

EHRFlow: A Large Language Model-Driven Iterative Multi-Agent Electronic Health Record Data Analysis Workflow



GitHub Code

Hao Wu*, Yinghao Zhu*, Zixiang Wang, Xiaochen Zheng, Ling Wang, Wen Tang, Yasha Wang, Chengwei Pan, Ewen M. Harrison, Junyi Gao†, Liantao Ma†



HDRUK
Health Data Research UK



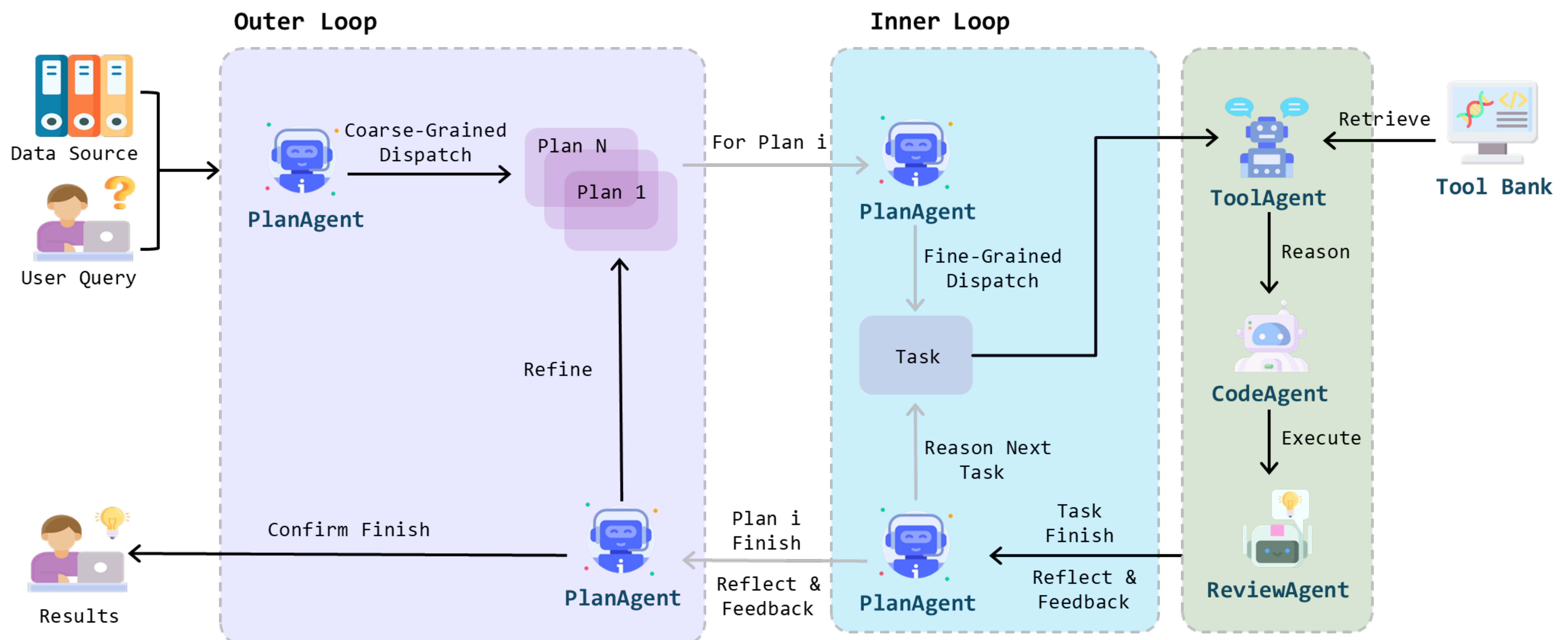
ETH

Boost EHR analysis pipeline for clinicians and medical researchers with LLMs

- Reduce the use barrier of data processing and deep learning for physicians and data scientists in EHR analysis.
- Avoid hallucinations and privacy concerns of applying LLMs in medical settings

Objective: Complete complex EHR data analysis works using natural language. Clinicians set the targets, and the platform understand the targets, make plans and write codes, without **human intervention**.

