

# LLMs bring flexibility to the OMOP ecosystem, but remove the guardrails



## WHAT LLMs ADD

### Access without a technical background

Clinical researchers can interrogate an OMOP database directly, without SQL or R expertise.

### Beyond the OHDSI tools

Combine the existing OHDSI stack (ATLAS, HADES R, DARWIN EU R) and gain the flexibility to query the database directly for questions the tools were never built to answer.

## WHAT LLMs REMOVE

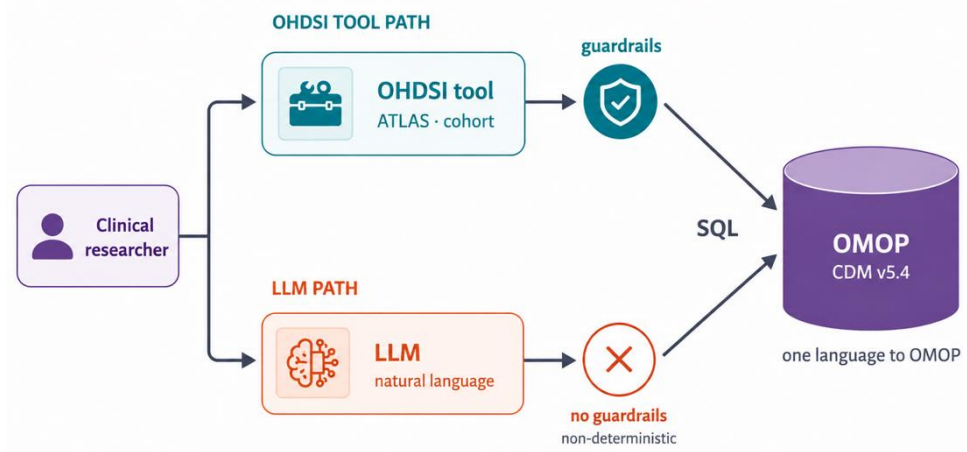
### Engineered tools had built-in guardrails

OHDSI tools also turn user intent into SQL, but expert-engineered. We know in advance how they behave, so guardrails are baked in.

### LLMs remove that guarantee

LLMs are non-deterministic. The expert-designed behaviour disappears, and the implicit guardrails go with it.

The flexibility is real; so is the loss of guardrails



## Semantic Risk

LLM-generated SQL can misinterpret OMOP semantics, leading to incorrect results and hidden patient safety risks.

# Example: A Simple Clinician Question



Clinician asks

Database with 27 patients

“Find patients with type 2 diabetes on metformin.”



LLM generates

```
SELECT c.person_id
FROM condition_occurrence c
JOIN drug_exposure d
  ON c.person_id = d.person_id
WHERE c.condition_source_concept_id = 201826
AND c.condition_start_date >= '2015-01-01';
```



LLM response

“We have 13 patients with diabetes, of which 11 are currently on metformin.”

Answer looks plausible and ready for analysis

```
SELECT c.person_id
FROM condition_occurrence c
JOIN drug_exposure d
  ON c.person_id = d.person_id
WHERE c.condition_source_concept_id = 201826
AND c.condition_start_date >= '2015-01-01';
```

**Wrong vocabulary semantics**

Uses source\_concept\_id instead of standard concept mapping

**Missing observation-period anchor**

Date filter is not anchored to the person's observation period

The SQL is syntactically valid and executes without error, silently returning incorrect results

# FastSSV: a semantic guardrail for LLM-generated OMOP SQL

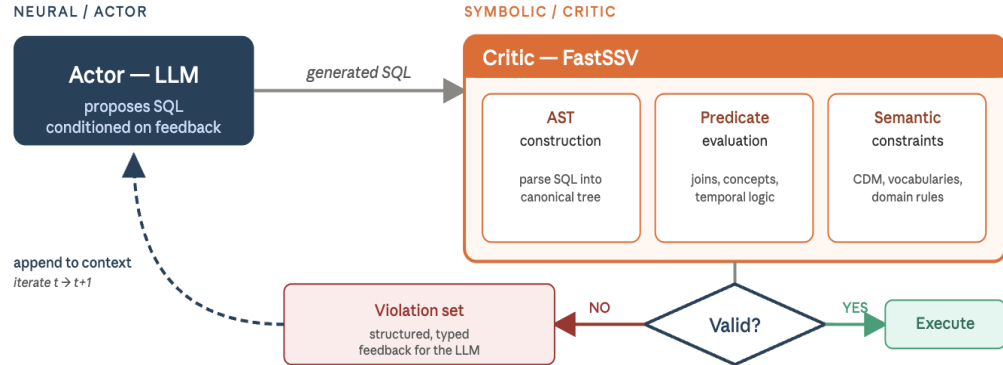


Rules derived from OHDSI SQL gold standards on OMOP CDM v5.4



FastSSV

155 rules



- Whether SQL comes from an LLM or from an OHDSI tool, the language spoken to OMOP CDM is the same: SQL. So that is where we put the guardrail.
- FastSSV ensures that, even when LLM outputs vary, they remain guided by the same analytical constraints and converge towards consistent, reproducible results.

# Team members



Vishnu Chandrabalan



Richard Dobson



Shihao Shenzhang



Zhangshu Joshua Jiang



Niko Moeller-Grell

---

[Live Demo](#)

## Acknowledgements:

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) [EP/Y035216/1] Centre for Doctoral Training in Data-Driven Health (DRIVE-Health) at King's College London, with additional support from the National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre (BRC) [NIHR203318] and Lancashire Teaching Hospitals NHS Foundation Trust. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care, or Lancashire Teaching Hospitals NHS Foundation Trust.