LLM-Symbolic Integration for Robust Temporal Tabular Reasoning

Atharv Kulkarni[1], Kushagra Dixit[1], Vivek Srikumar[1], Dan Roth[2], Vivek Gupta[3]

[1]University of Utah, [2]University of Pennsylvania, [3]Arizona State University

Scan Me

Introduction

- Large language models excel with free-form text, but fail on queries about time on tabular data.

- Large tables, filters, and counterfactuals bloat the prompt and lead to incorrect reasoning.

Can we give models a structured view of the data and then let them reason correctly via SQL?
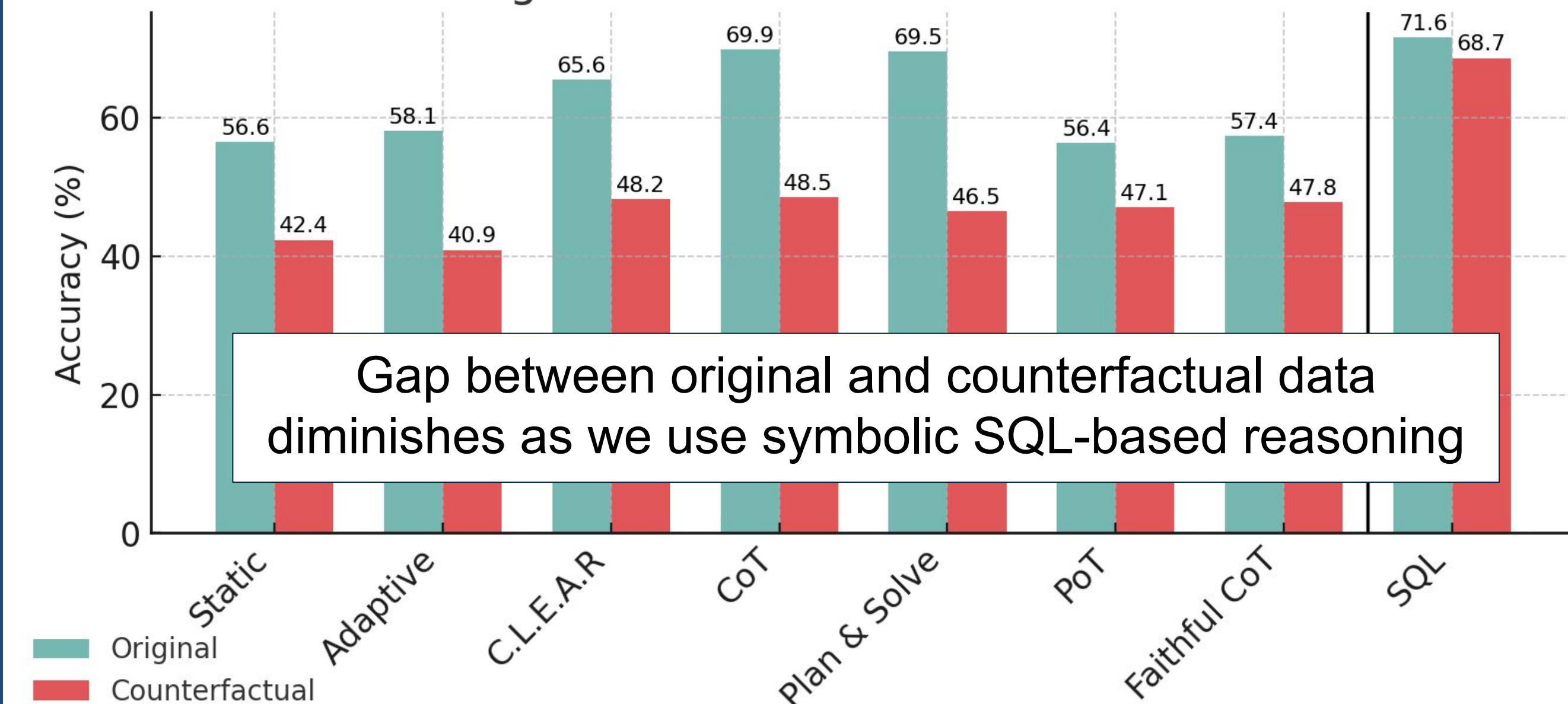
Approach

