



Jerry Yan



Leo Hu

Problem Statement and Solution

Our aim is to develop a **robust, easy-to-maintain LLM-based tool** to enable Accenture consultants to efficiently query **proprietary life sciences data**, thereby helping **Accenture clients** in the **pharmaceutical industry** better-understand their competitive landscape.

Accenture Life Science Consultants need **timely, accurate insights** into the competitive landscapes of Accenture clients, drawn from complex relational databases.

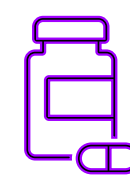
Interacting with consultants, manually **pulling data**, and iteratively refining queries can take up to **2-3 business days** per analysis for Life Science researchers, delaying strategic recommendations and exposing teams to manual-entry errors.

We propose a **multi-agent SQL Agent** that translates plain-English prompts into SQL code, leverages additional **semantic understanding** to navigate schema relationships, and **self-refines** its performance in a closed loop, delivering reproducible, high-confidence insights **in ~1 minute**.

Data & Scope: The *EvaluatePharma* Dataset



Corporate Financial Data
Revenue, Sales, R&D Cost



Product Data
Sales, Technology, Indication



Clinical Trials Data
Pipeline, Cost, Timeline

Industry Leading **Life Science** Market Intelligence Dataset
2000-2030 (Forecast)

