



OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

FAKULTÄT FÜR
INFORMATIK

SQLAssist:

A RAG unified framework for Natural Language to SQL translation



Basavaraj Sangayya Hiremath Dhanashree Gunda Mallika Manam

Niharika Ramanath Supriya Upadhyaya



Agenda

- Motivation and Research Objective
- Literature Review
- Methodology
- Results
- Evaluation
- Experiments
- Challenges & Limitations
- Conclusion
- Future Work
- References



Motivation and Research objectives

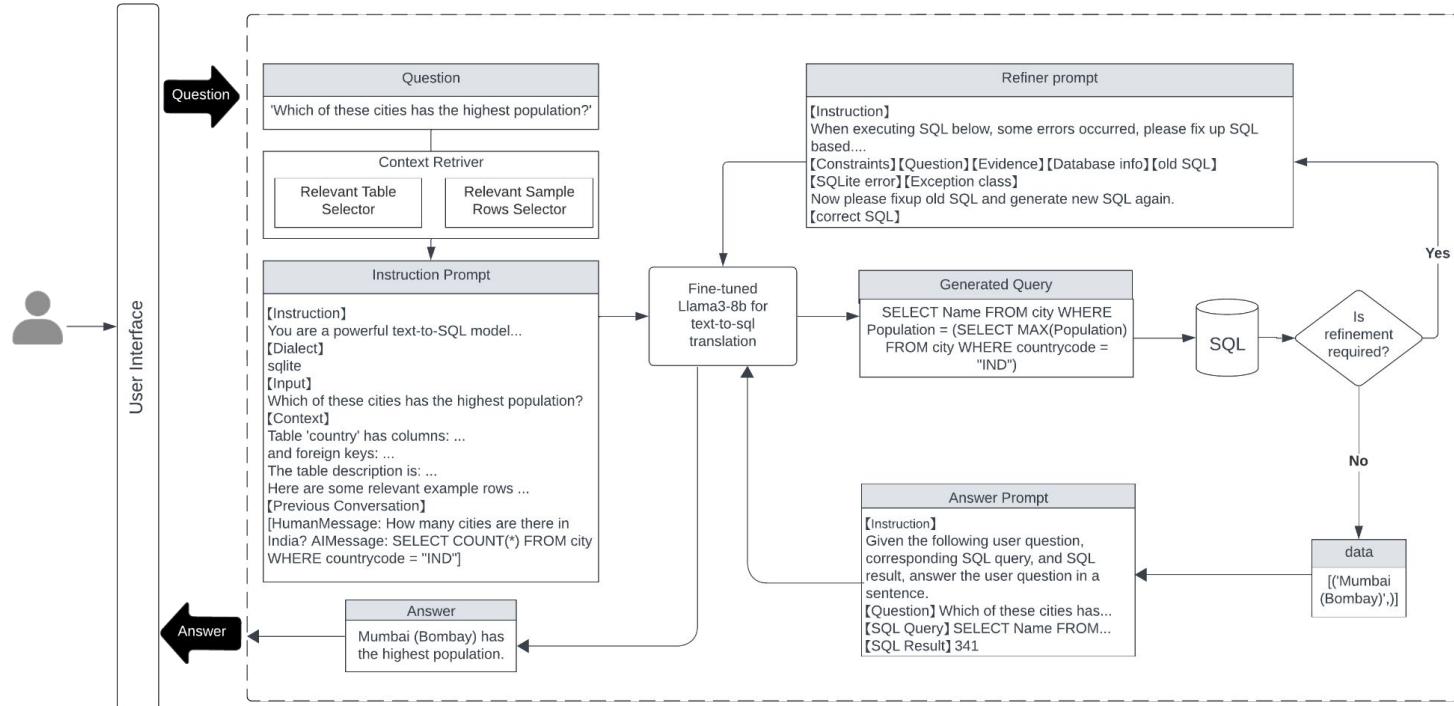
- **Challenges:** SQL query - significant barrier to non-expert users - hindering effective communication with databases.
- **Solution:** Automates SQL query generation from natural language - enhancing accessibility to databases.
- **Advancements:** LLMs - GPT4 exhibit remarkable performance in NLP tasks, particularly in the context of Text-to-SQL.
- **Research Objectives:** Our research aims to address the challenges in deploying LLMs(Llama-3) for Text-to-SQL tasks, focuses on the efficiency of text-to-SQL translation and advances the usability of databases for non-expert users.

- **Prompt Designs:**
 - LLM-based text-to-SQL tasks - refining prompt designs, recognizing the crucial role of prompts in utilizing LLMs for SQL generation.
 - Strategies such as **few-shot chain-of-thought reasoning** and **chain-of-thought prompting** to enhance LLMs' reasoning capabilities.
- **Multi-Stage Frameworks:**
 - Multi-agent collaboration architecture - Feature core decomposer agents, Refiners and auxiliary agents - collaborate to **generate SQL queries** from text, **fetch sub databases**, and **rectify incorrect SQL queries** - a holistic approach to text-to-SQL processing.
- **STAR for Context-dependent Text-to-SQL:**
 - Pre-trained TaLM designed for **context-dependent** text-to-SQL - highlights its focus on capturing the intricate relationships between natural language utterances and SQL queries throughout a conversation.
 - **Utterance dependency tracking** and **Schema state tracking** - approach leads to significant performance improvements on challenging text-to-SQL tasks.