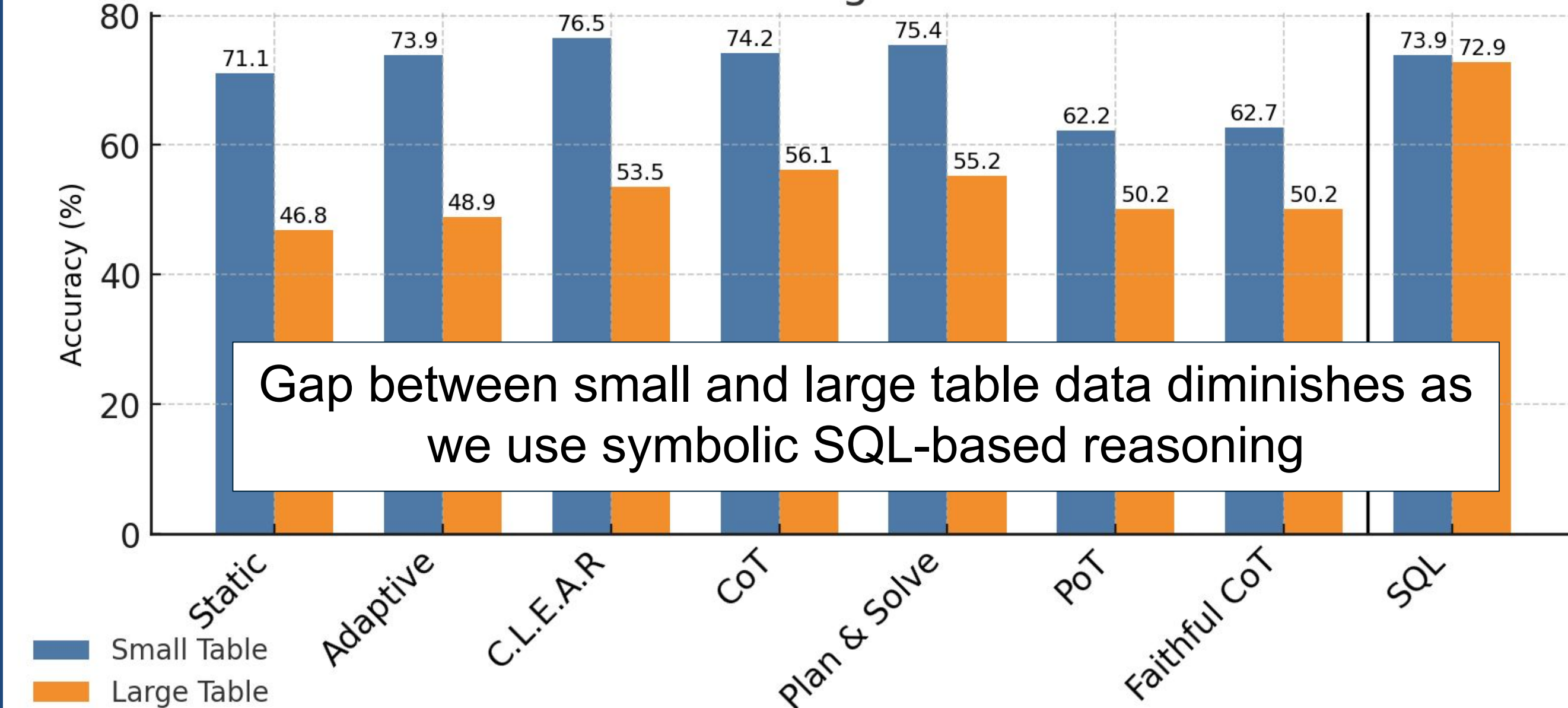### Symbolic Intermediate Representation

- **Hide the noise:** Rather than show all the data in the table, we only show the schema of the data.

