



Evidence-Grounded Tree-Based Hypothesis Generation for Scientific Discovery

Changhee Cho^{*1}, YeJun Ji^{*1}, Su-Young Jin^{*1}, Sumin Lee^{*2}, Jiho Sim^{*2}

^{*} Equal Contribution

¹ Graduate School of Data Science, Seoul National University, Seoul, Korea, ² Department of Chemistry, Seoul National University, Seoul, Korea



TL;DR: Evidence-Grounded AI Scientist(EGAS) decouples retrieval and reasoning into a Context Optimizer and Tree-of-Thoughts Hypothesis Generator, enabling open-ended discovery, falsifiable hypotheses, and a knowledge base that grows across cycles.

Introduction

1. LLM priors bias the evidence base

The LLM priors still influence what information is retrieved and how the retrieved information is interpreted and weighted.

2. Ideation is often treated as single-pass generation

Generating hypotheses in one step makes the process overly dependent on the model's prior knowledge and limits the discovery of unexpected connections.

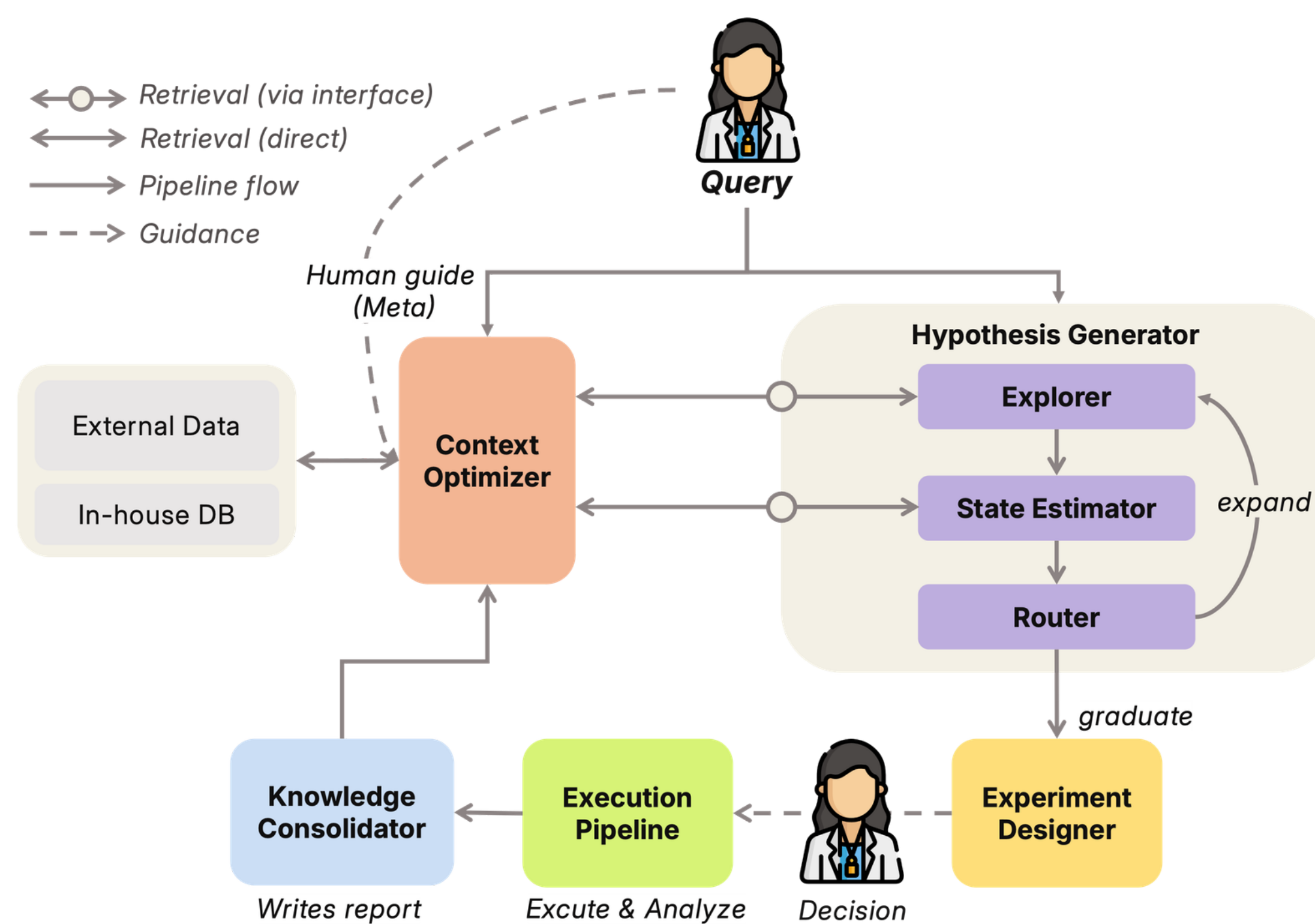
1. Curate Evidence without interpretation

EGAS strictly separates evidence curation from reasoning, so hypotheses are built on a **prior-free evidence base**.

2. Ideation as iterative grounded search

A Tree-of-Thoughts generator explores **multiple reasoning paths** over curated evidence, and experimental outcomes are **fed back** to extend the search.

EGAS Pipeline



Hypothesis Generator (HG)

HG performs a grounded Tree-of-Thoughts search over evidence curated by CO, independently of stochastic sampling.

Key idea 1: Structurally induced diversity

- The **Explorer** induces diversity by partitioning retrieved evidence into distinct subsets and exploring each subset through different reasoning modes. This produces sibling hypotheses that differ along interpretable dimensions.

Key idea 2: Grounding-aware evaluation

- The **State Estimator** evaluates each hypothesis at two levels:
 - (1) Node-level checks whether each hypothesis is evidence-backed and testable.
 - (2) Trajectory-level checks whether the reasoning step is justified by the evidence.
- These evaluation signals guide the **Router** in deciding whether a hypothesis should be refined, removed, or advanced.

Context Optimizer (CO)

CO acts as a "**Librarian**" that curates and indexes evidence **without interpreting, ranking, or summarizing it**.

Key idea: Interpretation-free curation

- All evidence is stored as atomic claims with provenance pointers, accessible to HG only through restricted retrieval interfaces.
- This strict separation preserves an unnarrowed inference space for HG, and makes hypothesis failures traceable to either the evidence layer or the reasoning layer.

Figure 1. Overall pipeline of EGAS.

Evidence curation (by the Context Optimizer) is strictly separated from reasoning (by the Hypothesis Generator). Human guidance is supported at multiple points in the pipeline.

The Explorer generates and expands hypotheses, the State Estimator evaluates them, and the Router decides the next search action.

How Would EGAS Work?