Motivation



Overview framework of EHRFlow

Four main agents:

- PlanAgent, ToolAgent, CodeAgent, and ReviewAgent

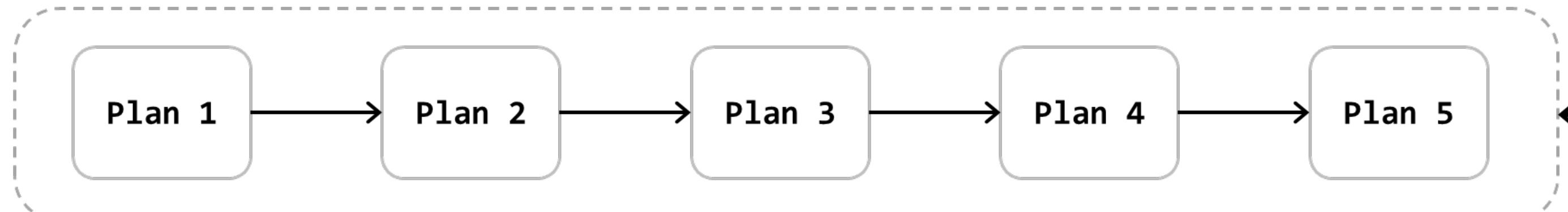
Dual-loop mechanism:

- Outer loop: Task decomposition into course-grained plans
- Inner loop: For each plan, fine-grained tasks dispatch, execution, and reflection

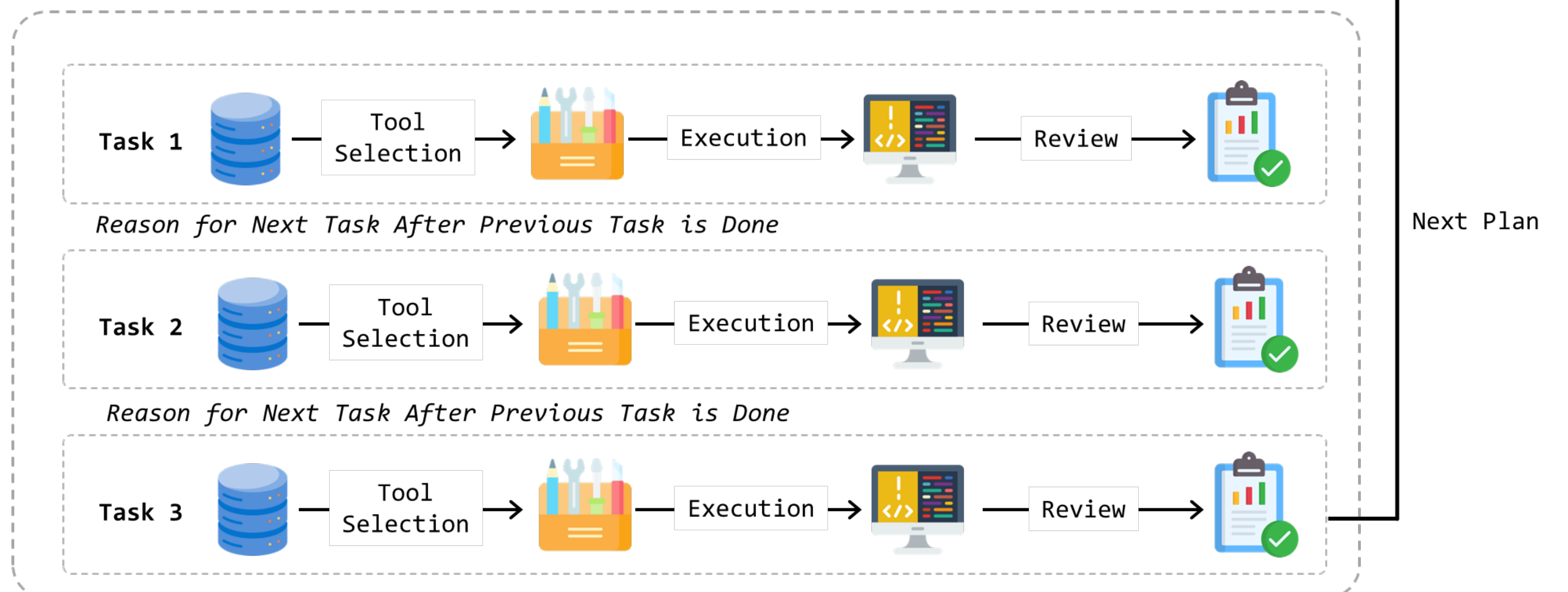
Privacy protection:

