Natural Language to SQL:

Analyzing Netflix Movies with LLMs

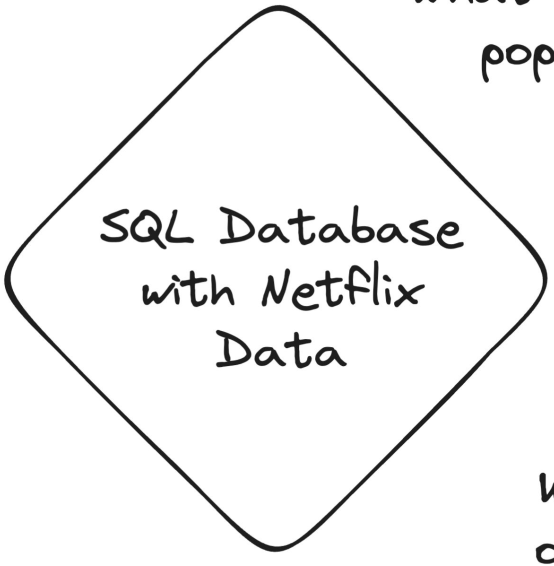
Analytics Vidhya
Data Hour Webinar by Seb Duerr

Wednesday, 7 February 2024 from 8:00 PM - 9:00 PM IST /
6:30 AM-7:30 AM Pacific Time (PT)

Agenda

1. **Approaching Ambiguity**
2. **State of the Art: SQL with LLMs**
3. **Implementation**

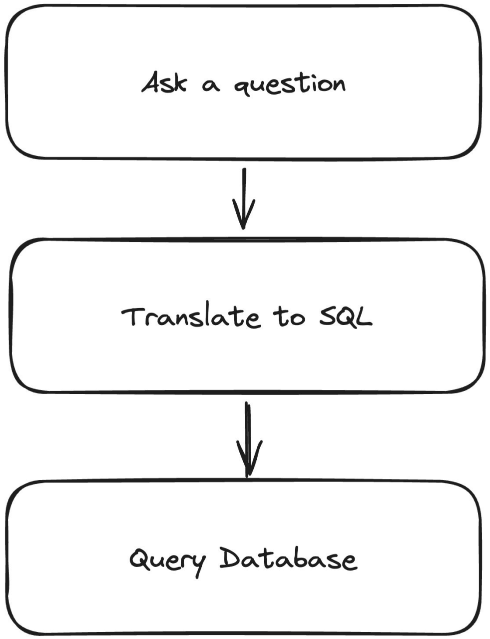
The Requirements



What was the most popular movie?

What is the oldest movie?

The Approach



Getting Ideas on Implementing A Human Language to SQL Translator

🔥 Check relevant blogs & docs: LangChain, Hugging Face, and OpenAI.

📑 Review Articles on medium.com (Analytics Vidhya) or Google Scholar.

👨‍💻 Look for Example Code on Github (look for similar dependencies).

Agenda

1. How to Research & Build Features
2. **State of the Art: SQL with LLMs**
3. Implementation

2 | State of the Art: SQL with LLMs

SPIDER, DAIL & GPT-4

What is Spider?

Spider is a large-scale *complex and cross-domain* semantic parsing and text-to-SQL dataset annotated by 11 Yale students. The goal of the Spider challenge is to develop natural language interfaces to cross-domain databases. It consists of 10,181 questions and 5,693 unique complex SQL queries on 200 databases with multiple tables covering 138 different domains. In Spider 1.0, different complex SQL queries and databases appear in train and test sets. To do well on it, systems must *generalize well to not only new SQL queries but also new database schemas.*

Why we call it "Spider"? It is because our dataset is complex and cross-domain like a spider crawling across mutiple complex(with many foreign keys) nests(databases).

XLang Lab for building LM agents!

Spider Paper (EMNLP'18)

Spider Post

Leaderboard - Execution w

Our current models do not predict any v execution accuracies. However, we enc value prediction, your model should be a from the database content (database co "LIMIT 3"). *Notice:* Test results after May some annotation errors).