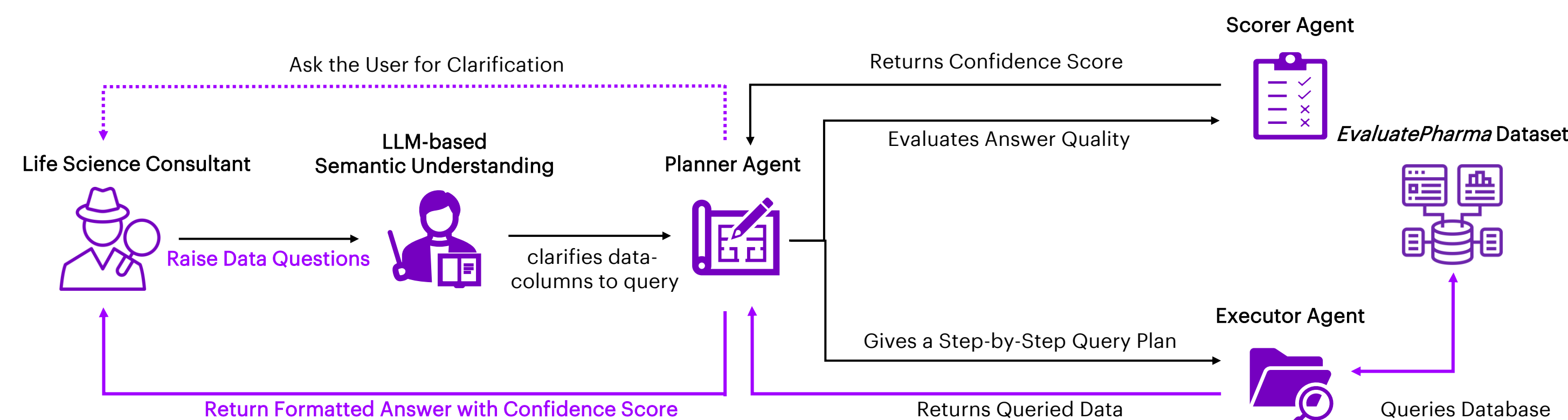
32K
Companies

104K
Products

850
Indications

Methodology

1. Framework: Data-Retrieval AI Chat Agent



2. Performance Statistics

1 Minute

Query Turnaround

99%

Wait Time Saved

+ 3.5%

Projected Productivity Boost

Novel LLM-based Data Retrieval Chat Agent

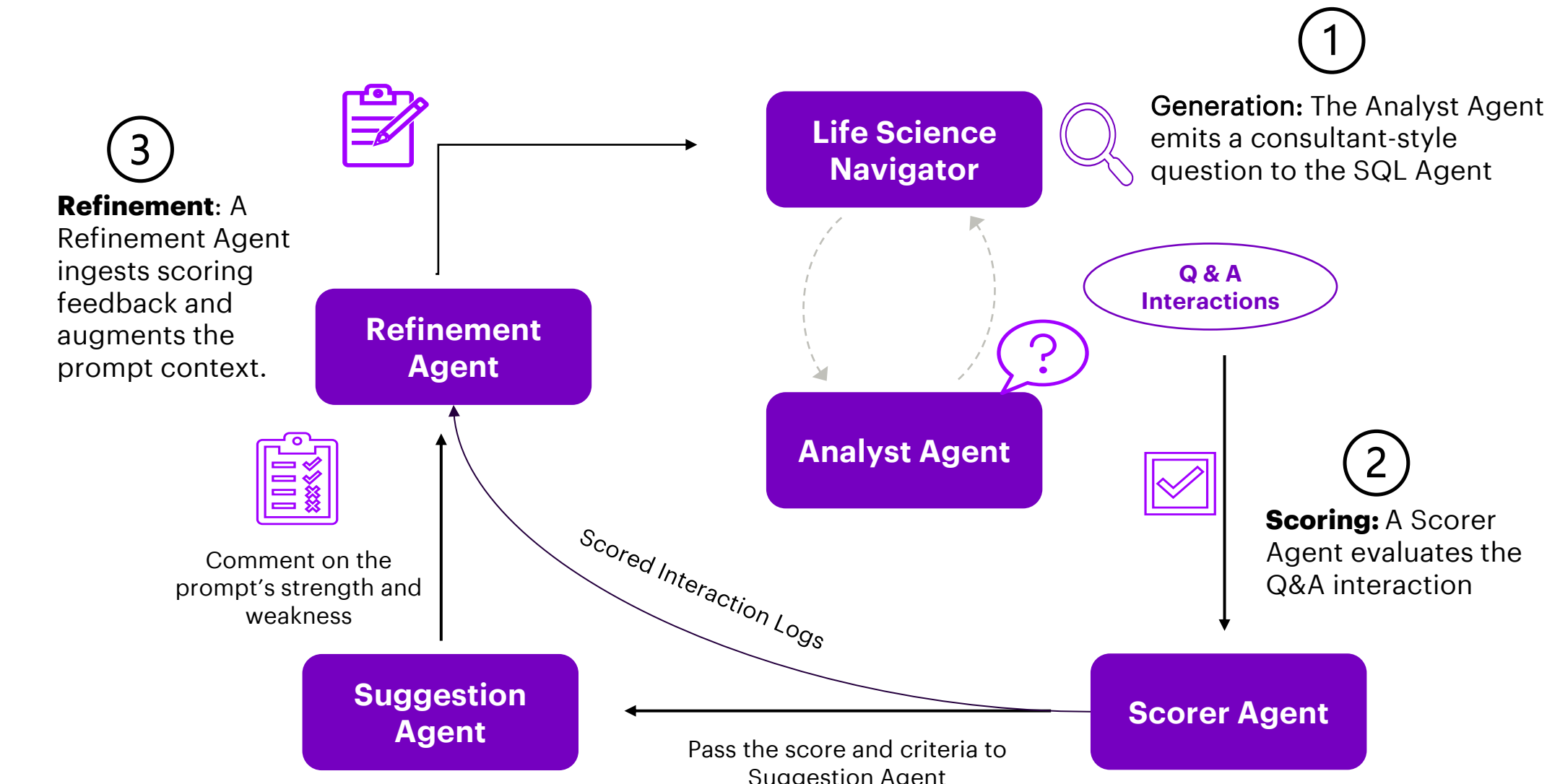
Ready to use Pipeline and User Interface

