

# Understanding Performance of LLM to Generate SQL Queries

Minhyuk Ko  
(minhyukko@vt.edu)

Dibyendu Brinto Bose  
(brintodibyendu@vt.edu)

Weilu Wang  
(weilu@vt.edu)

Mohammed Seyam  
(seyam@vt.edu)

Chris Brown  
(dcbrown@vt.edu)

## Goal

The goal of this research is to understand how effective off-the-shelf AI assistants, such as GitHub Copilot and ChatGPT, can be in generating SQL queries when the programmer requests a query in plain English. This process can democratize access to databases for non-experts and reduce keystrokes and cognitive load for database experts.

## Methodology

We used AI programming assistants, like GitHub CoPilot and Tabnine, and LLM-based chatbots, like ChatGPT and Google Gemini, to generate SQL queries based on a database from Kaggle. We measured the accuracy of SQL queries through test cases developed by ChatGPT, and we used our custom-built web application to evaluate the quality of SQL queries. Generated SQL queries were able to pass 18.78% (n = 120/639) of test cases on average.

Off-the-shelf AI Assistants produce low-quality SQL queries using natural language prompts. However, better prompting can possibly lead to better SQL query generation.



Take a picture to download the full paper

## Query Prompt Examples

- Display the id, birthday, height, weight who is a player.
- List all the teams that have won more than half of the matches that they played
- Adam Brown has gained 20 pounds due to eating too many hamburgers. Update the player table accordingly.

## Accuracy Results

AI Tool/LLM	Accuracy
ChatGPT	55.0% ( $n = 39$ )
Amazon Q	50.7% ( $n = 36$ )
Tabnine	21.1% ( $n = 15$ )
Blackbox	16.9% ( $n = 12$ )
Bing Chat	14.1% ( $n = 10$ )
Google Gemini	11.3% ( $n = 8$ )
GitHub Copilot	0% ( $n = 0$ )
Codeium	0% ( $n = 0$ )
CodeGeeX	0% ( $n = 0$ )
Total:	18.78% ( $n = 120/639$ )

## SQL Query Evaluation

The following criteria were used to evaluate the quality of the SQL query:

- Indentation
- Naming Convention
- Number of Comments
- Error Handling Mechanism
- Updateability

We developed a web application that would generate a readability report when the users provide their SQL queries. The web application is publicly available on the GitHub link provided in the paper.

