Generative AI for real-world evidence

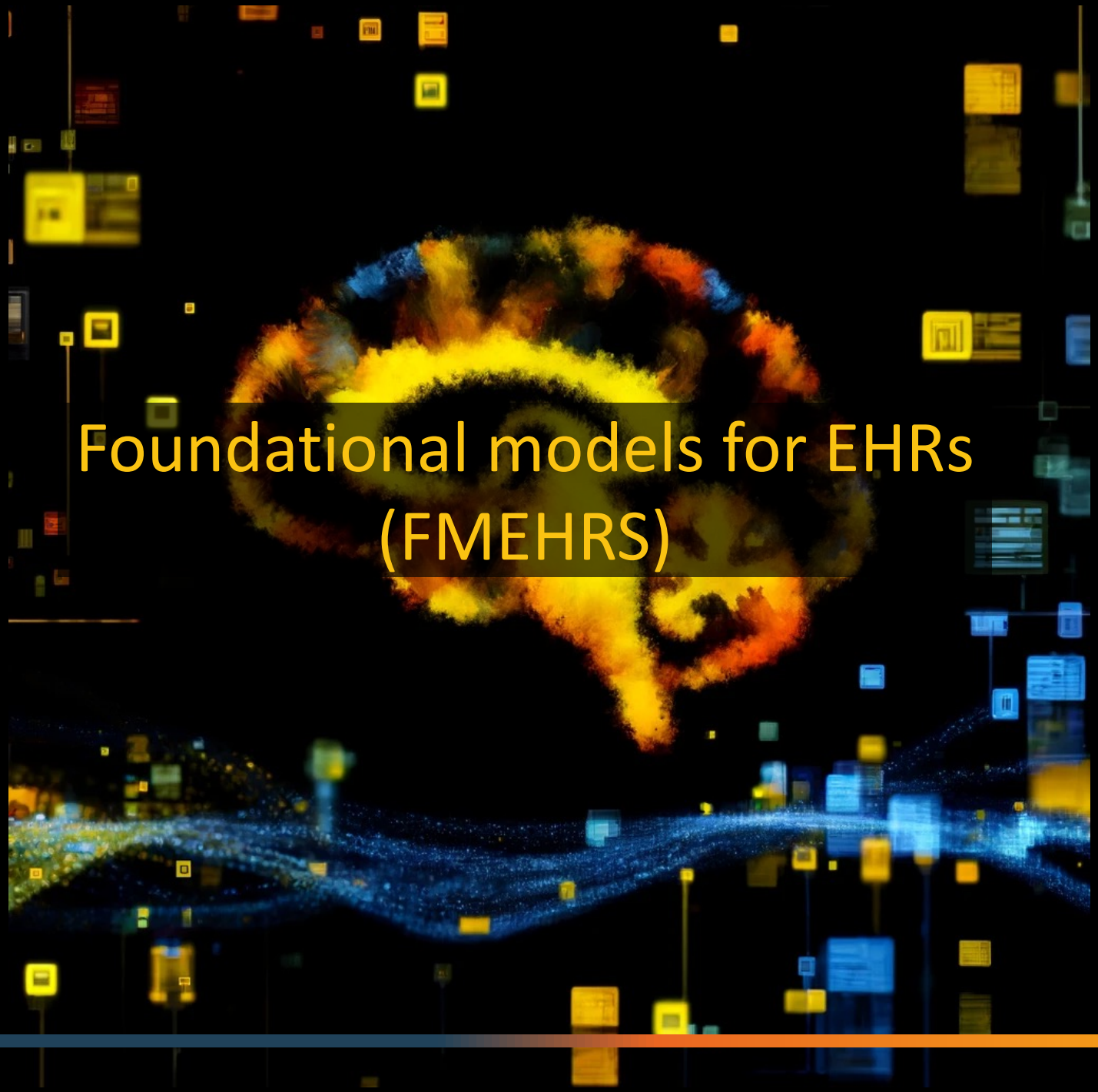
Martijn Schuemie

OHDSI
Observational Health Data Sciences and Informatics

Generative AI and Foundational Models Workgroup

- Monthly meetings
  - Second Tuesday of the month
  - Noon Eastern Time
- OHDSI Teams channel