Rank		
1 Nov 2, 2023	MiniSeek *Anonymous* Code and paper coming soon	91.2
1 Aug 20, 2023	DAIL-SQL + GPT-4 + Self-Consistency *Alibaba Group* (Gao and Wang et al.,'2023) code	86.6
2 Aug 9, 2023	DAIL-SQL + GPT-4 *Alibaba Group* (Gao and Wan	86.2
3 October 17, 2023	DPG-SQL + GP An Code and p	

Text-to-SQL Empowered by Large Language Models: A Benchmark Evaluation

Dawei Gao* Alibaba Group gaodawei.gdw@alibaba-inc.com
Haibin Wang* Alibaba Group binke.whb@alibaba-inc.com
Yaliang Li Alibaba Group yaliang.li@alibaba-inc.com
Xiuyu Sun Alibaba Group xiuyu.sxy@alibaba-inc.com
Yichen Qian Alibaba Group yichen.qyc@alibaba-inc.com
Bolin Ding Alibaba Group bolin.ding@alibaba-inc.com
Jingren Zhou Alibaba Group jingren.zhou@alibaba-inc.com

20 Nov 2023

ABSTRACT

Large language models (LLMs) have emerged as a new paradigm for Text-to-SQL task. However, the absence of a systematical benchmark inhibits the development of designing effective, efficient and economic LLM-based Text-to-SQL solutions. To address this challenge, in this paper, we first conduct a systematical and extensive comparison over existing prompt engineering methods, including

lacks a systematic study for prompt engineering in LLM-base to-SQL solutions. Specifically, for question representation, m isting research textualize structured knowledge as schema, a ther add task instructions and foreign keys to form prompts [Besides, some studies [7, 29] represent tables as several "CI TABLE" SQL statements, and prompt LLMs to answer the question in comments. However, even with similar represen

Spider 1.0

Yale Semantic Parsing and Text-to-SQL Challenge

Gao et al. 2023

🍬 Sweet spot: 5-Shot Learning

Few-shot	Selection	Question Similarity	Query Similarity	GPT-4		GPT-3.5-TURBO		TEXT-DAVINCI-003		Vicuna-33B	
				EM	EX	EM	EX	EM	EX	EM	EX
0-shot	-	-	-	22.1	72.3	34.6	74.4	31.7	71.7	6.9	43.7
1-shot	Random	0.23	0.47	41.7	77.4	45.9	73.9	38.2	70.6	14.4	47.9
	Question Similarity selection	0.39	0.65	53.3	78.8	51.9	74.3	44.1	72.3	16.5	48.5
	Masked Question Similarity selection	0.57	0.80	58.2	79.1	57.4	76.0	47.9	75.0	21.4	48.7
	DAIL selection	0.56	0.95	62.1	80.2	59.5	75.5	51.9	76.9	22.8	49.2
	Upper Limit	0.56	0.98	63.7	81.0	61.4	77.2	53.1	77.5	22.7	49.4
3-shot	Random	0.23	0.48	48.9	79.4	49.0	73.6	41.7	71.6	16.8	46.9
	Question Similarity selection	0.37	0.63	56.3	79.2	53.8	74.7	52.2	74.1	21.1	47.1
	Masked Question Similarity selection	0.54	0.78	66.1	81.5	61.1	77.3	59.7	77.0	27.7	52.3
	DAIL selection	0.53	0.94	69.1	81.7	63.9	77.8	64.4	79.5	30.7	53.6
	Upper Limit	0.53	0.98	71.5	83.4	66.2	79.2	66.7	81.1	31.2	54.4
5-shot	Random	0.23	0.48	51.6	79.5	52.9	75.7	49.0	72.1	-	-
	Question Similarity selection	0.36	0.61	58.2	79.9	55.9	75.1	54.8	73.2	-	-
	Masked Question Similarity selection	0.52	0.77	66.8	82.0	62.3	77.9	64.7	78.6	-	-
	DAIL selection	0.52	0.94	71.9	82.4	66.7	78.1	67.7	80.5	-	-
	Upper Limit	0.51	0.97	74.4	84.4	68.8	79.6	70.7	82.4	-	-