- **Tuning Techniques:**
 - The two primary categories - **reinforcement learning** and **supervised fine-tuning**.
 - Supervised fine-tuning focuses on training models on task-specific datasets conducive to NLU, the choice of tuning strategies and the quality of data are crucial factors influencing the performance of LLMs.
- **NLP Fine-tuning Challenges:**
 - MFTCoder - enhance both efficiency and performance.
 - Focus - PEFT techniques to optimize model parameters for maximal impact while minimizing computational overhead.
- **PEFT Methods:**
 - Two notable PEFT methods - **LoRA** and **QLora**
 - LoRA aims at adapting pretrained models to specific tasks or domains by augmenting them with an auxiliary branch.
 - QLora quantizes the pretrained model to 4 bits using dual quantization and a high-precision quantization method called NF4 - enhancing efficiency without compromising performance.

Methodology

Overview



The overview of our SQLAssist framework

Methodology

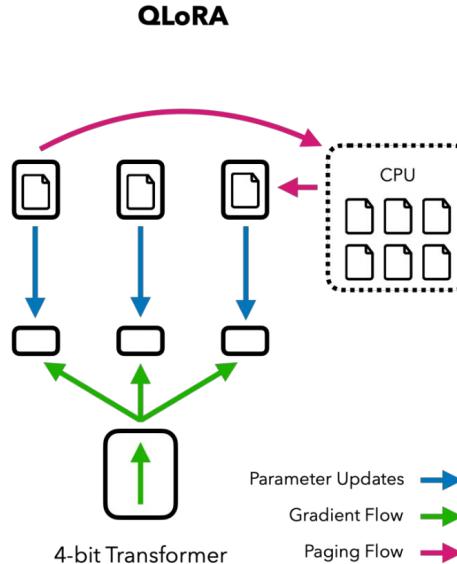
Dataset for training

- Clinton/Text-to-sql-v1
- 262k records
- Composed of multiple datasets like wikiSQL, sql-create-context, etc

Datasets: Clinton/Text-to-sql-v1		like 40	Dataset card	Viewer	Files and versions	Community
Split (1)						
train · 262k rows						
Search this dataset						
instruction string · lengths  0-106 85.5%	input string · lengths  33-744 71.6%	response string · lengths  2-446 94.5%	source string · classes  sql_create... 29.7%	text string · lengths  302-1.12k 71.5%		
Name the home team for carlton away team	CREATE TABLE table_name_77 (home_team VARCHAR, away_team VARCHAR)	SELECT home_team FROM table_name_77 WHERE away_team = "carlton"	sql_create_context	Below are sql tables schemas paired with instruction that describes a task. Using valid SQLite, write a...		
what will the population of Asia be when Latin America/Caribbean is 783 (7.5%)?	CREATE TABLE table_22767 ("Year" real, "World" real, "Asia" text, "Africa" text, "Europe" text, "Latin..."	SELECT "Asia" FROM table_22767 WHERE "Latin America/Caribbean" = '783 (7.5%)'	wikisql	Below are sql tables schemas paired with instruction that describes a task. Using valid SQLite, write a...		
How many faculty members do we have for each gender? Draw a bar chart, order by the Y-axis in descending.	CREATE TABLE Student (StuID INTEGER, LName VARCHAR(12), FName VARCHAR(12), Age INTEGER, Sex VARCHAR(1), Major...	SELECT Sex, COUNT(*) FROM Faculty GROUP BY Sex ORDER BY COUNT(*) DESC	nvbench	Below are sql tables schemas paired with instruction that describes a task. Using valid SQLite, write a...		
List the record of 0-1 from the table?	CREATE TABLE table_14656147_2 (week VARCHAR, record VARCHAR)	SELECT week FROM table_14656147_2 WHERE record = "0-1"	sql_create_context	Below are sql tables schemas paired with instruction that describes a task. Using valid SQLite, write a...		

Methodology

Instruction-Finetuning



Instruction-Finetuning



```
### Instruction:
```

You are a powerful text-to-SQL model. Your job is to answer questions about a database. You are given a question and context regarding one or more tables.

You must output the SQL query that answers the question.

```
### Dialect:
```

```
{dialect}
```

```
### Input:
```

```
{input}
```

```
### Context:
```

```
{context}
```

```
### Response:
```