3.1. Automatic Prompt-Refinement

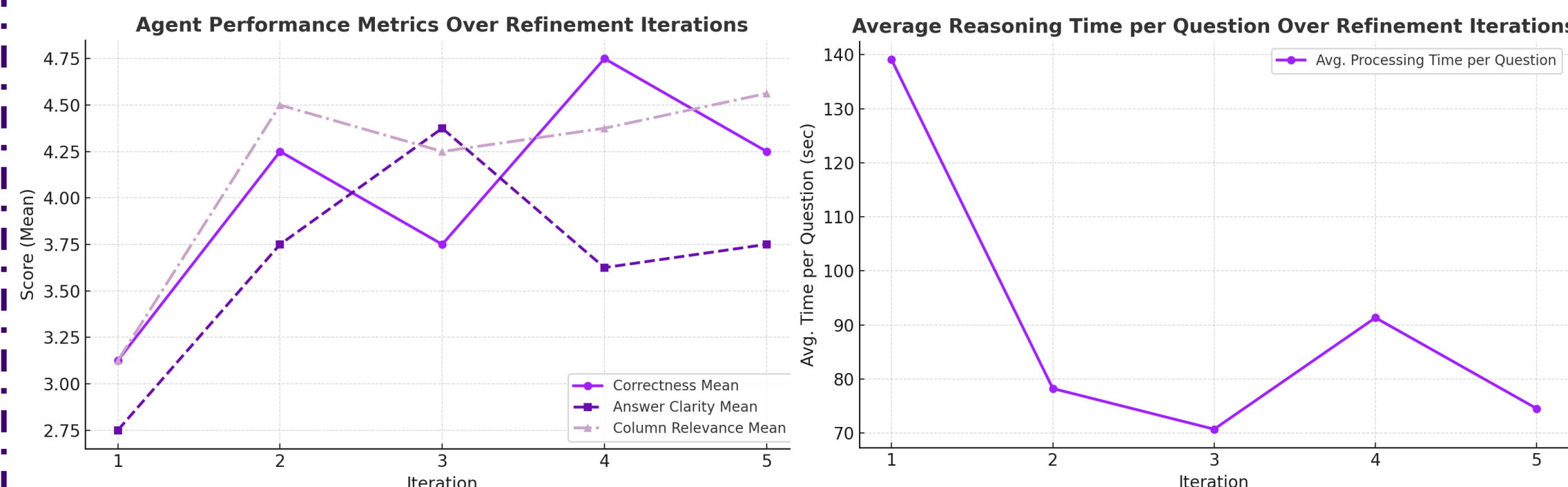
Motivation:

Manual prompt engineering for LLM-powered SQL generation is **laborious** and **fragile**: engineers must iteratively tweak system messages, few-shot examples, and parameters to handle schema nuances and edge cases.

Solution: We equip the SQL Agent with a closed-loop **self-refinement** pipeline:



3.2. Performance Uplift Through Self-Refinement



Performance Gain On Failure Cases After 5 Iterations:

+40%

Use Cases:

Self-improvement over failed queries.
Significantly **faster scaling** to new datasets.

Average Reasoning Time Per Question:

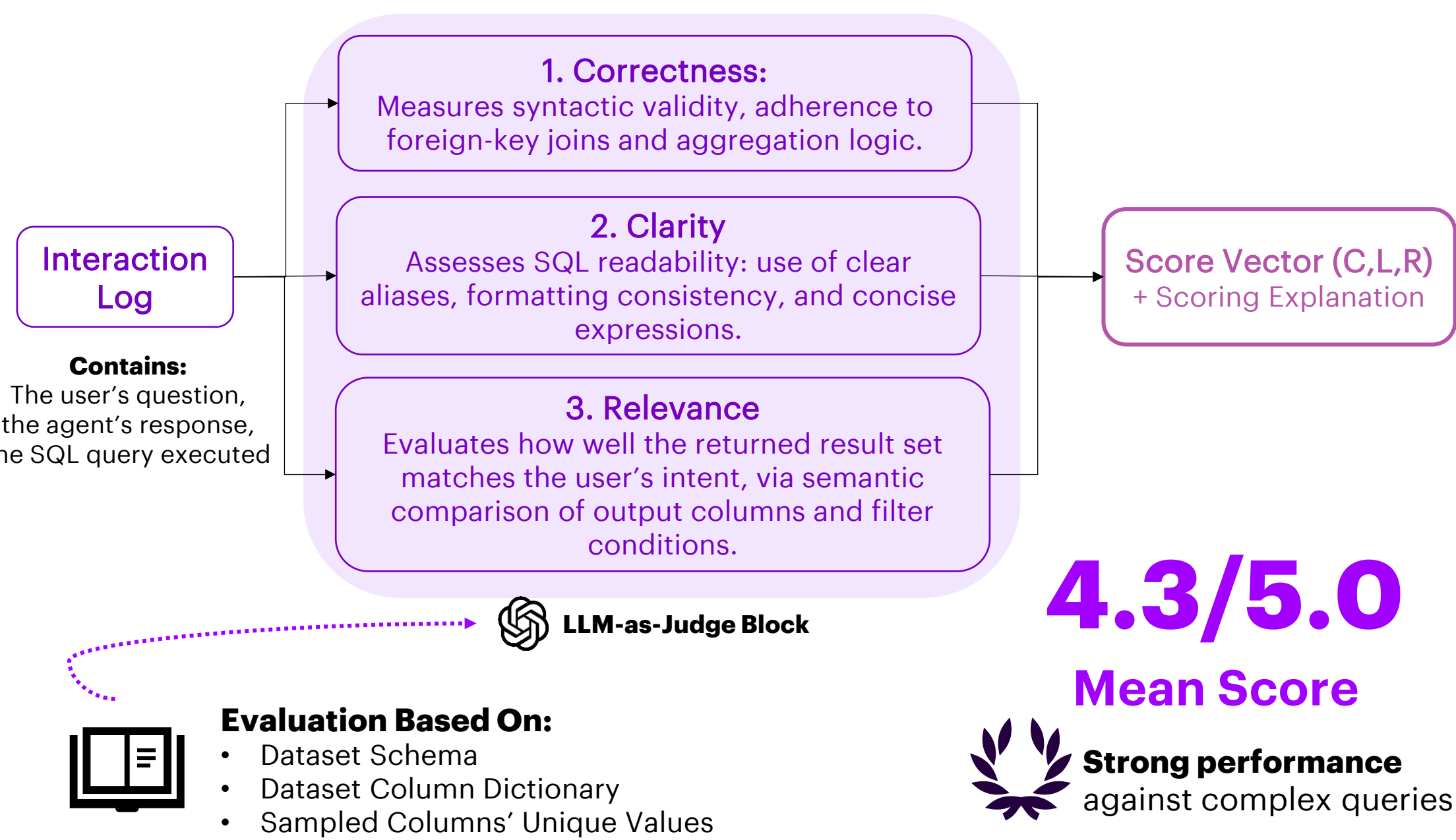
-46%

Advantage:

Significant **reduction of engineering time** spent on manual prompt-engineering.

3.3. Answer Quality Evaluation Framework

We employ an LLM-as-Judge to quantify SQL answer quality across three dimensions: **Correctness**, **Clarity**, and **Relevance**. For each user query, the SQL Agent's output is fed into a structured judge prompt that returns **numeric scores (1-5)** on each metric.