- Only data schema provided, sensitive operations executed locally

Coarse-Grained Plans in the Outer Loop

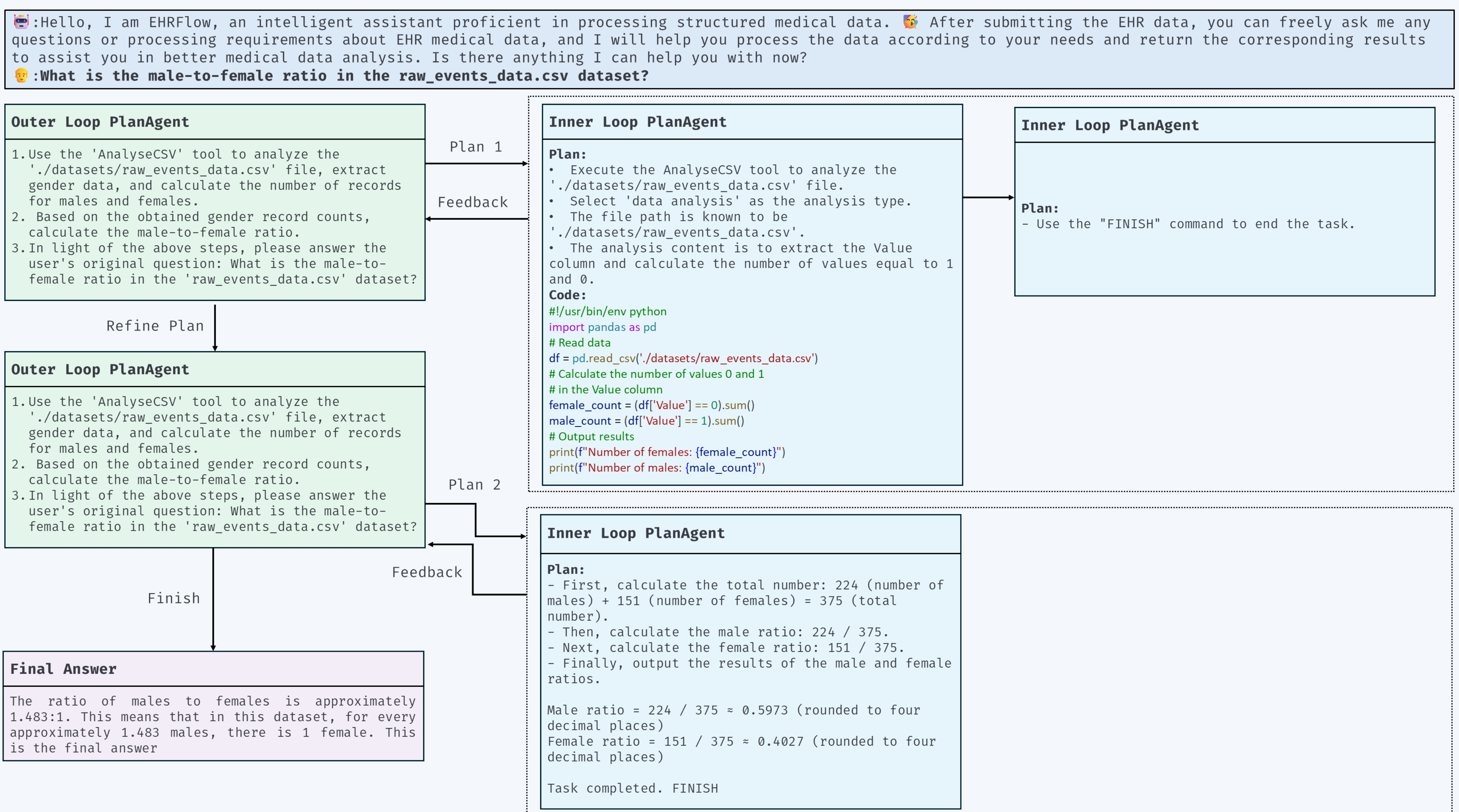


Fine-Grained Tasks in the Inner Loop



Plan and task decomposition and execution pipeline

Pipeline



Case Study

