

DAILY ENGLISH READING

1 August 2026 · 30-minute active reading pack for Davin

TODAY'S QUESTION

When should an LLM enter a deterministic clinical workflow?

WHY SELECTED

This paper presents a practical hybrid architecture: structured clinical logic remains deterministic, while an LLM is invoked only for selected free-text questions. The same boundary can guide AI-assisted statistical programming.

Difficulty: C1 · LLM-in-the-loop + FHIR + validation boundaries

READING ORDER · 30 MINUTES

Min	Task	Focus
0-3	Preview	Predict which steps should remain deterministic.
3-15	Main article	Read the Abstract, Architecture, Implementation, Feasibility Test and Discussion.
15-20	FDA principles	Focus on context of use, risk, standards and lifecycle management.
20-25	CDISC direction	Connect AI-assisted rules with automated validation.
25-30	Output	Design one hybrid SP workflow with a narrow LLM boundary.

SOURCES · OPEN ACCESS

MAIN · JAMIA Open · 13 Jun 2026

LLM-in-the-Loop execution of clinical quality language

[Open article](#)

SUPPORT 1 · FDA and EMA · Jan 2026

Guiding Principles of Good AI Practice in Drug Development

[Official page](#)

SUPPORT 2 · CDISC

2026 standards-based automation themes

[Official page](#)

BRIEF BACKGROUND

CQL expresses computable clinical logic over structured healthcare data. The proposed pipeline is triggered only when a rule references unstructured text.

The system maps formal logic to a natural-language query, extracts relevant facts from clinical notes and returns a binary result to the standard execution engine.

On 100 balanced test statements, Phi-3-Mini achieved 72% accuracy and an F1 score of 0.74. Llama 3.1 8B achieved 93% accuracy and an F1 score of 0.93.

The narrow interface matters: the model is not asked to generate arbitrary logic or a long narrative. It returns one constrained value to a deterministic process.

Performance depends on model, prompt and hyperparameters. Higher-risk use requires uncertainty handling, targeted validation and human escalation.

KEY VOCABULARY

Term	中文	Meaning / use
Clinical Quality Language (CQL)	临床质量语言	A domain-specific language for expressing computable clinical logic.
unstructured data	非结构化数据	Free text that does not follow a fixed tabular or coded schema.
execution engine	执行引擎	Software that interprets and runs formal logic.
LLM-in-the-Loop	大模型嵌入执行环	A design in which an LLM is called only for selected steps.
FHIR resource	FHIR 资源	A standardized unit of healthcare information exchange.
DocumentReference	文档引用资源	A FHIR resource pointing to clinical documents or notes.
binary classification	二元分类	A decision between two outcomes, such as true or false.
query-relevant facts	与查询相关的事实	Evidence extracted specifically for one question.
backward compatibility	向后兼容性	The ability to preserve existing tools and workflows.
model-agnostic	模型无关的	Designed to work with different models.
false positive	假阳性	A positive prediction when the condition is actually absent.
false negative	假阴性	A negative prediction when the condition is actually present.
runtime configuration	运行时配置	Settings supplied when the system executes.
human escalation	人工升级处理	Routing uncertain or high-risk cases to qualified reviewers.
context of use	使用情境	The specific purpose and conditions under which a model is used.

USEFUL PHRASES

1. invoke the LLM only when needed

The engine invokes the LLM only when unstructured text is referenced.

2. preserve the existing deterministic pathway

Structured data preserves the existing deterministic pathway.

3. translate formal logic into a natural-language query

The pipeline translates formal logic into a natural-language query.

4. extract query-relevant facts

The model first extracts query-relevant facts from the note.

5. return a constrained binary response

The LLM returns a constrained binary response to the execution engine.

6. reduce the surface area for hallucination

Narrow task scope can reduce the surface area for hallucination.

7. remain sensitive to prompt and model choice

Performance remains sensitive to prompt and model choice.

8. require domain-targeted evaluation

Each use case requires domain-targeted evaluation.

9. escalate uncertain outputs for human review

The system should escalate uncertain outputs for human review.

10. fit the declared context of use

Validation evidence must fit the declared context of use.

COMPREHENSION QUESTIONS

1. Which data continue through normal CQL execution?
2. What steps connect an unstructured note to a binary result?
3. How did the two local models differ?
4. Why can a narrow binary task reduce hallucination risk?
5. What additional controls are needed for high-risk use?

RETELLING PROMPTS

30 seconds: Problem -> boundary -> result.

45 seconds: Formal rule -> natural-language query -> facts -> true/false -> execution.

60 seconds: Apply the pattern to ADaM or TFL work.

5-MINUTE SPEAKING / OUTPUT TASK

Your role: Propose one bounded LLM step inside a deterministic programming system.

Minute 1: Choose one narrow task.

Minutes 2-3: Define the deterministic pathway, trigger, evidence and output schema.

Minute 4: Add gold-standard testing, uncertainty handling, regression tests and escalation.

Minute 5: State acceptance criteria and mandatory human-review cases.

ONE SENTENCE TO KEEP

The safest role for an LLM in a regulated workflow is often to resolve one narrow unstructured-data problem and return a constrained, testable result.