Gao et al. 2023

✋ Not for Now: Open Source Models

LLM	BS $_P$		TR $_P$		OD $_P$		CR $_P$		AS $_P$		Average	
	EM	EX	EM	EX	EM	EX	EM	EX	EM	EX	EM	EX
LLaMA-7B	6.5	9.6	3.1	4.9	3.6	9.0	4.8	16.3	1.3	5.9	3.9	9.1
LLaMA-13B	8.8	18.4	4.5	15.2	8.2	21.8	5.6	25.0	8.9	26.9	7.2	21.5
LLaMA-33B	9.6	26.7	12.0	25.9	13.6	36.4	12.2	**42.8**	**13.8**	38.1	12.2	34.0
Falcon-40B	0.3	11.7	0.2	0.9	0.3	7.6	0.1	21.9	0.0	5.0	0.2	9.4
Alpaca-7B	15.1	25.1	13.5	23.8	14.7	25.7	16.0	32.1	8.9	19.9	13.6	25.3
GPT4ALL-7B	7.8	19.4	8.8	24.6	8.1	27.0	8.5	25.9	6.5	21.8	7.9	23.7
Vicuna-7B	7.5	15.6	1.2	9.9	6.2	21.5	5.6	24.0	0.9	5.4	4.3	15.3
Vicuna-13B	8.2	21.7	10.1	24.4	11.2	31.4	5.8	33.5	4.7	20.0	8.0	26.2
Vicuna-33B	10.8	28.9	18.3	37.1	19.1	42.7	6.9	43.7	8.6	30.6	12.7	36.6
LLaMA-2-CHAT-7B	14.3	23.4	7.2	15.5	6.3	12.3	12.2	25.5	5.0	20.5	9.0	19.4
LLaMA-2-CHAT-13B	18.8	32.6	15.4	30.5	11.1	22.3	20.7	40.0	16.9	36.2	16.6	32.3
LLaMA-2-CHAT-70B	21.8	46.2	11.9	33.9	21.4	45.5	12.4	44.0	8.4	28.6	15.2	39.6
CodeLLaMA-34B	**27.8**	65.5	15.9	40.3	25.8	65.3	24.3	**68.5**	22.4	61.5	23.2	60.2

Gao et al. 2023

👌 Striving for Prompt Perfection

```
/* Given the following database schema: */
${DATABASE_SCHEMA}
/* Answer the following: How many authors are there? */
SELECT count(*) FROM authors

/* Given the following database schema: */
${DATABASE_SCHEMA}
/* Answer the following: How many farms are there? */
SELECT count(*) FROM farm

${TARGET_QUESTION}
```

Listing 6: Example of Full-Information Organization.

```
/* Some SQL examples are provided based on similar
↳ problems: */
SELECT count(*) FROM authors

SELECT count(*) FROM farm

${TARGET_QUESTION}
```

Listing 7: Example of SQL-Only Organization.

```
/* Some example questions and corresponding SQL queries
↳ are provided based on similar problems: */
/* Answer the following: How many authors are there? */
SELECT count(*) FROM authors

/* Answer the following: How many farms are there?. */
SELECT count(*) FROM farm

${TARGET_QUESTION}
```

Listing 8: Example of DAIL Organization.

Gao et al. 2023

Prompt Principles for Instructions

#	Principle
2	Integrate the intended audience in the prompt.
5	When you need clarity or a deeper understanding of a topic, idea, or any piece of information, utilize the following prompts: - ***"Explain [insert topic] in simple terms."*** - ***"Explain to me like I'm 11 years old."*** - ***"Explain to me as if I'm a beginner in the field."*** - ***"Write the [essay/text] using simple English [...]"***
14	Allow the model to elicit precise details and requirements from you by asking you questions until it has enough information. ***"From now on, I would like you to ask me question to …"***
26	To write any texted intended to be similar to a provided sample, include specific instructions: ***"Please use the same language based on the provided paragraph."***

Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4

Sondos Mahmoud Bsharat*, Aidar Myrzakhan*, Zhiqiang Shen*
*joint first author & equal contribution
VILA Lab, Mohamed bin Zayed University of AI

