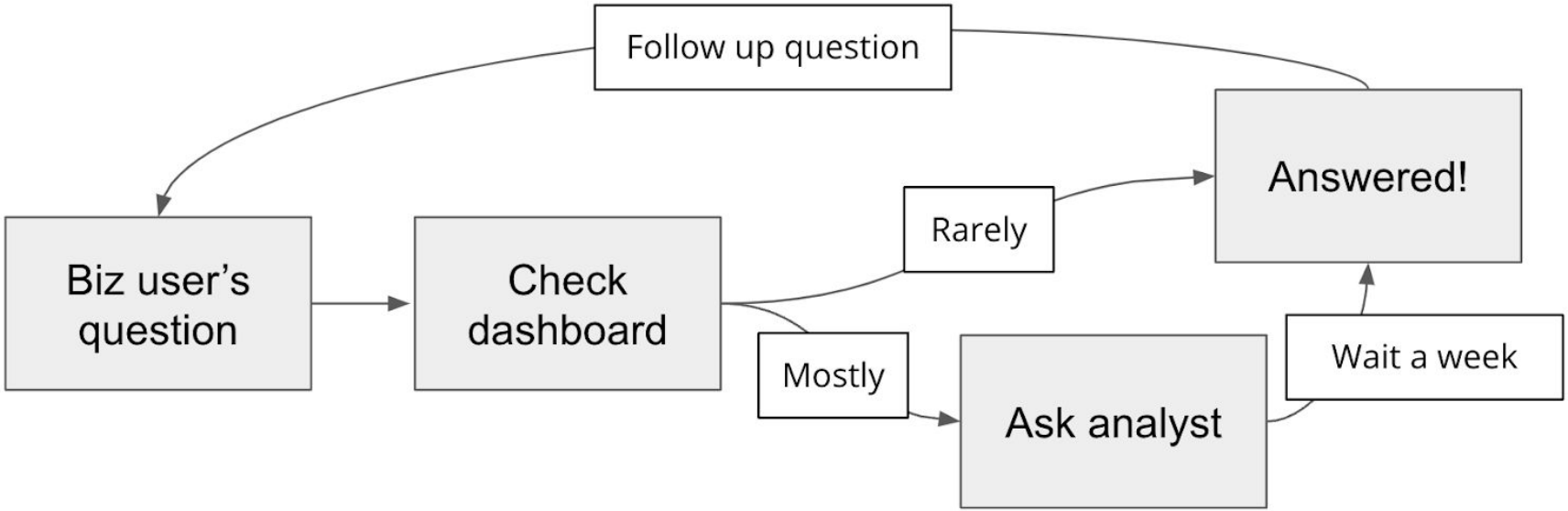
LLM - Text To SQL

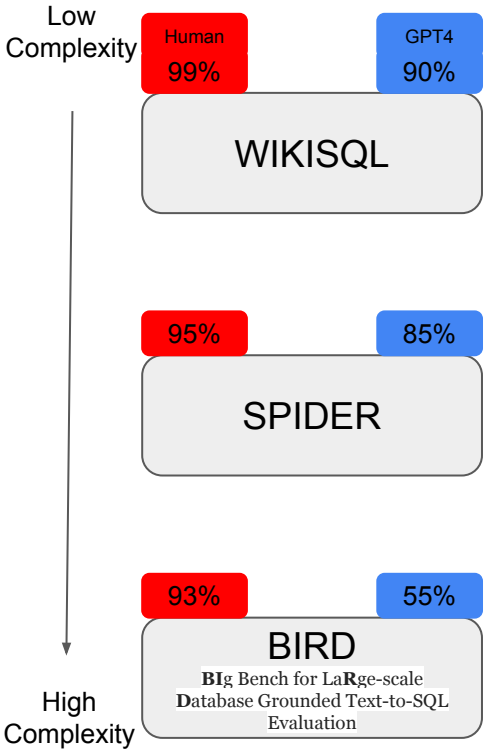
@jai-llm
24-Sep-2023

LLM Accuracy Key to Speeding Up Decision Making



Source: Vanna Link

However, LLM Performance Varies Widely from 55%- 90%



Dataset Details

- 80654 hand-annotated examples of 24241 tables from Wikipedia.
- SQL queries are exceedingly simple, with only SELECT, FROM, and WHERE clauses. No linkages to other tables.

- **Complex**: Covers GROUP BY, ORDER BY, and HAVING clauses, Nested queries, and JOINS across multiple tables linked through foreign keys.
- **Cross-Domain**: Has 200 complex databases across a high number of domains, Spider is able to include unseen databases in the test set, allowing us to test the model's generalizability.