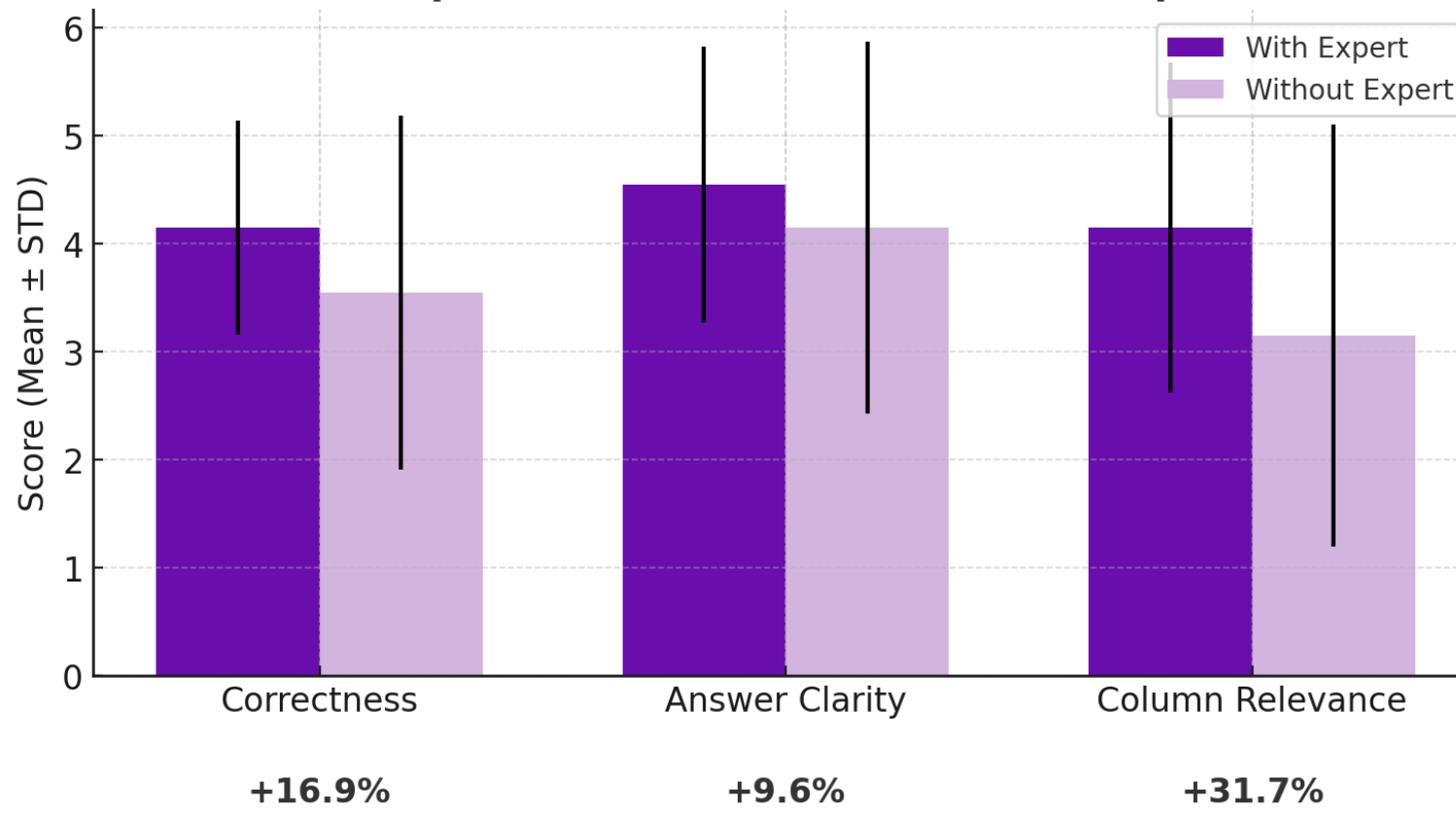


3.4. Semantic Understanding Module

We augment the core SQL pipeline with an **intelligent helper agent**, the Table Expert. Upon receiving the analyst's natural-language question, it internally:

- 1. Ingests** the natural language query.
- 2. Consults** its prompt memory of schema definitions and representative column values.
- 3. Performs** a lightweight semantic match to pick the top relevant tables and columns.
- 4. Emits** a structured hint string appended to the Planner Agent's context.

Performance Comparison With vs Without Table Expert Intervention



Average Performance Score:

+19%

Insight:

Semantic disambiguation significantly enhances SQL Agent performance.

Results and Impacts

- Client Engagement Enhancement** - Delivers on-demand, minute-latency analytics to client teams - data that would otherwise arrive too late for high-stakes meetings.
- Consultant Productivity Gains** - Shrinking data-pull turnaround from 2-3 business days to under one minute per query, giving consultants insights exactly when they need them, enabling more follow-up questions and deeper competitive analysis.
- Engineering Agility** - Saves **300 h/year** on performance tuning and dataset onboarding for the Accenture engineers, freeing capacity for feature development.
- Researcher Focus** - Reclaims up to **500 h/year** of data curation, a non-core task for researchers, allowing researchers to devote more effort to core Accenture Research tasks: model development, thought leadership, and strategic analysis.

Beta Release Interface

How many active companies are in the dataset?

There are 20949 active companies in the dataset.

Confidence Score: High

Show Plan and Query

Suggested Next Questions

- Can you show me the distribution of companies by their 'status' (e.g., 'Inactive (acquired)', 'Inactive (bankrupt)')?
- What are the top 5 classifications of these active companies?
- Which countries have the most active companies headquartered there?