Training instruction prompt template

Context Retriever

- Query-Time Table Retrieval: Dynamically retrieve relevant tables in the text-to-SQL prompt.
- Query-Time Sample Row retrieval: Embed/Index each row, and dynamically retrieve example rows for each table in the text-to-SQL prompt.

```
CREATE TABLE city (
    "ID" INTEGER NOT NULL,
    "Name" CHAR(35) DEFAULT '' NOT NULL,
    "CountryCode" CHAR(3) DEFAULT '' NOT NULL,
    "District" CHAR(20) DEFAULT '' NOT NULL,
    "Population" INTEGER DEFAULT '0' NOT NULL,
    PRIMARY KEY ("ID"),
    FOREIGN KEY("CountryCode") REFERENCES country ("Code")
)

/*
2 rows from city table:
ID Name CountryCode District Population
1 Kabul AFG Kabul 1780000
2 Qandahar AFG Qandahar 237500
*/
```

```
CREATE TABLE countrylanguage (
    "CountryCode" CHAR(3) DEFAULT '' NOT NULL,
    "Language" CHAR(30) DEFAULT '' NOT NULL,
    "IsOfficial" TEXT DEFAULT 'F' NOT NULL,
    "Percentage" DECIMAL(4, 1) DEFAULT '0.0' NOT NULL,
    PRIMARY KEY ("CountryCode", "Language"),
    FOREIGN KEY("CountryCode") REFERENCES country ("Code")
)
```

```
/*
2 rows from countrylanguage table:
CountryCode Language IsOfficial Percentage
ABW Dutch T 5.3
ABW English F 9.5
*/
```

Without Context Retriever

Input:
How many cities are in India and Germany?

Context:
Table 'city' has columns: ID (INTEGER), Name (CHAR(35)), CountryCode (CHAR(3)), District (CHAR(20)), Population (INTEGER), and foreign keys: ['CountryCode'] -> country,['Code']. The table description is: City Here are some relevant example rows (values in the same order as columns above)
(1236, 'Bahrain', 'IND', 'Uttar Pradesh', 135400)
(1203, 'Dhanbad', 'IND', 'Jharkhand', 151789)

Table 'country' has columns: Code (CHAR(3)), Name (CHAR(52)), Continent (TEXT), Region (CHAR(26)), SurfaceArea (DECIMAL(10, 2)), IndepYear (SMALLINT), Population (INTEGER), LifeExpectancy (DECIMAL(3, 1)), GNP (DECIMAL(10, 2)), GNPOld (DECIMAL(10, 2)), LocalName (CHAR(45)), GovernmentForm (CHAR(45)), HeadOfState (CHAR(60)), Capital (INTEGER), Code2 (CHAR(2)), and foreign keys: . The table description is: country Here are some relevant example rows (values in the same order as columns above)
('IND', 'India', 'Asia', 'Southern and Central Asia', 3287263, 1947, 1013662000, 62.5, 447114, 430572, 'Bharat/India', 'Federal Republic', 'Kocheril Raman Narayanan', 1109, 'IN')
('DEU', 'Germany', 'Europe', 'Western Europe', 357022, 1955, 82164700, 77.4, 2133367, 2102826, 'Deutschland', 'Federal Republic', 'Johannes Rau', 3068, 'DE')

Table 'countrylanguage' has columns: CountryCode (CHAR(3)), Language (CHAR(30)), IsOfficial (TEXT), Percentage (DECIMAL(4, 1)), and foreign keys: ['CountryCode'] -> country,['Code']. The table description is: countrylanguage Here are some relevant example rows (values in the same order as columns above)
('IND', 'Hindi', 'T', 39.9)
('USA', 'German', 'F', 0.7)

With Context Retriever

Methodology

Context Retriever

 How many cities are in Russia?

 There are 0 cities in Russia.

View Query

```
SELECT COUNT(*) FROM city WHERE countrycode = "RU"
```

 How many cities are in India and Germany?

 Sorry, could not retrieve the answer. Please rephrase your question more accurately.

View Query

```
SELECT COUNT(DISTINCT city."Name") FROM city WHERE city."CountryCode" IN (SELECT
country."Code" FROM country WHERE country."Continent" = 'Asia' AND country."Region" = 'Southern
and Central Asia') AND city."CountryCode" IN (SELECT country."Code" FROM country WHERE
```

Without Context Retriever

 How many cities are in Russia?

 189 cities are in Russia.

View Query

```
SELECT COUNT(Name) FROM city WHERE CountryCode = 'RUS'
```

 How many cities are in India and Germany?

 434 cities are in India and Germany.

View Query

```
SELECT COUNT(DISTINCT Name) FROM city WHERE CountryCode IN (SELECT Code FROM country
WHERE Name IN ('India', 'Germany'))
```

With Context Retriever



Methodology

Refiner

- To detect and fix the SQLite errors
- In case of error, initiates regeneration of SQL using refiner prompt template additionally consisting:
 - Special instruction and constraints to fix the error
 - SQLite error evidence and exception class

Algorithm 1 Refiner

Input: generated sql query gsql, input question q

Output: refined sql query rsq

```
1: count = 0
2: Max Retry = 4
3: while count ≤ MaxRetry do
4:   isRefineRequired = execute(gsql)
5:   if isRefineRequired then
6:     gsql ← regenerate(refinerTemplate(q, gsql))
7:     count += 1
8:   else
9:     rsq ← gsql
10:    break
11:  end if
12: end while
```

```
refiner_template = f"""
### Instruction:
When executing SQL below, some errors occurred, please fix up SQL based on query and database info.
Solve the task step by step if you need to. Using SQL format in the code block, and indicate script type in the code block.
When you find an answer, verify the answer carefully. Include verifiable evidence in your response if possible.
### Constraints:
- In `SELECT <column>`, just select needed columns in the [Question] without any unnecessary column or value
- In `FROM <table>` or `JOIN <table>`, do not include unnecessary table
- If use max or min func, `JOIN <table>` FIRST, THEN use `SELECT MAX(<column>)` or `SELECT MIN(<column>)`
- If [Value examples] of <column> has 'None' or None, use `JOIN <table>` or `WHERE <column> is NOT NULL` is better
- If use `ORDER BY <column> ASC|DESC`, add `GROUP BY <column>` before to select distinct values
### Query:
-- {query}
### Evidence:
{evidence}
### Database info:
{desc_str}
### old SQL:
{sql}
### SQLite error:
{sqlite_error}
### Exception class:
{exception_class}

Now please fixup old SQL and generate new SQL again.
### correct SQL:
"""