- **Ask for SQL**, not prose:

  Write an SQL query that answers the question using the database schema.

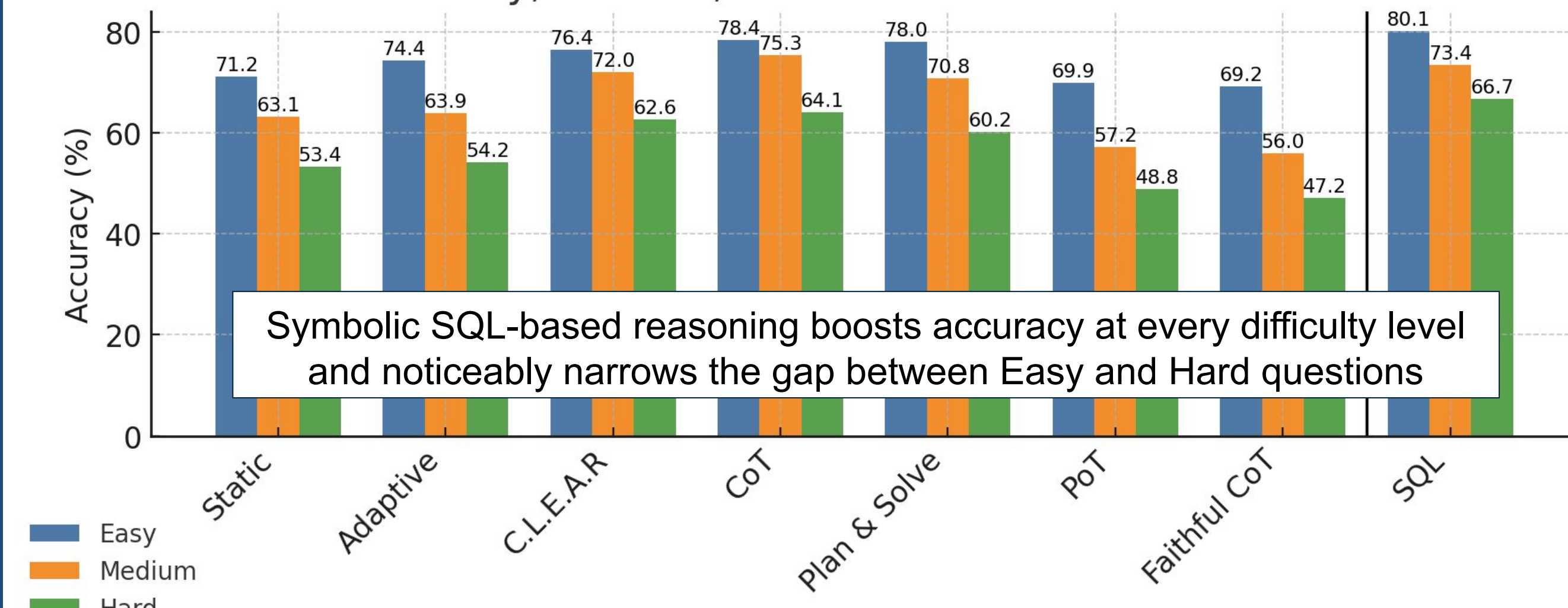- **Run:** Execute the query on the database and return the result.

Results

Original vs Counterfactual for GPT-4o.

Gap between original and counterfactual data diminishes as we use symbolic SQL-based reasoning

- Original
- Counterfactual

Static 56.6 / 42.4
Adaptive 58.1 / 40.9
C.L.E.A.R. 65.6 / 48.2
CoT 69.9 / 48.5
Plan & Solve 69.5 / 46.5
PoT 56.4 / 47.1
Faithful CoT 57.4 / 47.8
SQL 71.6 / 68.7

Small Table vs Large Table for GPT-4o.