OHDSI Generative AI &
Foundational Models Workgroup

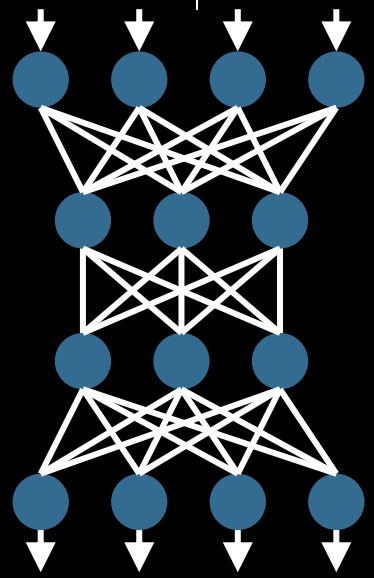


Foundational models for EHRs (FMEHRS)

OHDSI
Observational Health Data Sciences and Informatics

LLM pretraining: predict the next word

… said the Cat. “I don’t much care where—” said ?



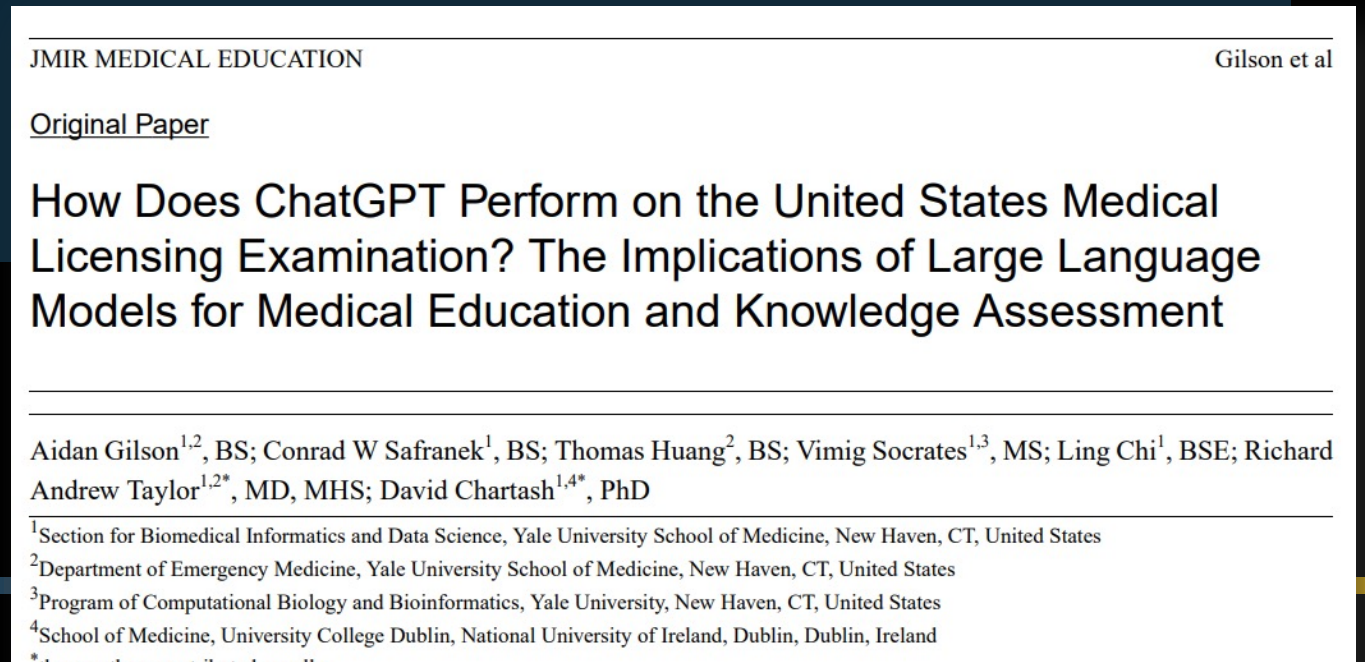
Predicting the next word requires:

- Grammar: next word probably is a (proper) noun
- Semantics: only some things can talk
- Context: this is conversation between the Cheshire Cat and Alice

Most likely next word: Alice

Fine-tuning world's most expensive auto-completion

- Pre-trained models can be used to predict the next word, and the next, and the next, generating text
- Can be further training to generate answers to questions (chat)
  - Supervised: Human-created training set
  - Reinforcement learning: human corrects output of LLM
- Initial work shows model learned important concepts in medicine



JMIR MEDICAL EDUCATION Gilson et al

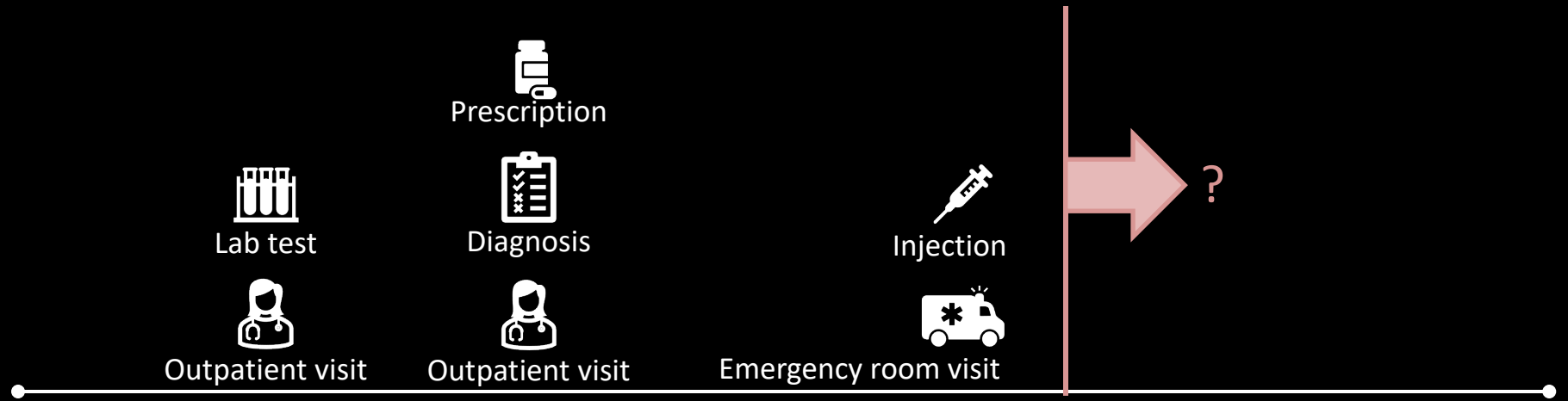
Original Paper

How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment

Aidan Gilson[1,2], BS; Conrad W Safranek[1], BS; Thomas Huang[2], BS; Vimig Socrates[1,3], MS; Ling Chi[1], BSE; Richard Andrew Taylor[1,2*], MD, MHS; David Chartash[1,4*], PhD

[1]Section for Biomedical Informatics and Data Science, Yale University School of Medicine, New Haven, CT, United States
[2]Department of Emergency Medicine, Yale University School of Medicine, New Haven, CT, United States
[3]Program of Computational Biology and Bioinformatics, Yale University, New Haven, CT, United States
[4]School of Medicine, University College Dublin, National University of Ireland, Dublin, Dublin, Ireland