```

Refiner instruction prompt template

Refiner

Refiner can fix errors that are caught by sql executor like syntactic sql or column name error

 How many cities are in Russia?

 189 cities are in Russia.

```
SELECT COUNT(Name) FROM city WHERE CountryCode = 'RUS'
```

```
User Question: How many cities are in Russia?
Selected Tables: ['city', 'country', 'countrylanguage']
Old Generated SQL Query: SELECT COUNT(Name) FROM city WHERE code = 'Russia'
Generated SQL Query: SELECT COUNT(Name) FROM city WHERE CountryCode = 'RUS'
SQL Result: [(189,)]
SQL Error before refining: no such column: code
SQL Exception before refining: <class 'sqlite3.OperationalError'>
Answer: 189 cities are in Russia.
Previous conversation : []
Is refined: True
Refined queries: ["SELECT COUNT(Name) FROM city WHERE CountryCode = 'RUS'"]
```

Refiner does not fix semantic errors

 List the name and population of the cities from country with language same as Russian Federation?

 Sorry, could not retrieve the answer. Please rephrase your question more accurately.

```
SELECT Name, Population FROM City WHERE CountryCode IN (SELECT CountryCode FROM CountryLanguage WHERE Language = 'Russian Federation')
```

```
User Question: List the name and population of the cities from country with la
Selected Tables: ['countrylanguage', 'city', 'country']
Generated SQL Query: SELECT Name, Population FROM City WHERE CountryCode IN (S
SQL Result: []
SQL Error before refining:
SQL Exception before refining:
Answer: Sorry, could not retrieve the answer. Please rephrase your question mor
Previous conversation : [HumanMessage(content='List the name and population of
Is refined: False
Refined queries: []]
```

Rephraser

- To present the data in a manner that's easily understandable for users
- Rephrasing SQL results into natural language answers
- Answer prompt with few-shot examples guides the model in rephrasing

 What is the surface area of India?

 The surface area of India is 3,287,263 square kilometers.

 View Query

 View Logs

```
User Question: What is the surface area of India?
Selected Tables: ['country', 'countrylanguage', 'city']
Generated SQL Query: SELECT SurfaceArea FROM country WHERE Name = 'India'
SQL Result: [(3287263,)]
Answer: The surface area of India is 3,287,263 square kilometers.
Previous conversation : [HumanMessage(content='What is the population of the c
Is refined: False
Refined queries: []]
```

Rephraser Illustration

answer_prompt = `f'''Given the user question, corresponding SQL query, and SQL result, answer the user question.`

Here is a typical example:

Question: List name and population of the 5 cities in country with Italian language?
SQL Query: `SELECT Name, Population FROM city WHERE CountryCode IN (SELECT Code FROM country WHERE Name = 'Italy') ORDER BY Population DESC LIMIT 5`
SQL Result: `[('Roma', 2643581), ('Milano', 1300977), ('Napoli', 1002619), ('Torino', 903705), ('Palermo', 683794)]`
Answer: Here's the list 5 cities in country with Italian language

Name,	Population
1. Roma,	2643581
2. Milano,	1300977
3. Napoli,	1002619
4. Torino,	903705
5. Palermo,	683794

Here is another typical example:

Question: What percentage of population speaks Kannada in the country Bangalore?
SQL Query: `SELECT Percentage FROM countrylanguage WHERE Language = "Kannada" AND CountryCode IN (SELECT CountryCode FROM city WHERE Name = "Bangalore")`
SQL Result: `[(3.9,)]`
Answer: 3.9% of the population speaks Kannada in the country Bangalore.

Here is a new example, please start answering:

Question: `{exec_result['question']}`
SQL Query: `{exec_result['sql']}`
SQL Result: `{exec_result['data']}`
Answer: '''

Rephraser instruction prompt template

Methodology

Rephraser



List the name and population of 6 cities from country with language Russian?



The cities with the largest population are Bakı, Gənəcə, Sumqayıt, Mingəçevir, Minsk, and Gomel.