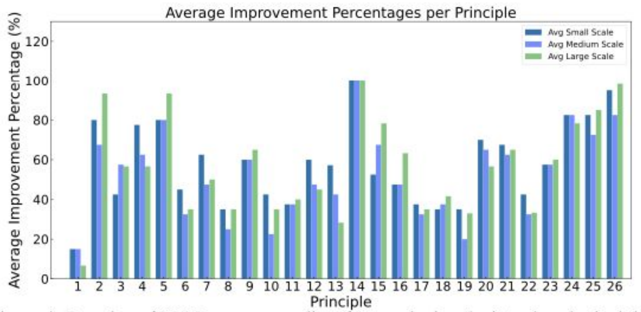
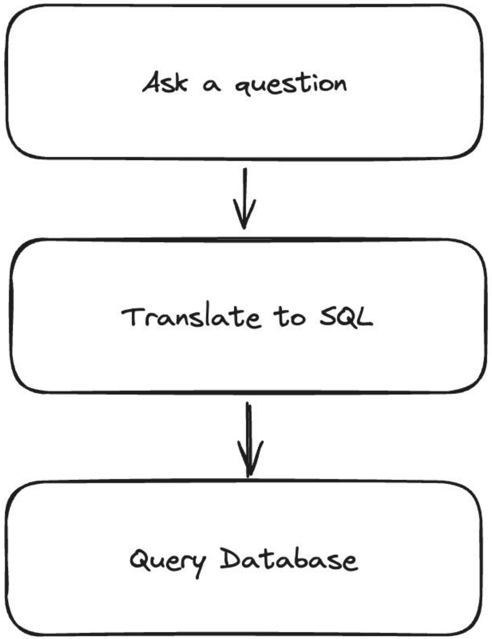


Figure 4: Boosting of LLM response quality after employing the introduced principles on prompts. *small-scale* indicates the 7B models, *medium-scale* indicates the 13B models and *large-scale* indicates the 70B and GPT-3.5/4 models.

