✦ Member-only story

SQLCoder-70b Becomes the Leading AI SQL Model

Meng Li · Follow
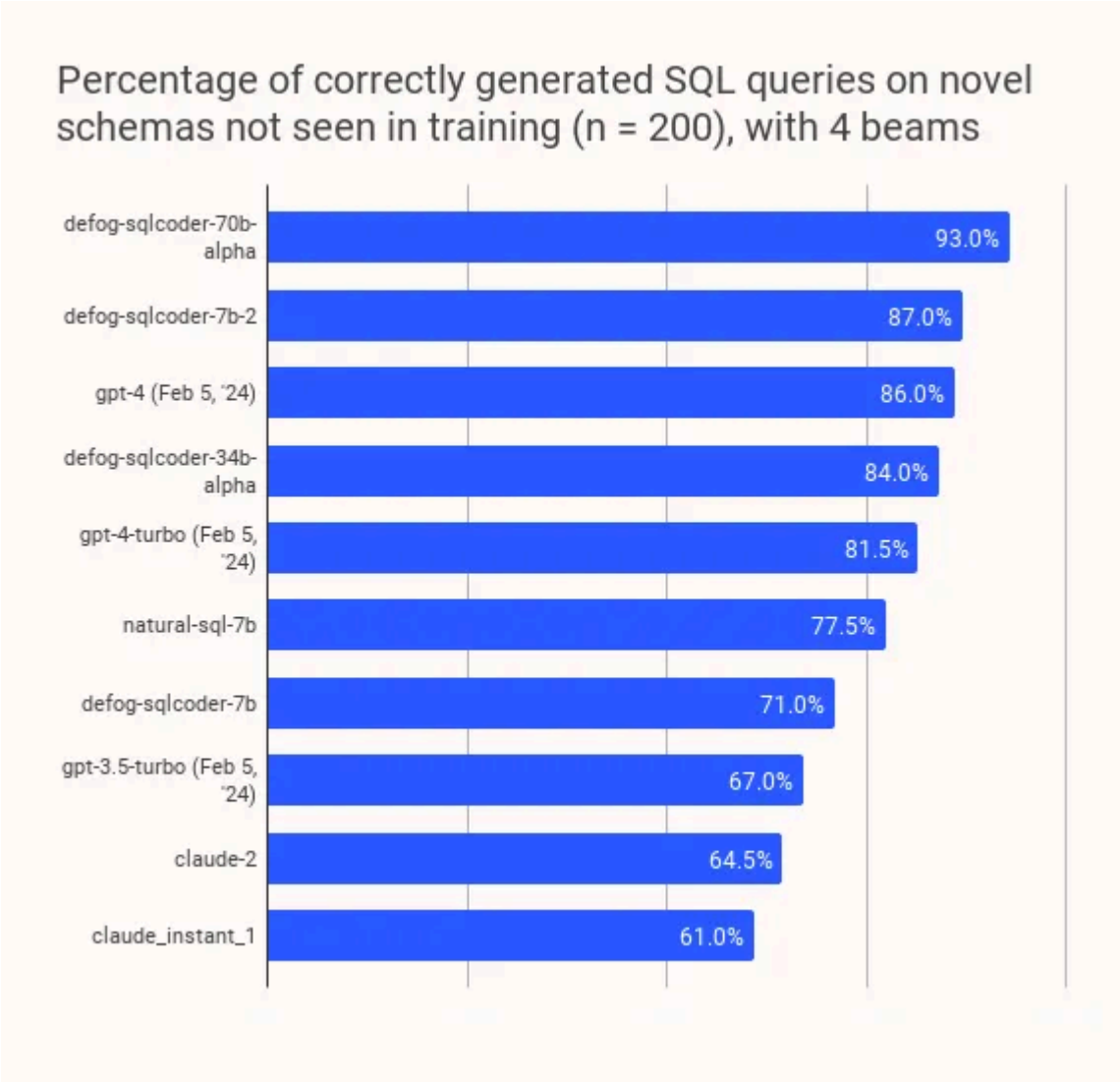
Published in Follower Booster Hub

4 min read · Feb 16, 2024

▶ Listen   ⬆ Share   ••• More

Percentage of correctly generated SQL queries on novel schemas not seen in training (n = 200), with 4 beams

Model	Percentage
defog-sqlcoder-70b-alpha	93.0%
defog-sqlcoder-7b-2	87.0%
gpt-4 (Feb 5, '24)	86.0%
defog-sqlcoder-34b-alpha	84.0%
gpt-4-turbo (Feb 5, '24)	81.5%
natural-sql-7b	77.5%
defog-sqlcoder-7b	71.0%
gpt-3.5-turbo (Feb 5, '24)	67.0%
claude-2	64.5%
claude_instant_1	61.0%

At the peak of the AI wave, a model named SQLCoder-7b was shockingly released on Hugging Face, rivaling GPT-4 in text-to-SQL generation and achieving remarkable breakthroughs in data processing speed and accuracy, even showing potential to surpass GPT-4.