View Query

```
SELECT Name, Population FROM City WHERE CountryCode IN (SELECT CountryCode FROM
Countrylanguage WHERE Language = 'Russian') LIMIT 6
```

View Logs

```
User Question: List the name and population of 6 cities from country with lang
Selected Tables: ['countrylanguage', 'city', 'country']
Generated SQL Query: SELECT Name, Population FROM City WHERE CountryCode IN (
SQL Result: [('Bakı', 1787800), ('Gənəcə', 299300), ('Sumqayıt', 283000), ('Min
SQL Error before refining:
SQL Exception before refining:
Answer: The cities with the largest population are Bakı, Gənəcə, Sumqayıt, Min
Previous conversation : [HumanMessage(content='List the name and population of
Is refined: False
Refined queries: []]
```



Here's the list of 6 cities from country with language Russian

1. Bakı, 1787800
2. Gənəcə, 299300
3. Sumqayıt, 283000
4. Mingəçevir, 93900
5. Minsk,

View Query

```
SELECT Name, Population FROM City WHERE CountryCode IN (SELECT CountryCode FROM
Countrylanguage WHERE Language = 'Russian') LIMIT 6
```

View Logs

```
User Question: List the name and population of 6 cities from country with lang
Selected Tables: ['countrylanguage', 'city', 'country']
Generated SQL Query: SELECT Name, Population FROM City WHERE CountryCode IN (
SQL Result: [('Bakı', 1787800), ('Gənəcə', 299300), ('Sumqayıt', 283000), ('Min
SQL Error before refining:
SQL Exception before refining:
Answer: Here's the list of 6 cities from country with language Russian
1. Bakı, 1787800
2. Gənəcə, 299300
3. Sumqayıt, 283000
4. Mingəçevir, 93900
5. Minsk,
Previous conversation : [HumanMessage(content='List the name and population of
Is refined: False
Refined queries: []]
```

Without few-shot

With few-shot

Memory Integration

- Enables model to handle follow-up questions by remembering the context of the conversation
- Maintain ChatMessageHistory list of no more than n messages
- Include ChatMessageHistory as previous conversation in instruction prompt

 How many cities are in Germany?

 93 cities are in Germany.

View Query

```
SELECT COUNT(Name) FROM city WHERE CountryCode = 'DEU'
```

View Logs

 Which city is the least populated amongst these?

 Adamstown is the least populated city.

View Query

```
SELECT Name FROM city WHERE Population = (SELECT MIN(Population) FROM city)
```

Without previous conversation history

```
### Instruction:  
You are a powerful text-to-SQL model. Your job is to answer  
given a question and context regarding one or more tables.  
  
You must output the SQL query that answers the question.  
### Dialect:  
{dialect}  
  
### Input:  
{input}  
  
### Context:  
{context}  
  
### previous conversation:  
{history}  
  
### Response:
```

Inference Instruction prompt template with history



Results

Natural Language SQL Translator with following features and capabilities:

- Includes query-relevant context in instruction prompt
- Can answer follow-up questions using memory
- Refines SQL execution errors
- Rephrases answers for enhanced clarity

 How many cities are in the USA?

 274 cities are in the USA.

 which city is the least populated among these?

 Charleston is the least populated city among these.

 which city is the most populated among these?

 New York is the most populated city among these.

Illustration 1

 List name and districts of 5 cities in the USA

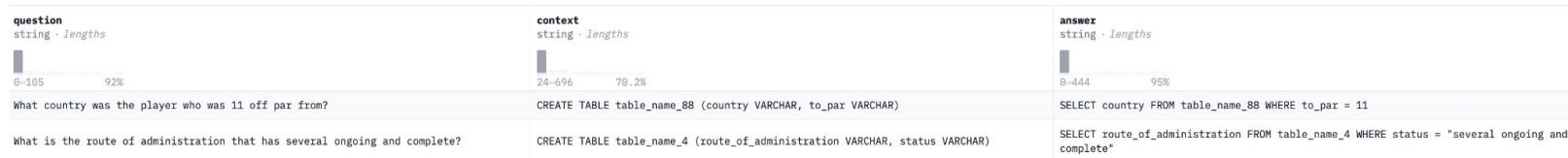
 Here's the list of 5 cities in the USA

1. New York, New York
2. Los Angeles, California
3. Chicago, Illinois
4. Houston, Texas
5. Philadelphia, Pennsylvania

Illustration 2

Datasets for evaluation

- Without query complexity labels
 - [hoangphu7122002ai/text2sql_en](#) - evaluated on 1000 records



- With query complexity - basic SQL, aggregation, single join, etc
 - [gretelai/synthetic_text_to_sql](#) - evaluated on 1000 records

id int32	domain string · classes	domain_description string · classes	sql_complexity string · classes	sql_complexity_description string · classes	sql_task_type string · classes	sql_task_type_description string · classes	sql_prompt string · lengths	sql_context string · lengths	sql string · lengths
1	100k 100 values	100 values	8 values	8 values	4 values	4 values	10 438	11 2.05k	16 761
5,097	forestry	Comprehensive data on sustainable...	single join	only one join (specify inner, outer, cross)	analytics and reporting	generating reports, dashboards, and...	What is the total volume of timber sold by each..	CREATE TABLE salesperson (salesperson_id INT, name TEXT,...	SELECT salesperson_id, name, SUM(volum...
5,098	defense industry	Defense contract data, military...	aggregation	aggregation functions (COUNT, SUM, AVG, MIN,..)	analytics and reporting	generating reports, dashboards, and...	List all the unique equipment types and...	CREATE TABLE equipment_maintenance (equ...	SELECT equipment_type, SUM(maintenanc...
5,099	marine biology	Comprehensive data on marine species...	basic SQL	basic SQL with a simple select statement	analytics and reporting	generating reports, dashboards, and...	How many marine species are found in the...	CREATE TABLE marine_species (name VARCHAR(50), common_name..	SELECT COUNT(*) FROM marine_species W...