FMEHR pretraining: predict the next or masked event

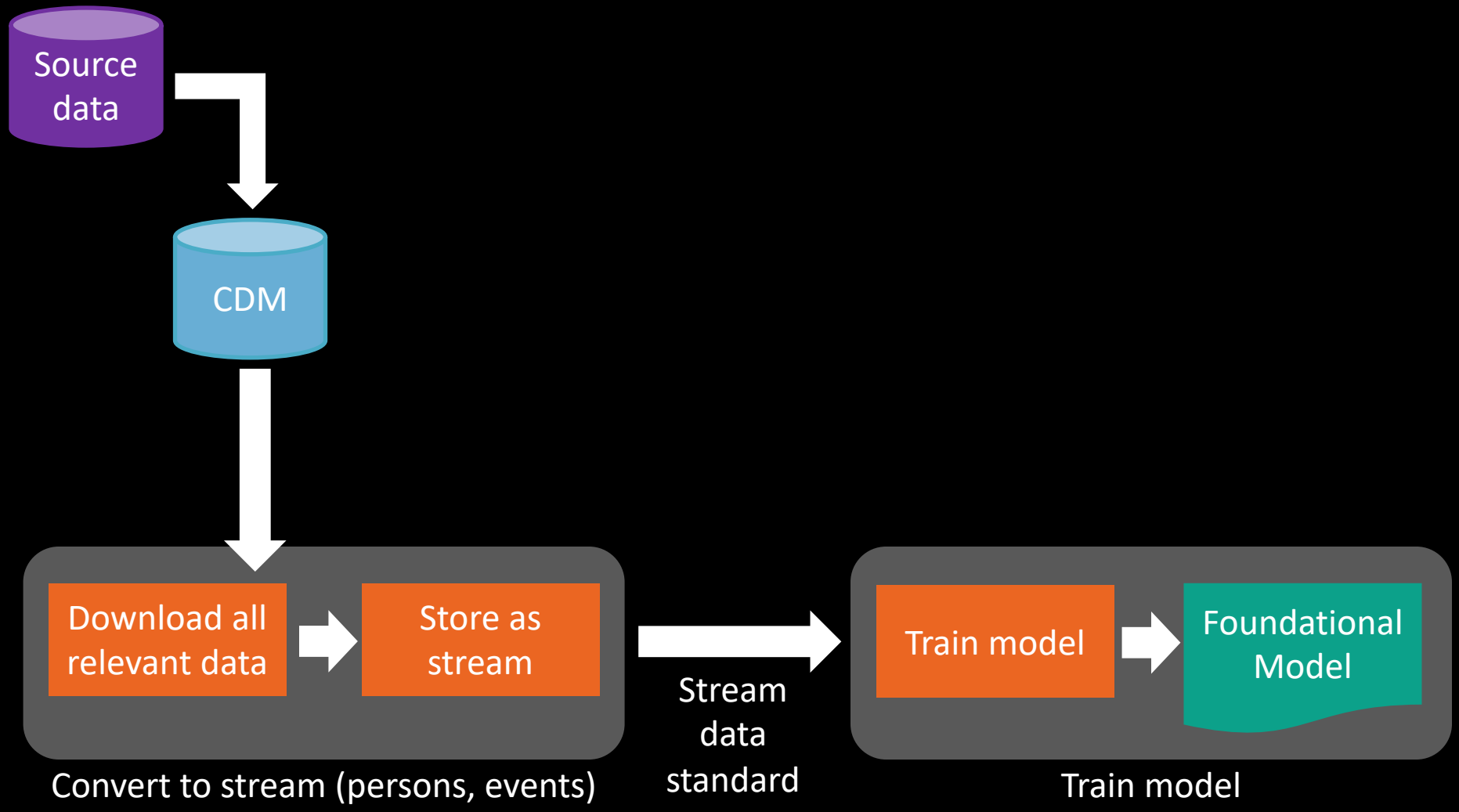


- Trained on all data in a database
  - unlike current PLP, which only uses target cohort
- Not a language model (healthcare model?)
- Pretrained model will contain deep knowledge
- Use cases:
  - Better prediction models
  - Counterfactual prediction
  - Clustering
  - …?