Gap between small and large table data diminishes as we use symbolic SQL-based reasoning

- Small Table
- Large Table

Static 71.1 / 46.8
Adaptive 73.9 / 48.9
C.L.E.A.R. 76.5 / 53.5
CoT 74.2 / 56.1
Plan & Solve 75.4 / 55.2
PoT 62.2 / 50.2
Faithful CoT 62.7 / 50.2
SQL 73.9 / 72.9

Easy, Medium, and Hard results for GPT-4o.

Symbolic SQL-based reasoning boosts accuracy at every difficulty level and noticeably narrows the gap between Easy and Hard questions

- Easy
- Medium
- Hard

Static 71.2 / 63.1 / 53.4
Adaptive 74.4 / 63.9 / 54.2
C.L.E.A.R. 76.4 / 72.0 / 62.6
CoT 78.4 / 75.3 / 64.1
Plan & Solve 78.0 / 70.8 / 60.2
PoT 69.9 / 57.2 / 48.8
Faithful CoT 69.2 / 56.0 / 47.2
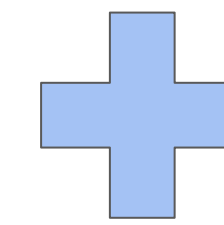SQL 80.1 / 73.4 / 66.7

Direct Prompting

Question

Q. *In which year did Áron Szilágyi achieve his personal highest number of gold medal wins?*
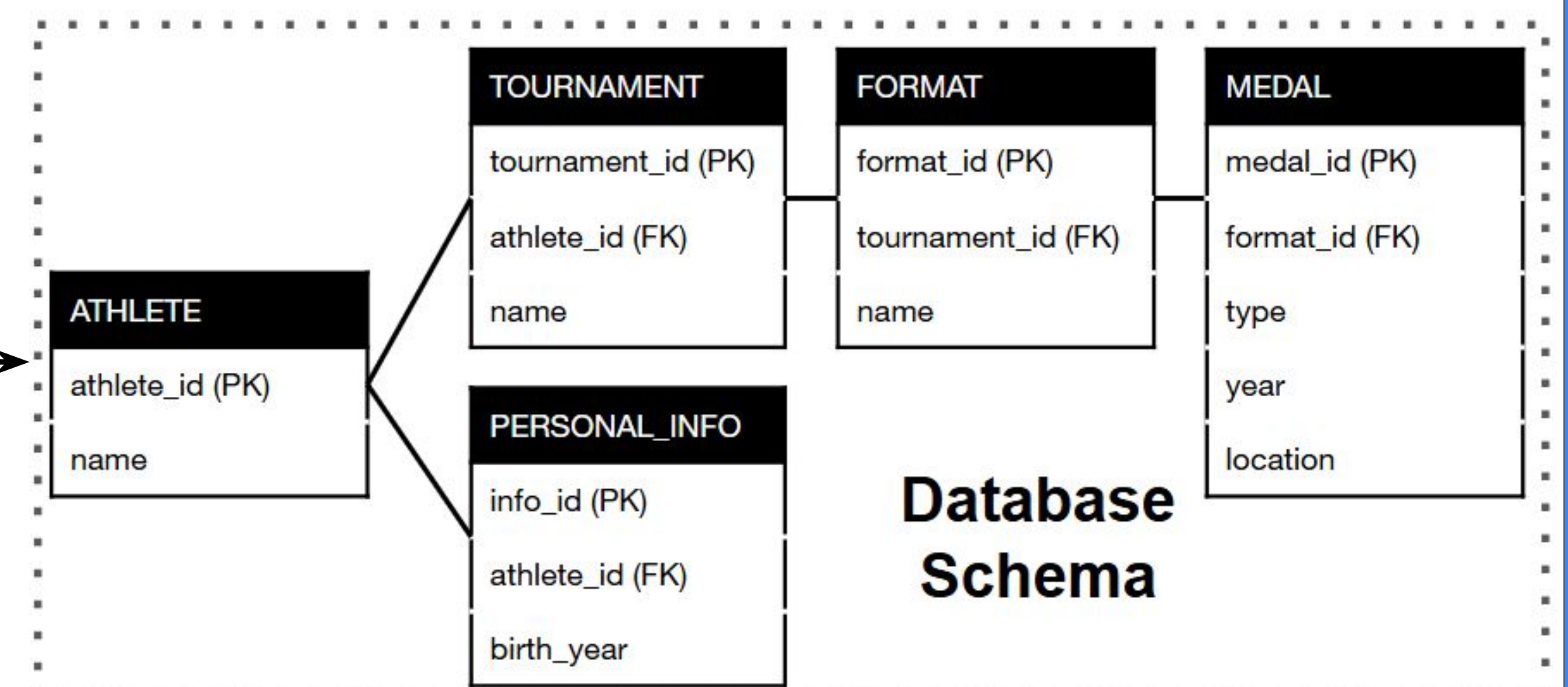
Correct answer: 2022