However, it raises curiosity about how this model achieved such a leap and what its secret weapon might be.

Model Features

Have you ever felt confused while trying to understand data in SQL databases? With the SQLCoder-7B-2 and SQLCoder-70B-Alpha models, these issues are effectively addressed.

		date	group_by	order_by	ratio	join	where
1							
2		--------------	----	--------	--------	-----	----	-----
3	sqlcoder-70b	96	91.4	97.1	85.7	97.1	91.4
4	sqlcoder-34b	80	94.3	85.7	77.1	85.7	80
5	gpt-4	64	94.3	88.6	74.2	85.7	80
6	sqlcoder2-15b	76	80	77.1	60	77.1	77.1
7	sqlcoder-7b	64	82.9	74.3	54.3	74.3	74.3
8	gpt-3.5	68	77.1	74.2	34.3	65.7	71.4
9	claude-2	52	71.4	74.3	57.1	65.7	62.9

Designed for non-technical users, these powerful tools allow for easy analysis of database contents without the need for deep expertise.

Excitingly, SQLCoder-70B-Alpha surpasses all general models, including GPT-4, in text-to-SQL conversion capability, meaning it can understand your needs more accurately and generate corresponding SQL queries.

SQLCoder-7B-2 also stands out, excelling in natural language to SQL generation.

However, these models are intended for read-only access to ensure data security and are not trained to handle malicious write requests.

Open-source spirit plays a crucial role in driving technological advancements.

The SQLCoder2 and SQLCoder-7B models are now open to the public, broadening access to their capabilities.

SQLCoder2 significantly improves upon the original SQLCoder, while SQLCoder-7B debuts with 7 billion parameters, performing comparably to its predecessor.

Accuracy by category on out-of-training-set schemas

	date	group_by	order_by	ratio	join	where
gpt-4	72	91.4	82.9	80	82.9	80
sqlcoder2-15b	76	80	77.1	60	77.1	77.1
sqlcoder-7b	64	82.9	74.3	54.3	74.3	74.3
gpt-3.5	68	77.1	68.6	37.1	71.4	74.3
claude-2	52	71.4	74.3	57.1	65.7	62.9
claude-instant	48	71.4	74.3	45.7	62.9	60
gpt-3	32	71.4	68.6	25.7	57.1	54.3

Behind all this is the development team's deep understanding of user feedback and continuous improvements.

They addressed the original SQLCoder's difficulties with date-time functions and the occasional generation of incorrect column or table names. Additionally, there was a significant community demand for models capable of running on smaller GPUs.

Thus, SQLCoder2 and SQLCoder-7B were developed, not only resolving these issues but also being more suitable for production environments.

How does the SQLCoder evaluate AI models with high accuracy?

Evaluation and Performance

SQLCoder classifies each generated question into six major categories. The following table will show you the accuracy rate of each model in answering questions across these categories.

	date	group_by	order_by	ratio	join	where
sqlcoder-70b	96	91.4	97.1	85.7	97.1	91.4
sqlcoder-7b	96	85.7	97.1	85.7	82.8	77.1
sqlcoder-34b	80	94.3	85.7	77.1	85.7	80
gpt-4	72	94.3	97.1	80	91.4	80
gpt-4-turbo	76	91.4	91.4	62.8	88.6	77.1
natural-sql-7b	56	88.6	85.7	60	88.6	80
sqlcoder-7b	64	82.9	74.3	54.3	74.3	74.3
gpt-3.5	72	77.1	82.8	34.3	65.7	71.4
claude-2	52	71.4	74.3	57.1	65.7	62.9

In the sea of data, SQL is the net for fishing. How reliable are SQL statements generated by LLMs? SQL-Eval is here to judge!

As its name suggests, SQL-Eval's mission is to evaluate the correctness of SQL statements generated by LLMs.

During development, developers often face a dilemma:

> *how to determine whether an SQL query is "correct"? After all, there may be multiple correct ways to write SQL for the same question.*

For example, if you want to find the last 10 users from Toronto, both of the following queries are correct:

1. Query A retrieves the user's ID, username, and creation time;

2. Query B selects the user's ID, full name, and creation time.

```
1   -- query 1
2   SELECT userid, username, created_at
3   from users
4   where city='Toronto'
5   order by created_at
6   DESC LIMIT 10;
7
8   -- query 2
9   SELECT userid, firstname || ' ' || lastname, created_at
10  from users
11  where city='Toronto'
12  order by created_at
13  DESC LIMIT 10;
```
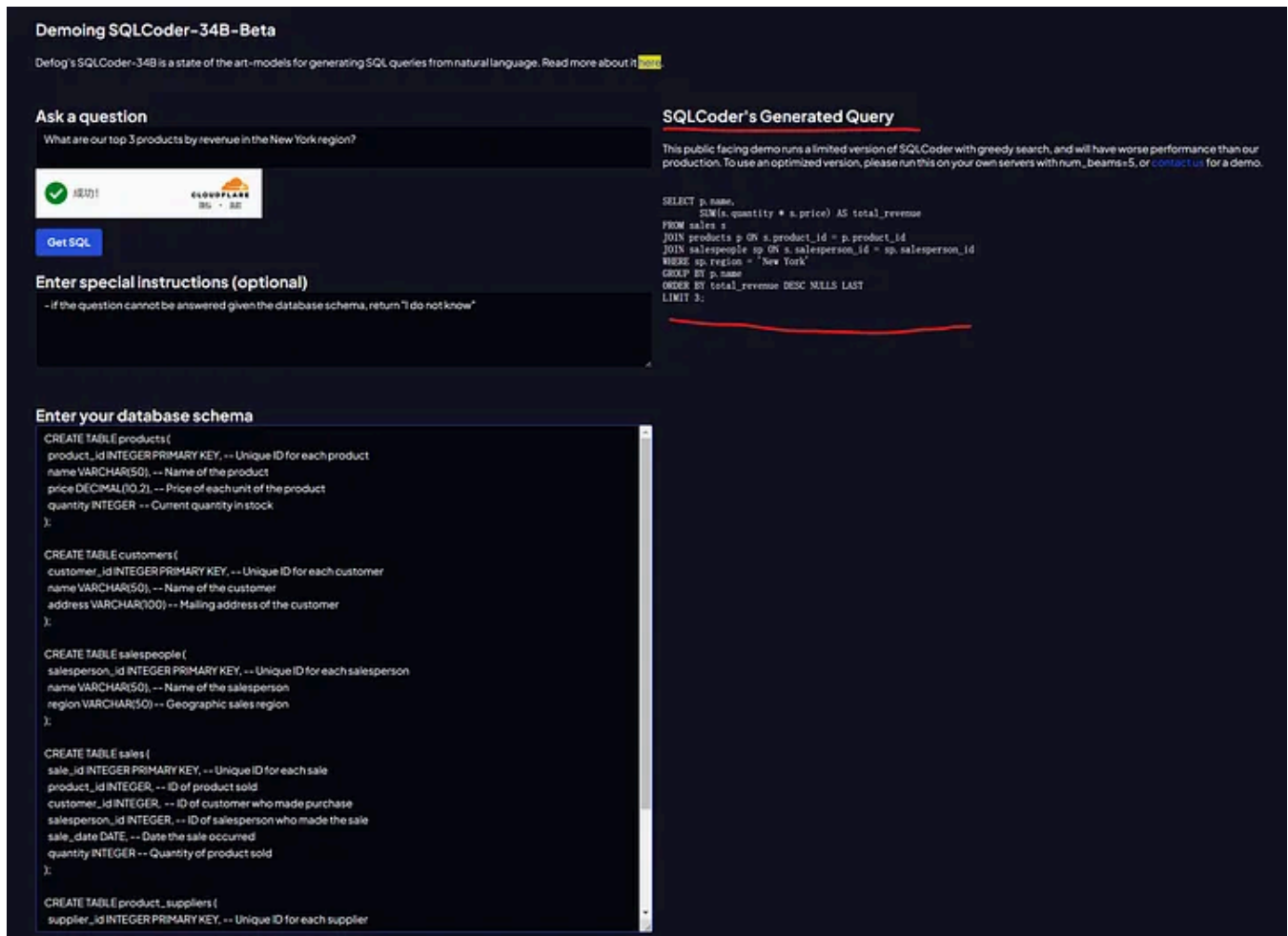
How to Get Started with SQLCoder-7b Quickly

Prompt format

```
### Task
Generate a SQL query to answer [QUESTION]{user_question}[/QUESTION]

### Database Schema
The query will run on a database with the following schema:
{table_metadata_string_DDL_statements}

### Answer
Given the database schema, here is the SQL query that [QUESTION]{user_question}
[SQL]
```

Use Environment

Example:

https://colab.research.google.com/drive/1z4rmOEiFkxkMiecAWeTUlPl0OmKgfEu7

Demo: https://defog.ai/sqlcoder-demo/

Model: https://huggingface.co/defog/sqlcoder-7b-2

Github: https://github.com/defog-ai/sqlcoder/

Conclusion

The rise of SQLCoder-7b represents not only a technological leap but also an important step towards the democratization of data insights.

Want to discover more secrets of AI? Follow us for more exciting discoveries!

I am Meng Li, an independent open-source software developer, and author of the SolidUI AI painting project, very interested in new technologies, and focused on AI

and the data field.

Sql   Generative Ai Tools   AI   Llm   Open Source

Open in app ↗

Search

Youssef Hos… in Towards …

Building RAG Application using Gemma 7B LLM &…

✦ · 9 min read · Mar 1, 2024

Jesus Rodrig… in DataSer…

Five Key Facts Wu Dao 2.0: The Largest…

✦ · 4 min read · Aug 31, 2021

Ayush Th

Top LLM M HuggingFa

8 min read ·

View list

Follow

Written by Meng Li

222 Followers · Writer for Follower Booster Hub

AI Engineer,Full-time open source engineer, Apache Linkis Committer, initiator of the SolidUI AI painting project.

More from Meng Li and Follower Booster Hub