- Data was collected from real-world scenarios, retains their original, "dirty" format.
- It also provides external knowledge, similar to how real-world developers may have external knowledge from metadata, docs, or other existing context stores.

Source: Medium Link

Vanna Used CyberSyn Dataset to Compare LLMs

Dataset

- CyberSyn
- **Reasons**: Representative, Accessible, Understandable, and Maintained.

Models

- Bison (Palm2), GPT-3.5-Turbo, GPT-4, Llama-2 (no results)

Context

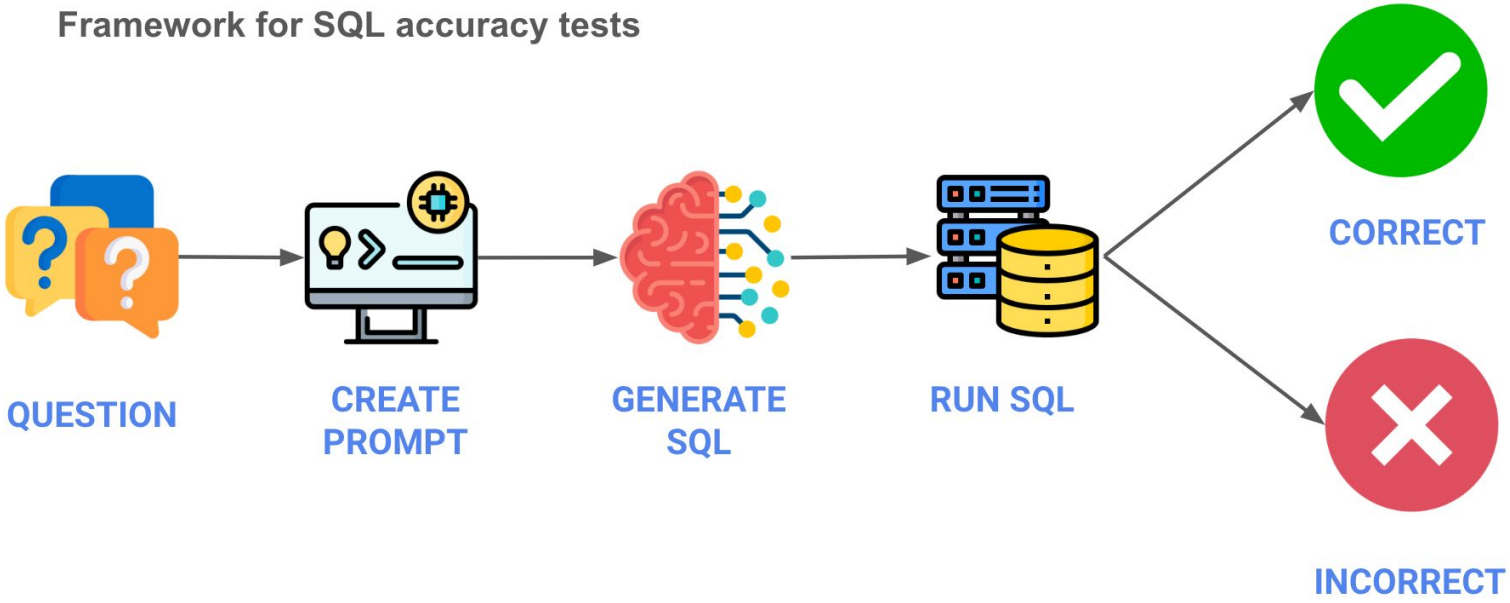
- **Schema Only** - We put the schema (using DDL) in the context window.
- **Static Examples** - We put static example SQL queries in the context windows.
- **Contextually Relevant Examples (RAG)** - Finally, we put the most relevant context (SQL / DDL / documentation) into the context window, finding it via a vector search based on embeddings.

Questions

- How many companies are there in the dataset?
- What annual measures are available from the 'ALPHABET INC.' Income Statement?
- What are the quarterly 'Automotive sales' and 'Automotive leasing' for Tesla?
- How many Chipotle restaurants are there currently?