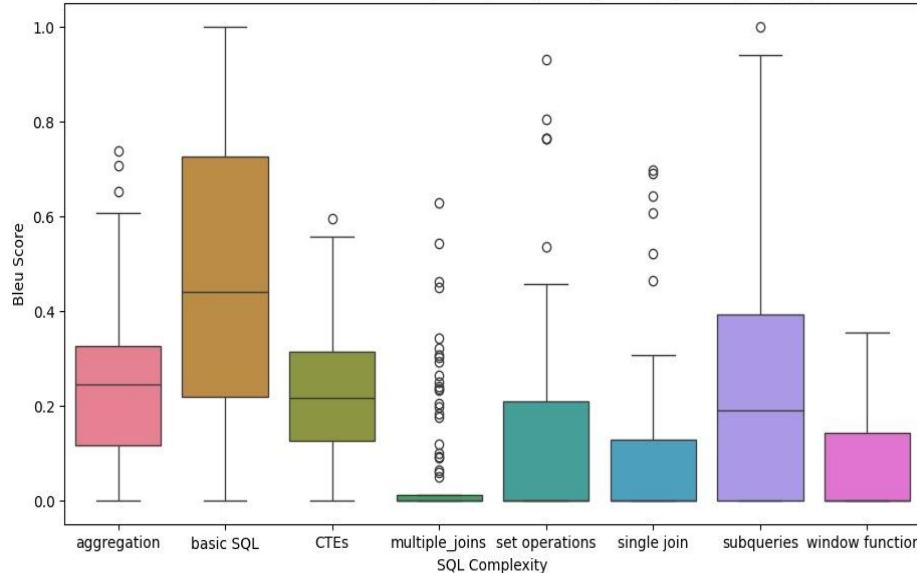
Evaluation Metrics:

1. Bleu Score (Bilingual Evaluation Understudy)
2. Rouge Score (Recall-Oriented Understudy for Gisting Evaluation)

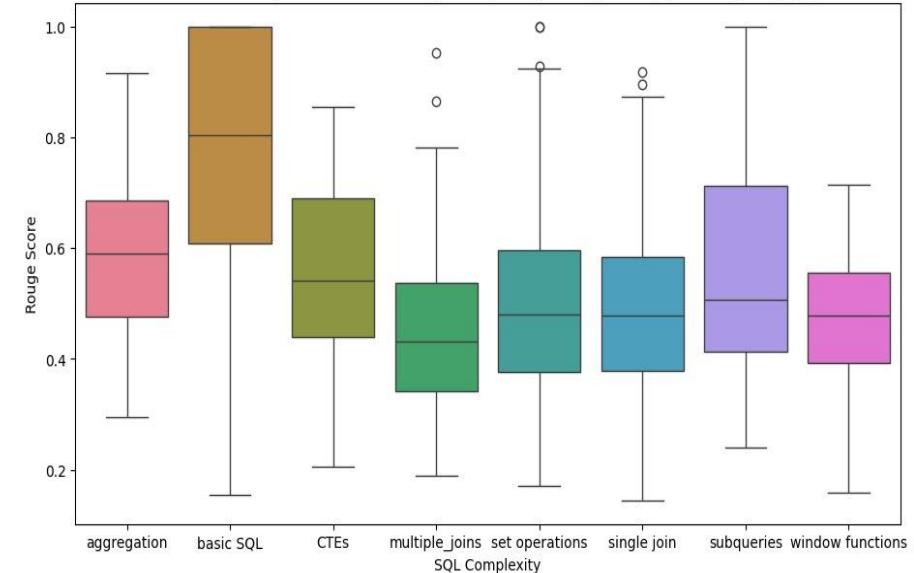
Both metrics rely on n-gram overlap to measure similarity between the model-generated output and the reference summaries or ground truth.

Evaluation

Box Plot of Bleu Score for Each SQL Complexity - For gretelai/synthetic_text_to_sql



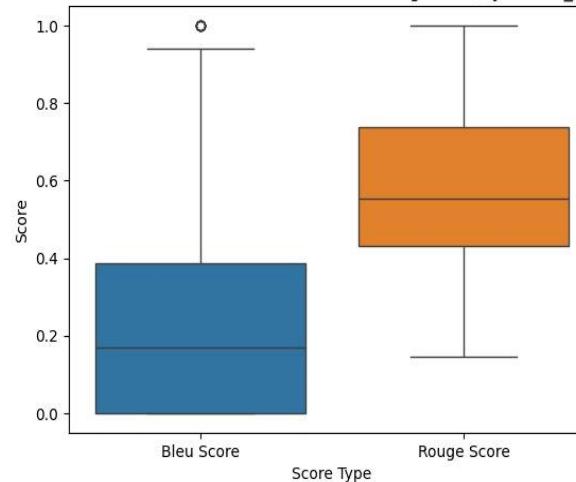
Box Plot of Rouge Score for Each SQL Complexity - For gretelai/synthetic_text_to_sql