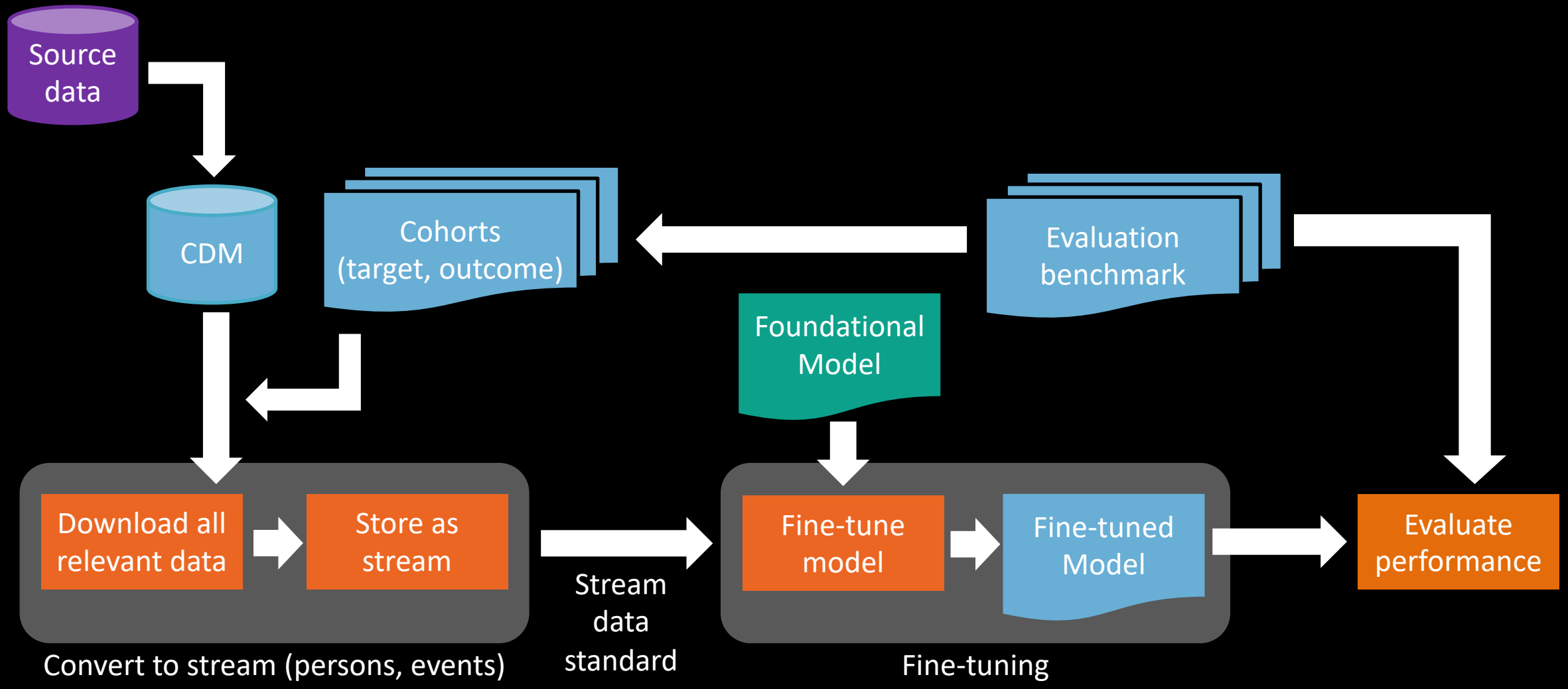
FMEHRs vs regular LLMs

- Regular LLM (e.g. GPT-4o):
  - Natural language model
  - Trained on all text on the internet
  - Pretraining task: predict next word
  - Fine-tuning task: chat, …
- FMEHRs:
  - ‘Healthcare model’?
  - Trained on all data in a CDM database
  - Pretraining task: predict next clinical ‘event’
  - Fine-tuning task: patient-level prediction, …