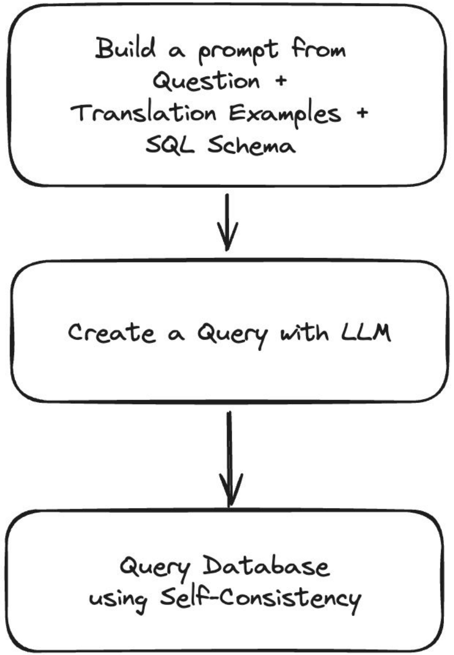
Bsharat et al. 2023

2 | State of the Art: SQL with LLMs

The Approach

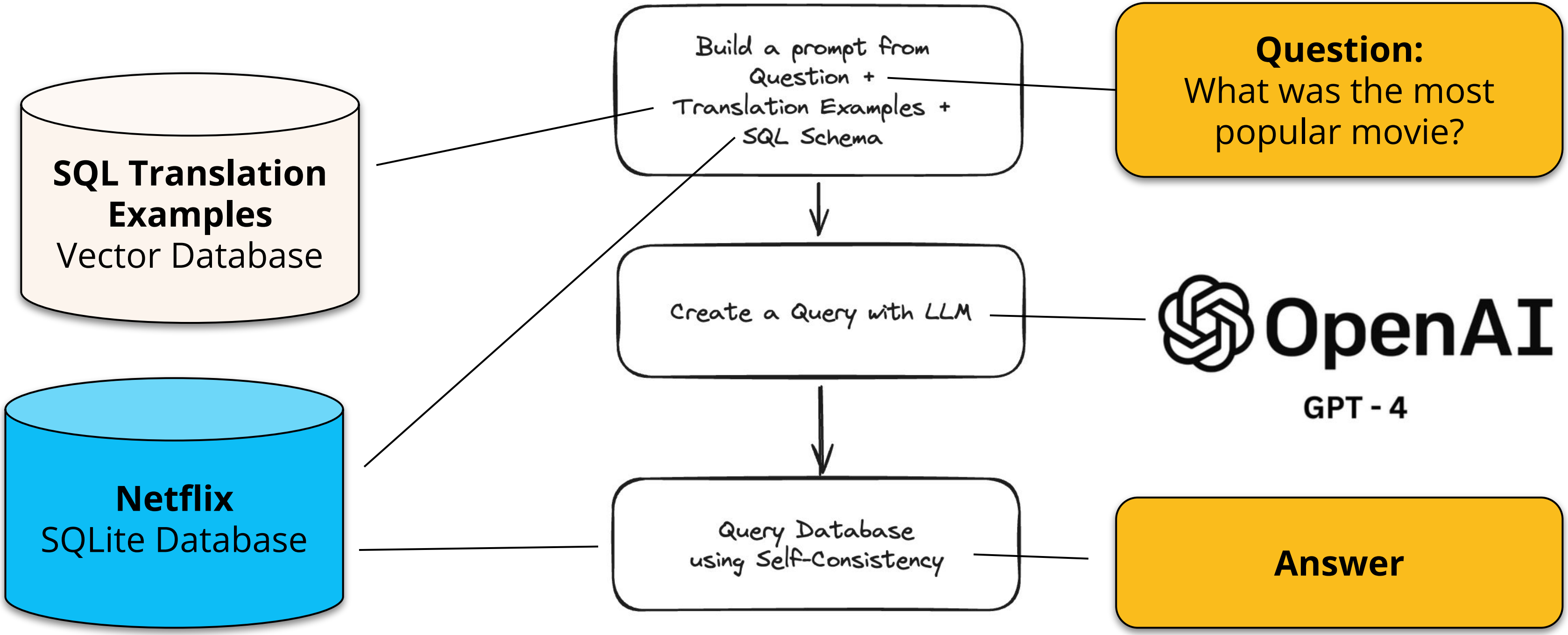


The Updated Approach



1) Ingestion

2) Inference

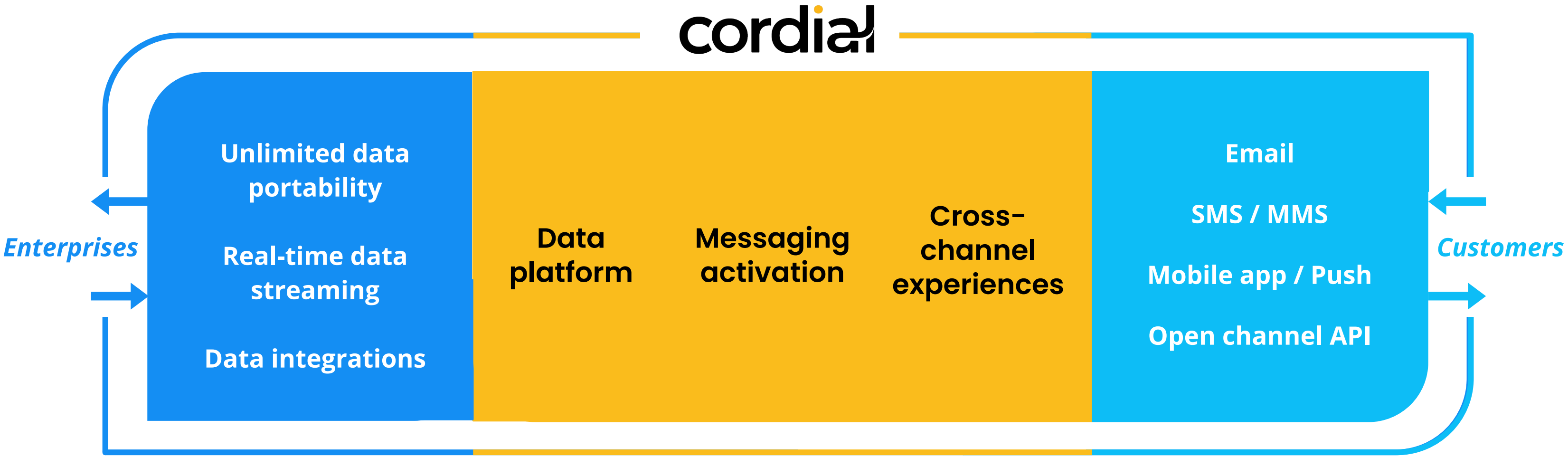


Agenda

1. **How to Research & Build Features**
2. **State of the Art: SQL with LLMs**
3. **Implementation**

📕 [Jupyter Notebook]

Cordial lets brands use all of their data to create personal, unified experiences across channels.



Leading brands trust Cordial.

PACSUN	realtor.com®	L.L.Bean	edible	Orveon
REVOLVE	Forbes	Virgin VOYAGES	JOANN	TILLYS
Abercrombie & Fitch	purple	Backcountry	ADORE ME	BOOT BARN®

Sebastian (Seb) Duerr

linkedin.com/in/sebastianduerr

Thank you.