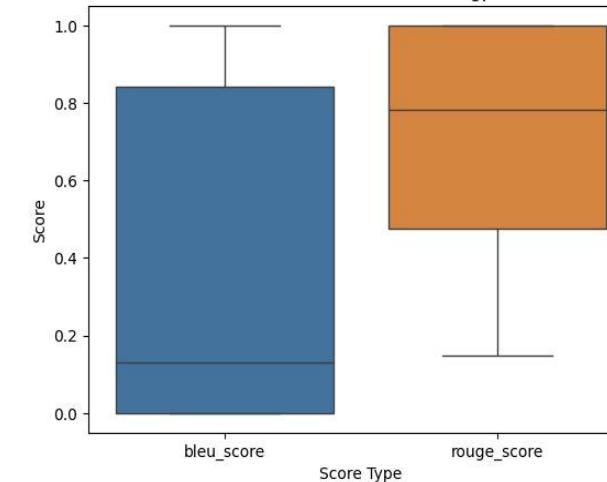
Evaluation

Dataset	No. of Items	Mean Bleu Score	Mean Rouge Score
gretelai/synthetic text-to-sql dataset (With query complexities)	1000	0.448	0.763
hoangphu7122002ai/text2sql-en (Without query complexities)	1000	0.385	0.729

Box Plot of BLEU Score and ROGUE Score - For gretelai/synthetic_text_to_sql



Box Plot of BLEU Score and ROGUE Score - For hoangphu7122002ai/text2sql_en



Experiments

1. Follow-up query :

Explored the ability to address follow-up questions following the generated SQL query.

 How many cities are in the USA?

 274 cities are in the USA.

 which city is the least populated among these?

 Charleston is the least populated city among these.

 which city is the most populated among these?

 New York is the most populated city among these.

2. Query Complexity :

Evaluated the model's proficiency in generating complex queries, including joins and aggregation functions.

 Which is the most spoken language of the largest country in Asian continent by surface area?

 Chinese

 View Query

```
SELECT Language FROM countrylanguage WHERE CountryCode IN (SELECT Code FROM country  
WHERE Continent = 'Asia' ORDER BY SurfaceArea DESC LIMIT 1)
```



Experiments

3. Poorly phrased question :

Designed ambiguous questions, incomplete questions, questions with grammatical errors to evaluate model's robustness

⌚ what is Bangalore

⌚ Sorry, could not retrieve the answer. Please rephrase your question more accurately.

⌚ what is Bangalore's population?

⌚ Bangalore's population is 2,660,088.

View Query

```
SELECT Population FROM city WHERE Name = "Bangalore"
```

4. Refinement :

Conducted experiments with refiner to detect and fix erroneous SQL queries.

Question : Who is the Head of State of Germany?

Generated SQL Query: SELECT HeadOfState FROM country WHERE Name =
SQL Result: <class'sqlite3.OperationalError>

Previous conversation: [HumanMessage(content='How many cities are in the USA?'),
AIMessage(content='SELECT COUNT(*) FROM city WHERE countrycode = "USA"')]

Is refined: True

Refined queries: ["SELECT HeadOfState FROM country WHERE Name = 'Germany'"]

SQL Result: [(Johannes Rau,)]

Answer: Johannes Rau is the Head of State of Germany.

Experiments

5. Relevant table extraction :

Investigated the effectiveness of the fine-tuned text-to-SQL model in identifying the most relevant table for a given query.

 How many cities are in India and Germany?

 434 cities are in India and Germany.

View Query

```
SELECT COUNT(*) FROM city WHERE CountryCode IN ('IND', 'DEU')
```

View Logs

User Question: How many cities are in India and Germany?

Selected Tables: ['city']

Generated SQL Query: SELECT COUNT(*) FROM city WHERE CountryCode IN ('IND', 'DEU')
SQL Result: [(434,)]

Answer: 434 cities are in India and Germany.

Previous conversation : [HumanMessage(content='How many cities are in India and Germany?'), AIMessage(content='434 cities are in India and Germany.'), HumanMessage(content='Is this refined?'), AIMessage(content='Yes, it is refined.'), HumanMessage(content='Refined queries: []')]

 Amsterdam lies in which country?

 Amsterdam lies in the Netherlands.

View Query

```
SELECT Name FROM city WHERE ID IN (SELECT Capital FROM country WHERE Name = 'Netherlands')
```

View Logs

User Question: Amsterdam lies in which country?

Selected Tables: ['city', 'country']

Generated SQL Query: SELECT Name FROM city WHERE ID IN (SELECT Capital FROM country WHERE Name = 'Netherlands')
SQL Result: [('Amsterdam',)]

Answer: Amsterdam lies in the Netherlands.