FMEHR pipeline: building a foundational model



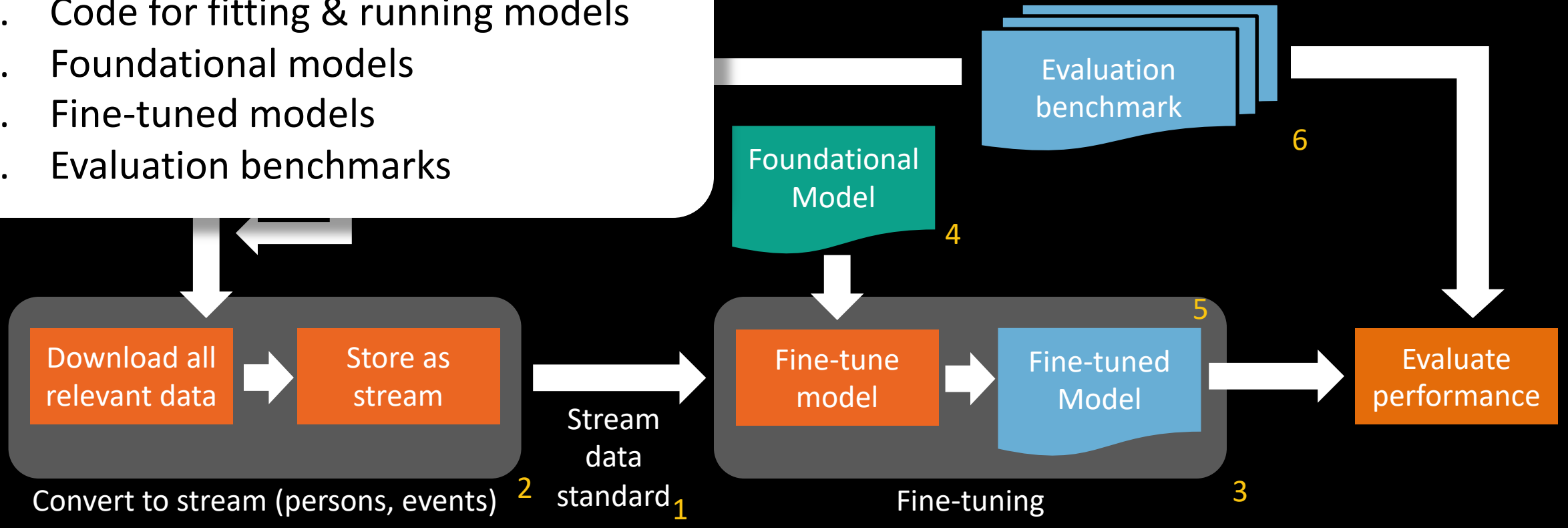
FMEHR pipeline: fine-tuning for prediction



FMEHR pipeline: fine-tuning for prediction

Shareable:

1. Stream data standard (MEDS)
2. Code for converting CDM to stream
3. Code for fitting & running models
4. Foundational models
5. Fine-tuned models
6. Evaluation benchmarks



Stanford's EHRShot

- https://som-shahlab.github.io/ehrshot-website/
- Dataset + benchmark
  - 15 clinical prediction tasks
- Foundational model

Current status

- Started OHDSI Assessment of Pre-trained Observational Large Longitudinal models in OHDSI (APOLLO) project
- Implemented several pretraining algorithms
  - Can now fit them on our data (needs big GPUs!)
- Integration with the PatientLevelPrediction package for fine-tuning & prediction
- Next steps:
  - Evaluate against OHDSI PLP benchmark
  - Support MEDS standard