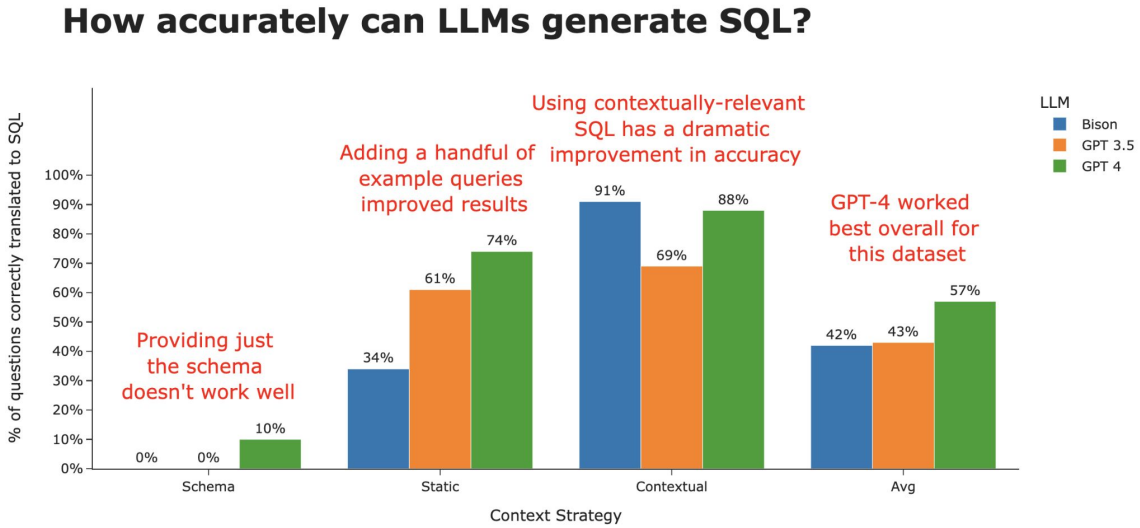
Source: Vanna Link

Vanna Evaluation Framework



Source: Vanna Link

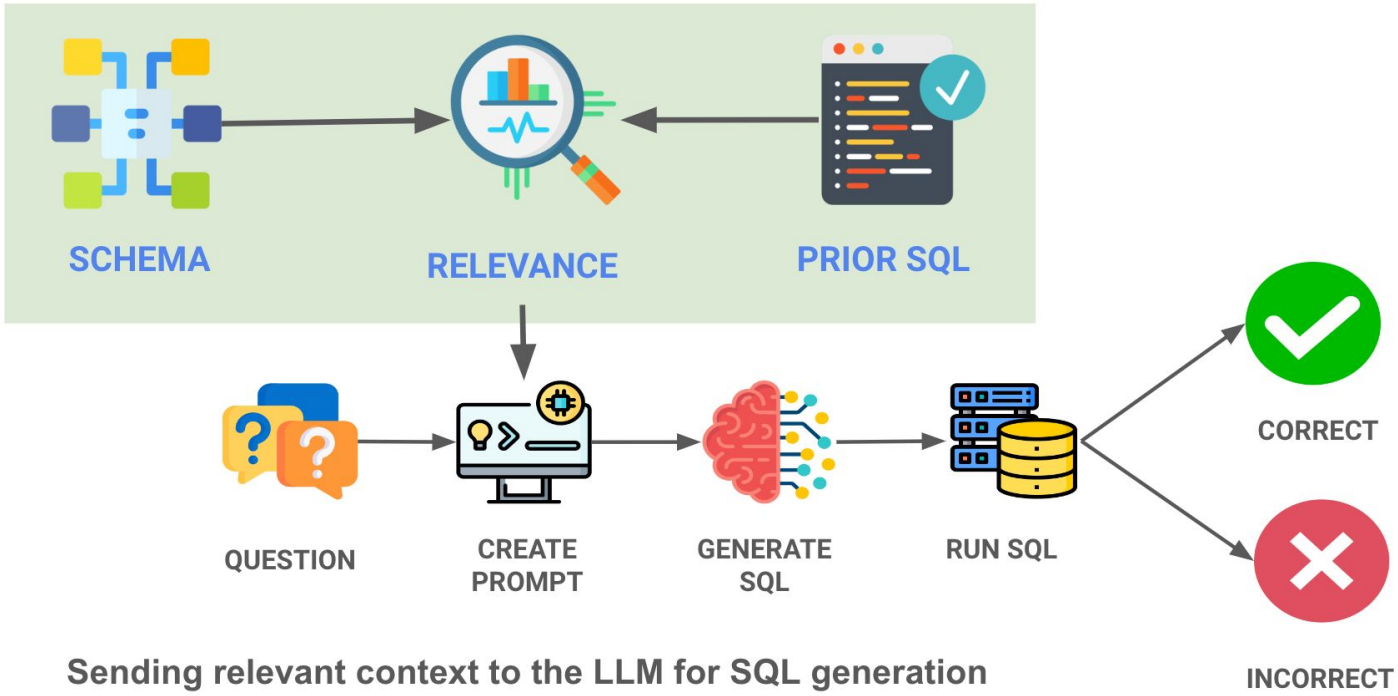
Vanna Showed RAG Dramatically Improves SQL Quality



