

DAILY ENGLISH READING

1 September 2026 - 30-minute active reading pack

TODAY'S QUESTION

Where should the LLM stop and deterministic TFL validation begin?

Difficulty: C1 | TFL QC | LLM semantics | deterministic validation | CDISC ARS

WHY SELECTED

PROVE uses LLMs to interpret varied table wording while reserving arithmetic and logical verdicts for programmed validators. Under label variation, LLM-assisted semantic matching raised recall from 0.588 to 0.993 and F1 from 0.735 to 0.996. The lesson: probabilistic interpretation is useful; exact statistical verdicts are executable.

READING ORDER

0-3 min Preview: what does the LLM do and what does the validator own?

3-14 min PROVE: synthetic packages, injected discrepancies, label variation, recall/F1.

14-20 min CDISC ARS: machine-readable analysis results and traceability.

20-25 min ARDS + TFL: structured result datasets separated from presentation.

25-30 min Output: design one semantic + deterministic TFL QC rule.

SOURCES

Main: PROVE Multi-Agent Framework for Clinical Trial Reporting

<https://arxiv.org/abs/2607.28218>

Support: CDISC Analysis Results Standard

<https://www.cdisc.org/standards/foundational/analysis-results-standard>

Support: 2026 CDISC US Interchange Program

BRIEF BACKGROUND

Ten synthetic oncology reporting packages were generated through SDTM, ADaM and TFL.

Each discrepancy package contained 15 randomly injected errors.

Exact-label rules worked well; semantic label variation exposed brittle matching.

LLM semantic matching improved recall to 0.993 while deterministic validators kept verdicts exact.

CDISC ARS and ARDS-first workflows move statistical truth into structured result data.

Architecture: ADaM -> result data -> validation -> semantic interpretation -> rendering.

KEY VOCABULARY

deterministic validator	确定性验证器	repeatable rule-based checker
semantic matching	语义匹配	match concepts by meaning
cross-output consistency	跨输出一致性	agreement across TFL outputs
discrepancy injection	差异注入	insert known errors for testing
synthetic reporting package	合成报告包	controlled SDTM/ADaM/TFL test package
recall	召回率	fraction of true errors detected
F1 score	F 1 分数	balance of precision and recall
label variation	标签变体	different wording, same meaning
exact-match rule	精确匹配规则	requires literal text equality
evidence link	证据链接	finding linked to source evidence
analysis results dataset	分析结果数据集	structured analysis-result data
machine-readable result	机器可读结果	result software can process
rendering layer	呈现层	turn structured results into TFLs
provenance	来源与执行谱系	record of data code and versions
semantic interpretation boundary	语义解释边界	where LLM ends and exact validation begins

USEFUL PHRASES

1. use the LLM to interpret wording, not to certify arithmetic
2. reserve numerical verdicts for executable checks
3. normalize semantically equivalent table labels
4. link every finding to source evidence
5. separate analysis results from presentation
6. validate the result before rendering the display
7. treat formatting variation as a semantic problem
8. make cross-table reconciliation machine-executable
9. measure false negatives on injected discrepancies
10. design the review path around traceability and reproducibility

COMPREHENSION QUESTIONS

1. Why is semantic label variation a good use case for LLM assistance?
2. Why should arithmetic and logical validation remain deterministic?
3. What does the recall improvement reveal about exact-match validation?
4. How does CDISC ARS change the role of a rendered TFL?
5. What QC becomes possible when results exist as machine-readable data?

RETELLING PROMPTS

30 sec: LLM interpretation -> programmed validation.

45 sec: synthetic package -> injected error -> semantic match -> exact check.

60 sec: explain why TFLs may become presentation layers over result data.

5-MINUTE SPEAKING / OUTPUT TASK

Minute 1: choose percentage, denominator, SOC/PT, population, label or lab-shift QC.

Minutes 2-3: split semantic interpretation from deterministic calculation.

Minute 4: add source, rule, program, model and reviewer traceability.

Minute 5: explain why final pass/fail must remain executable.

ONE SENTENCE TO KEEP

**In regulated TFL QC, language interpretation may be probabilistic,
but the statistical verdict should be executable, reproducible, and traceable.**