Previous conversation : [HumanMessage(content='Is Amsterdam in the Netherlands?'), AIMessage(content='Yes, it is in the Netherlands.'), HumanMessage(content='Is this refined?'), AIMessage(content='Yes, it is refined.'), HumanMessage(content='Refined queries: []')]

Challenges:

- Limited computational resources
- Indefinite/irrelevant generation due to poorly phrased question, incorrect prompt template, example rows included in context for small models
- Langchain CustomLLM class does not work well open source models
- No benchmark dataset for SQL complexity evaluation
- Exact match evaluation is not possible since evaluation datasets include ground truth query with alias table names

Limitations:

- Less capable of handling too complex queries
- Sometimes the model fails to efficiently utilize the conversation history to identify the root context for follow up question
- Handling poorly phrased questions



Conclusion

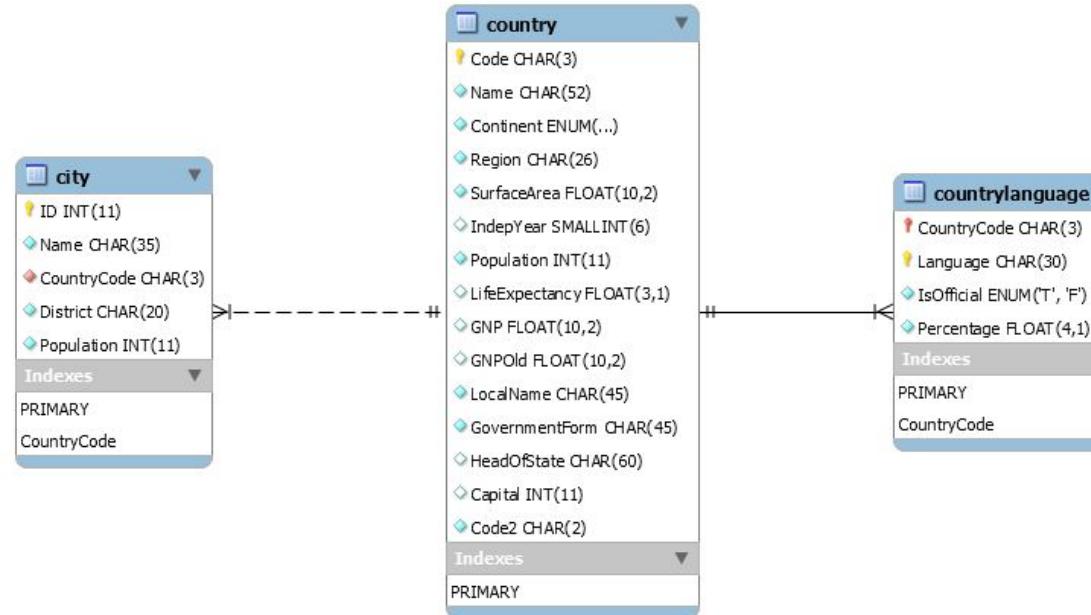
- Multi-module collaboration employing fine-tuned Llama-3-8b model with a collaborative architecture for robust Text-to-SQL translation.
- Utilization of Retriever-Augmenter-Generator (RAG) based context retriever, refiner, and dynamic memory.
- Demonstrates good performance in handling simple queries and immediate follow-up questions.
- Unified approach utilizing RAG, embedding context effectively for accurate SQL query generation.

Future Work

- Training the model to handle queries involving aliases, joins, and aggregate statements.
- Exploration of additional features or techniques to improve the model's understanding of the relationships between queries and database schema.
- Explore larger models and advanced training techniques to improve performance and scalability.
- Investigate methods to mitigate biases in the training data and improve the model's generalization capabilities for diverse query types and domains.



Demo



References

- [1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. arXiv preprint arXiv:2303.08774 (2023).
- [2] Zefeng Cai, Xiangyu Li, Binyuan Hui, Min Yang, Bowen Li, Binhu Li, Zheng Cao, Weijie Li, Fei Huang, Luo Si, and Yongbin Li. 2022. STAR: SQL Guided Pre-Training for Context-dependent Text-to-SQL Parsing. arXiv:2210.11888 [cs.CL]
- [3] Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2023. QLoRA: Efficient Finetuning of Quantized LLMs. arXiv:2305.14314 [cs.LG]
- [4] Prajpta Devgade. 2021. The Natural Language Query using NLP by Generation of SQL Query. International Journal for Research in Applied Science and Engineering Technology 9 (05 2021), 381–383.
<https://doi.org/10.22214/ijraset.2021.34145>
- [5] Chunxi Guo, Zhiliang Tian, Jintao Tang, Shasha Li, Zhihua Wen, Kaixuan Wang, and Ting Wang. 2024. Retrieval-Augmented GPT-3.5 Based Text-to-SQL Framework with Sample-Aware Prompting and Dynamic Revision Chain. In Neural Information Processing, Biao Luo, Long Cheng, Zheng-Guang Wu, Hongyi Li, and Chaojie Li (Eds.). Springer Nature Singapore, Singapore, 341–356.
- [6] Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin de Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. 2019. Parameter-Efficient Transfer Learning for NLP. arXiv:1902.00751 [cs.LG]
- [7] Jeremy Howard and Sebastian Ruder. 2018. Universal Language Model Fine-tuning for Text Classification. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Iryna Gurevych and Yusuke Miyao (Eds.).



OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

FAKULTÄT FÜR
INFORMATIK

Thank You!