LLM

Direct prompt (Chain-of-Thought) — Wrong

In 2020, Áron Szilágyi won a Gold Medal in the Individual format at the Tokyo Olympics. This was his most successful year in terms of highest Gold Medal wins since he did not win multiple Gold Medals in any other format or competition.
Predicted answer: 2020 ✗ (Wrong)

Symbolic Integration

Database Schema

Q. In which year did Áron Szilágyi achieve his personal highest number of gold medal wins?

LLM

```sql
WITH gold_medal_counts AS (
    SELECT m.year, COUNT(m.medal_id) AS gold_medals
    FROM Medal m
        JOIN Format f ON m.format_id = f.format_id
        JOIN Tournament t ON f.tournament_id = t.tournament_id
        JOIN Athlete a ON t.athlete_id = a.athlete_id
    WHERE a.name = 'Áron Szilágyi' AND m.type = 'MedalGold'
    GROUP BY m.year)
SELECT year
FROM gold_medal_counts
WHERE gold_medals =
(SELECT MAX(gold_medals) FROM gold_medal_counts);
```

Execute

Answer: 2022
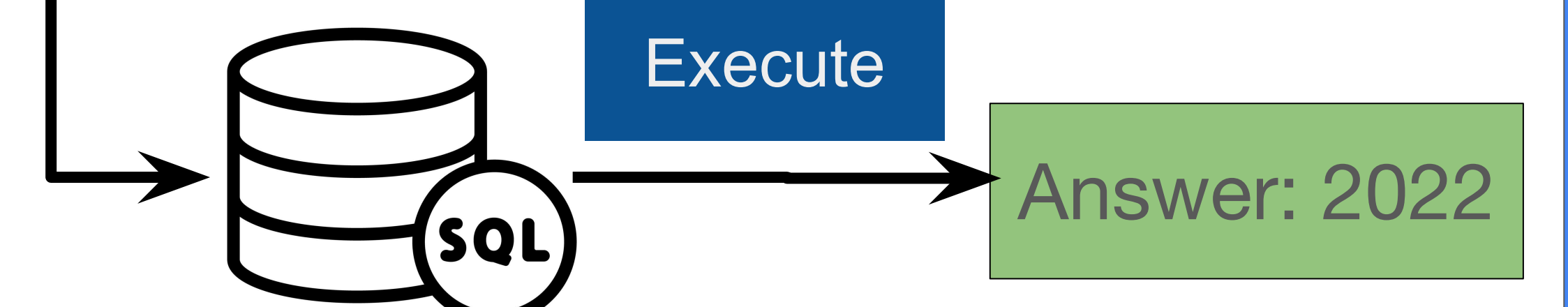
Conclusion

For temporal reasoning over tables, converting tables into symbolic schemas improves LLM reasoning by reducing reliance on memorized patterns.