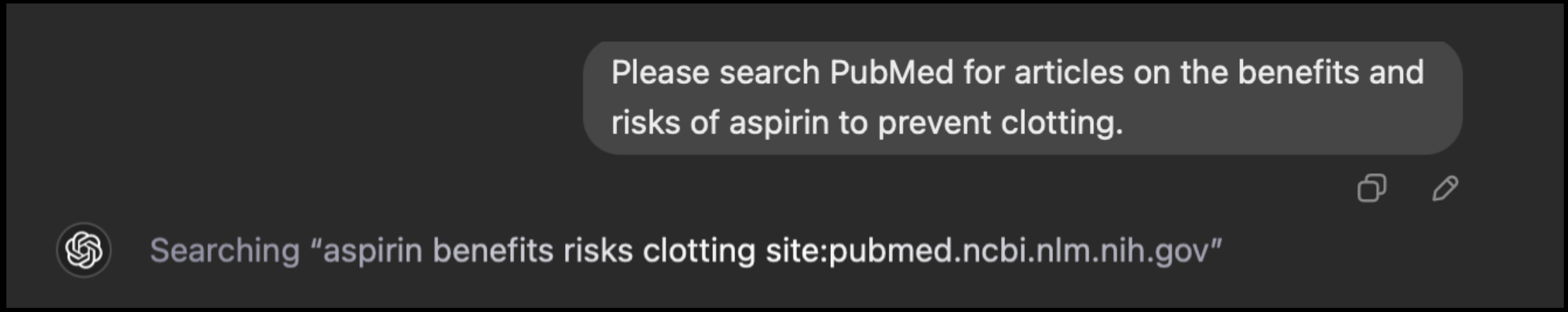
OHDSI

Observational Health Data Sciences and Informatics

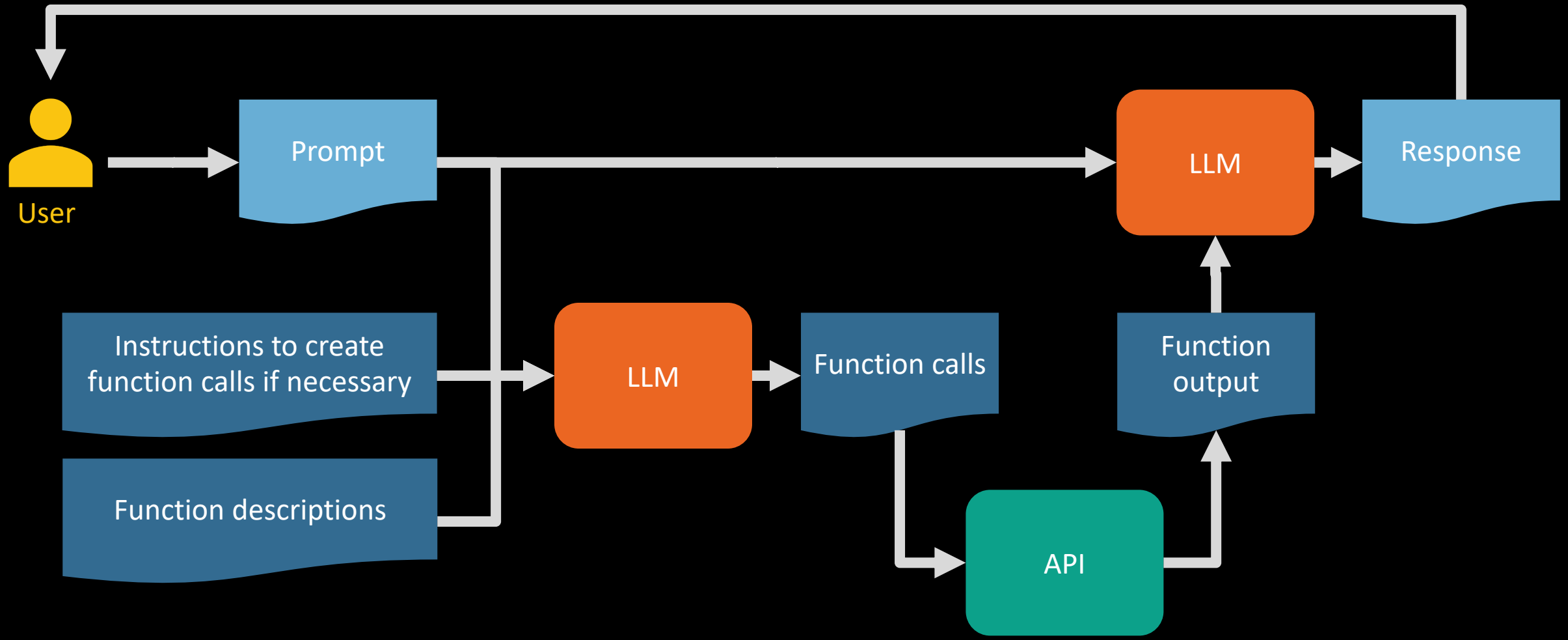
Providing tools for AI

LLMs can use tools

- LLMs can be instructed to use tools
  - To find information (e.g fetch the latest news)
  - To make changes in the world (e.g. to book a hotel)
- Anyone can make tools, and provide them to AI agents