Accuracy	Bison	GPT 3.5	GPT 4	Avg
Schema	0%	0%	10%	**3%**
Static	34%	61%	74%	**56%**
Contextual	91%	69%	88%	**83%**

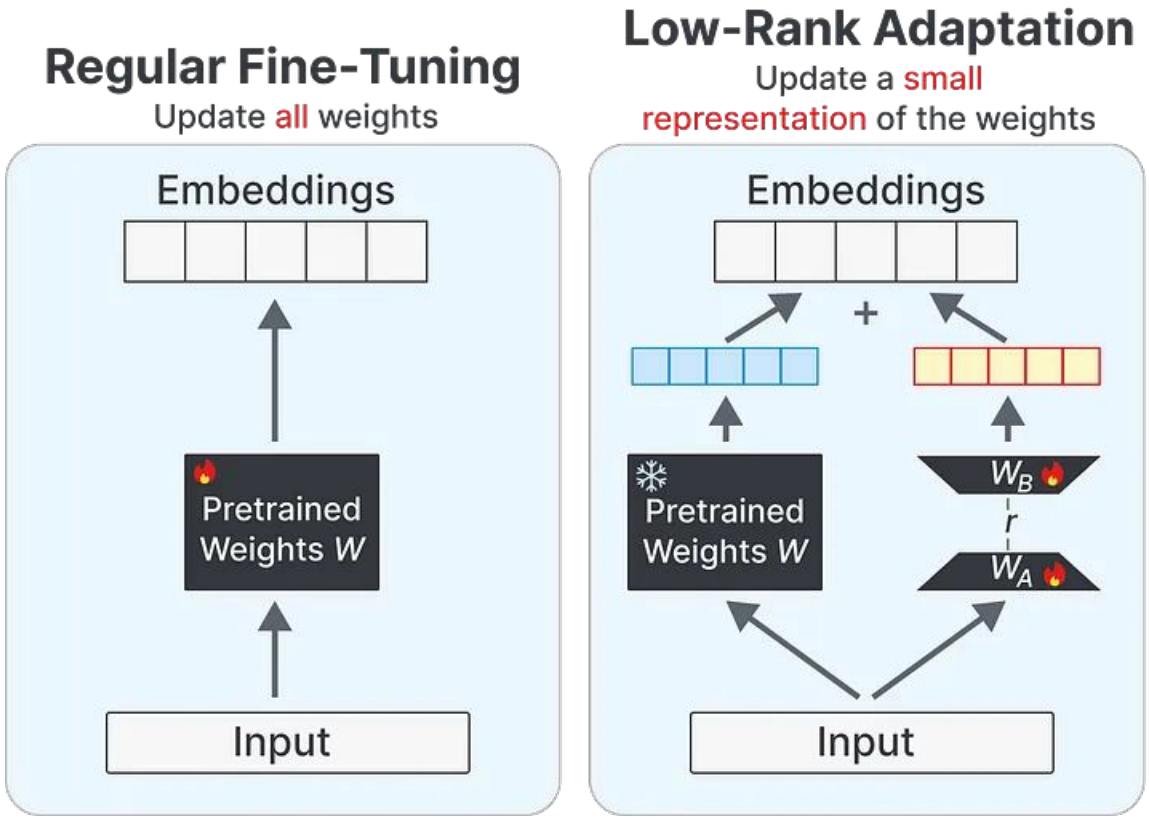
Source: Vanna Link

Vanna RAG Consisted of Schema + Prior SQL



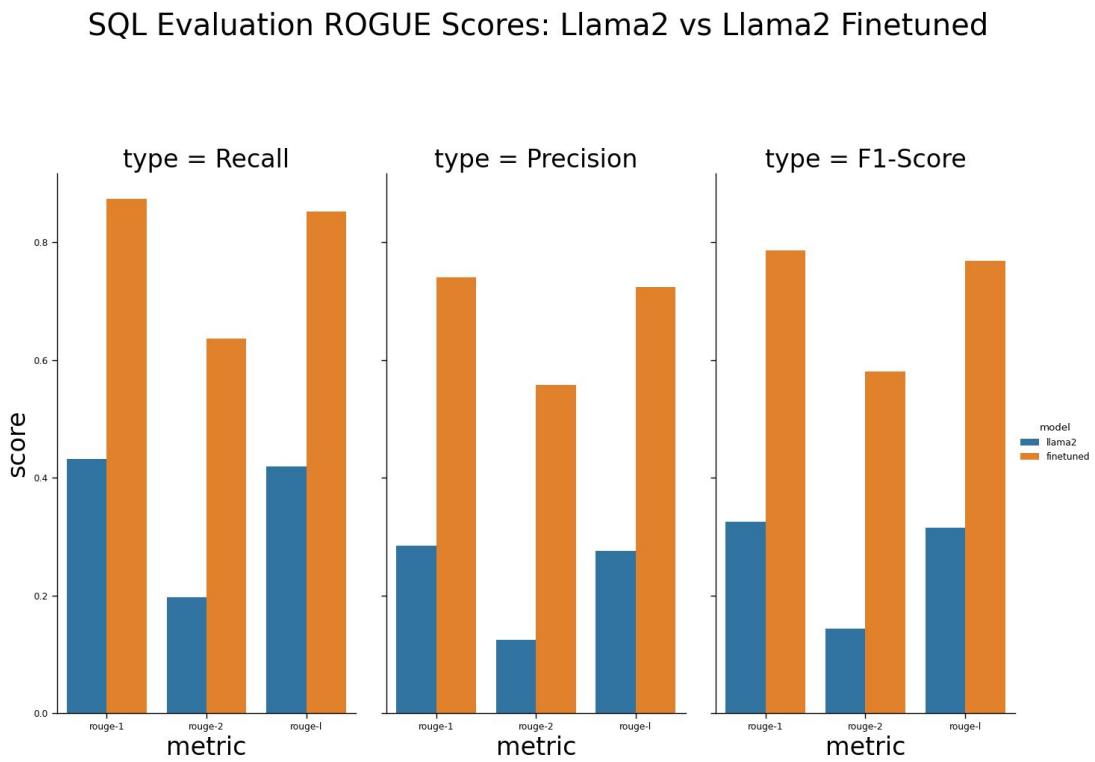
Source: [Vanna Link]

