

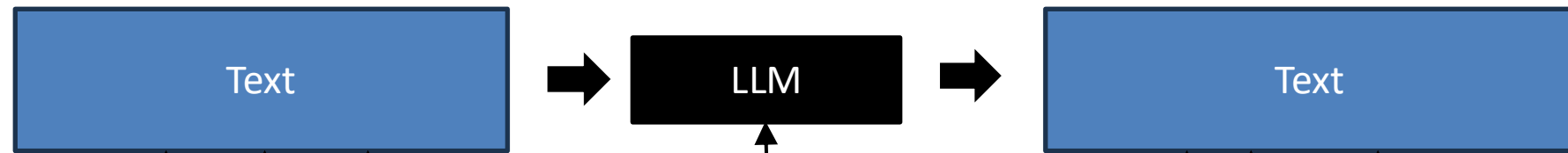


# Generative AI and Foundational Models Workgroup

---



# Large language models in 2026



## Retrieval-augmented generation (RAG)

Add relevant information (e.g. search results) to the prompt.

## Skills

Add re-usable instructions to the prompt.

## Leaner & meaner LLMs

LLMs are getting better and run on smaller hardware. But they are static. Nobody changes the weights, and every time you call them, they start from the same point.

## Tool use (MCP)

Conventions on how to parse LLM output for function calls and execute them. Requires documentation of available tools in LLM input prompt.

## Reasoning

LLMs have been fine-tuned to first generate 'thoughts' in a separately-marked output section.

## Structured output

LLMs can be forced to generate valid json, improving application reliability.



# Types of AI applications in OHDSI

## 1. LLM in a workflow

- Mostly traditional code, preparing LLM input and processing LLM output
- Keeper: In: patient profile → Out: case adjudication
- Ariadne: In: Source term + candidate target concepts → Out: Exact match or no match

## 2. LLM as an agent

- LLM can use tools
- Depending on results of tools, LLM can use more tools
- (Implied: LLM is called in a loop)
- Agentic coding: Prompt → Modify code → Run test → Fix code





# Generative AI use cases in OHDSI

## LLM in a workflow

- NLP (e.g. convert clinical notes to concepts)
- Vocabulary mapping
- Phenotype evaluation

## LLM as an agent

- Phenotype design
- Study design

## Foundational models

- Patient-level prediction
- Data synthesis



# Challenges in generative AI for OHDSI

- Hallucinations
  - Not really, just need good grounding
- Preserve patient privacy
  - Need local models when processing patient data
- Achieve synergy
  - Most OHDSI collaborators working in silos
- Avoid RWE slop
  - As shown through OMOP and OHDSI methods research, most published observation studies are of poor quality
  - GenAI allows 10x bad practices



# OHDSI Best Practices for Observation Research

- Validate your assumptions through **objective diagnostics**
  - Negative controls!
  - Validate your phenotype algorithm
- Don't assume you (or the LLM) know the full **causal structure**
  - Test covariate balance on *all* covariates
- Use a **diverse network of databases**
  - Evidence from 1 database is no evidence at all

Shoutout to Rich Boyce, who is developing the StudyAgent around existing OHDSI tools, fully following these best practices



# Opportunities

- LLMs are great at following instructions
  - Can follow best practices (better than humans ;-))
- Foundational models could lead to completely new analytics
  - Causal transformers?
- Collaborate!



# Collaboration opportunities

- Validation (avoid slop)
  - Minds Meet Machines challenges
    - Concept set creation
    - Procedure code mapping
  - Gold standards
    - Exact condition concept mapping
- Skills
  - Currently debating in the Workgroup
- Tools
  - For both workflow (REST API) and agent (MCP)
    - Hecate Vocabulary vector search (validated using gold standard)
  - Also currently being debated in the Workgroup