AI function calling



API definition

OpenAPI format (recognized by ChatGPT).
Describes:

- Function purpose
- Input parameters
- URL
- Output

```yaml
openapi: 3.0.1
info:
  title: TODO Action
  description: An action that allows the user to create and manage a TODO list using a GPT.
  version: 'v1'
servers:
  - url: https://example.com
paths:
  /todos:
    get:
      operationId: getTodos
      summary: Get the list of todos
      responses:
        "200":
          description: OK
          content:
            application/json:
              schema:
                $ref: '#/components/schemas/getTodosResponse'
components:
  schemas:
    getTodosResponse:
      type: object
      properties:
        todos:
          type: array
          items:
            type: string
          description: The list of todos.
```

OHDSI tools for AI agents?

Use case 1: Support researchers

- Fetch information on databases in the OHDSI network
- Fetch information about the CDM and its conventions
- Fetch information about concepts in the vocabulary
- Fetch information about concepts in a (set of) database(s)
- Generate concept sets and cohort definitions
- Generate study specifications

Don’t just do “Write SQL that computes the incidence of diabetes”. Collaborate with LLM to follow best practices, such as the use of validated phenotype algorithms

OHDSI tools for AI agents?

Use case 2: Support clinical decision making

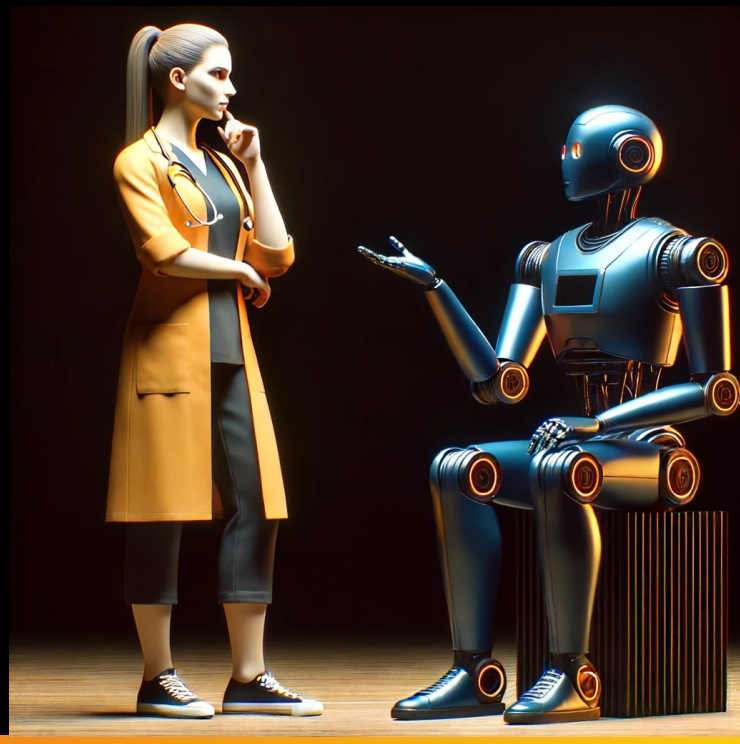
- Fetch results from HowOften
- Fetch results from LEGEND studies

Fits OHDSI's call for improved dissemination

Example:

"When prescribing X, should I be concerned about side-effect Y?"

"LEGEND's meta-analysis found a hazard ratio of … for Y when comparing X to Z. About …% of people experience Y while on X."



Should we generate additional results that fit this use case?


OHDSI
Observational Health Data Sciences and Informatics

Thank you!