Fine-tuned Llama2 Using PEFT + LoRA



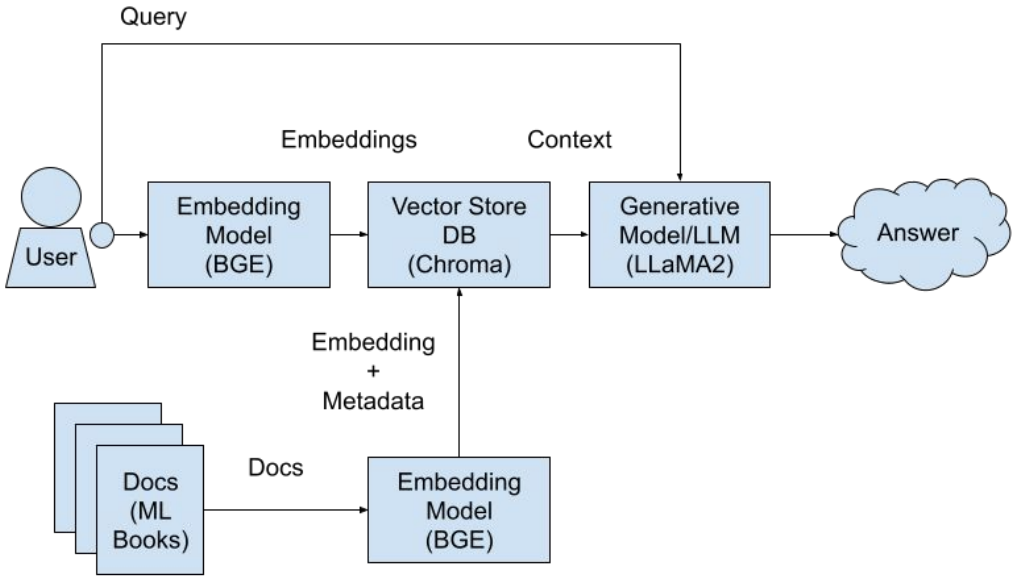
Source: Medium

Finetune Shows 2-3x Better ROGUE Scores on Text2SQL



Source: [Jai Github]

Next Steps: Use Fine-Tuned LLM + RAG



- **Docs**:
  - Schema
  - SQL Queries
- **Fine-tuned LLaMA2**:
  - PEFT + LoRA Finetune