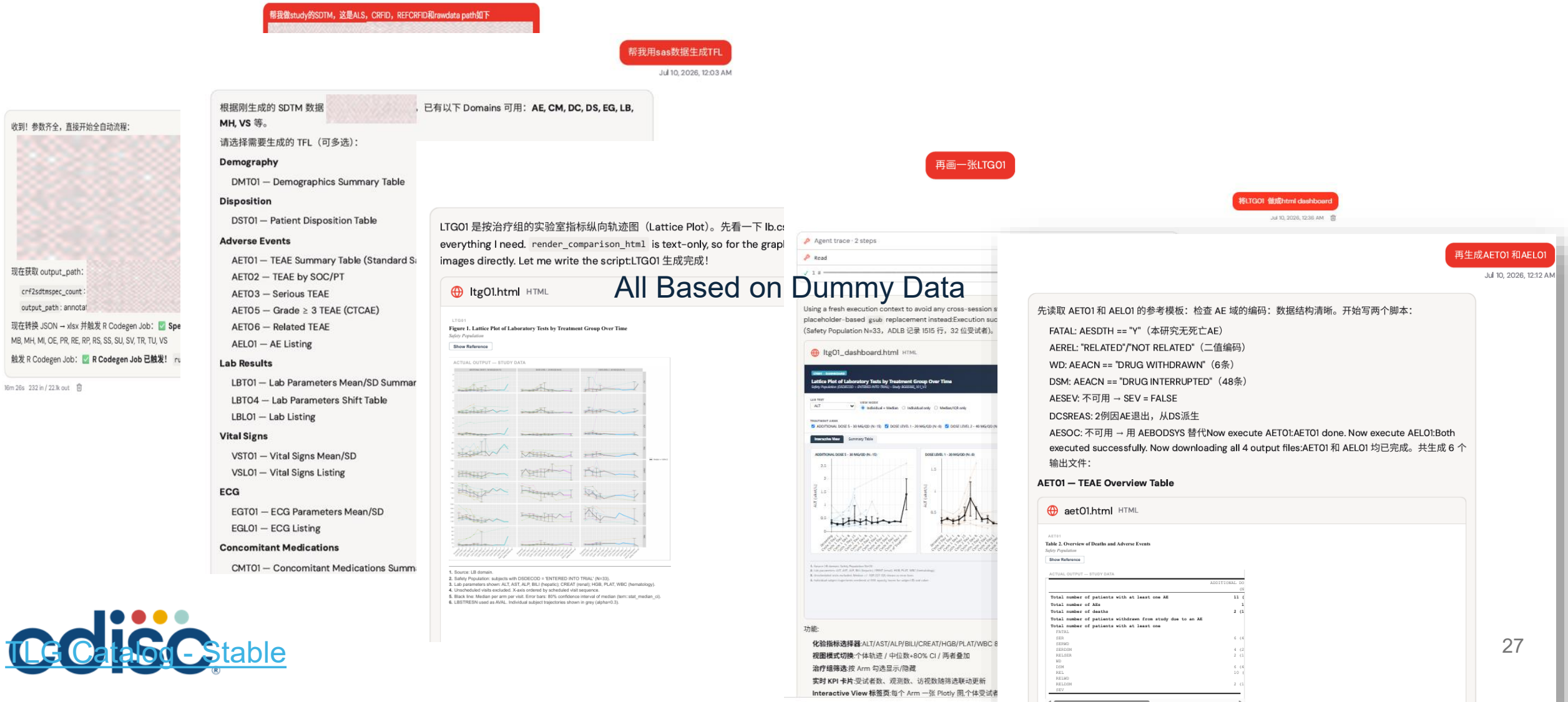


All Based on Dummy Data

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



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



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

cdisc®

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

cdisc®

Thank You

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

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

赵春鹏（Chunpeng）
上海 浦东新区

WeChat 扫二维码，添加我为朋友。

LinkedIn