AI-Powered SQL for Real-World Data:
From Questions to Insights

MSR20

Authors:
Kontonis Vasileios,MSc|Vasileios.Kontonis@pfizer.com,
Kountouris Nikolaos,MSc|nikolaos.kountouris@pfizer.com,
Bokou Dimitra,MSc|dimitra.bokou@pfizer.com
Presenting author: Kontonis Vasileios

INTRODUCTION

Accessing and analyzing real-world data (RWD) is essential for healthcare decision-making. However, many professionals—including physicians—lack SQL expertise, and even for those proficient, running queries can be time-consuming. To address this, we developed an AI-powered system that transforms natural language questions into SQL queries, enabling quick and easy database exploration and early insight generation.

OBJECTIVE

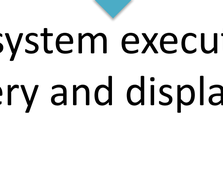
To assist non-technical users in exploring databases, identifying trends, and conducting basic feasibility assessments. The system also serves as a supportive tool that enhances understanding of database content and promotes early-stage insight generation. Technically skilled users can benefit as well by quickly generating and testing simple queries, improving efficiency.

METHOD

A Large Language Model (LLM) was fine-tuned for SQL generation, enabling it to interpret domain-specific natural language questions and produce accurate, executable SQL queries with visual outputs. The model was trained on curated SQL examples to ensure functionality, enhance its capabilities, and execute effectively on a synthetic database.

TECHNICAL SOLUTION

1. User enters a question in natural language

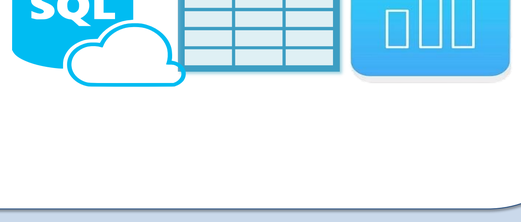


2. LLM translates this question to SQL query

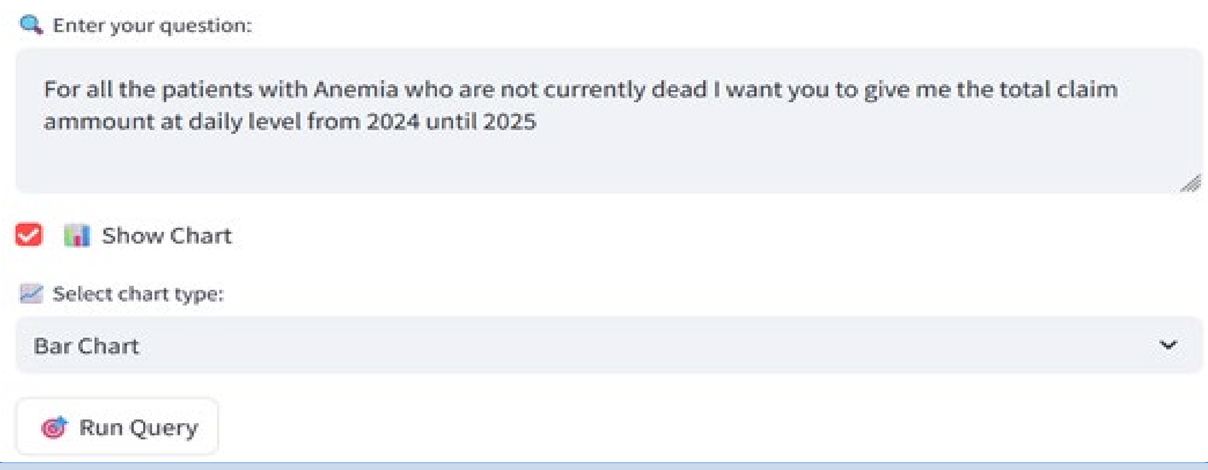
3. Execute the SQL query on the Database

4. The system executes the query and displays:

the SQL code | tabular output | graphs



USER INTERFACE DEMONSTRATION

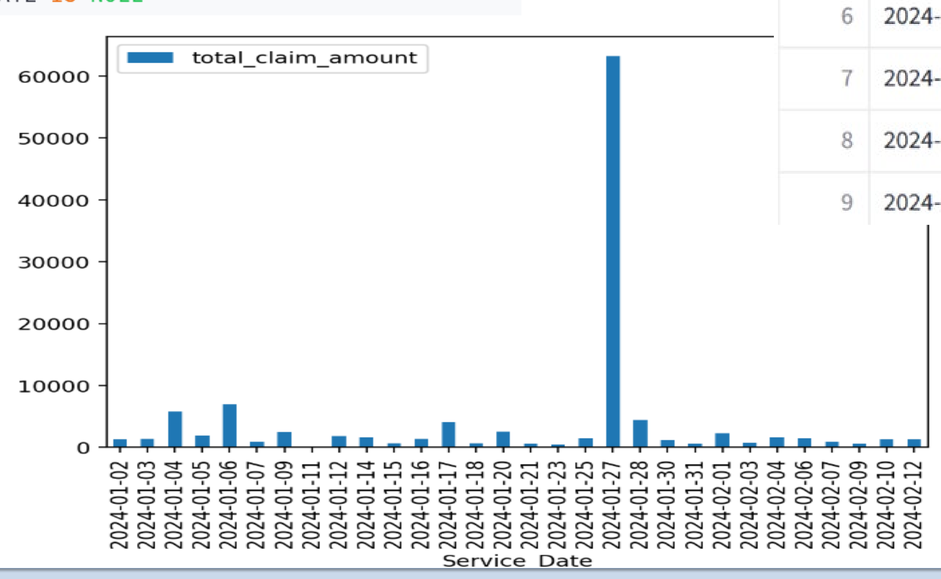


OUTPUT DEMONSTRATION

```sql
SELECT     DATE(claims.SERVICEDATE) AS Service_Date
          ,SUM(claims_transactions.AMOUNT) AS total_claim_amount
FROM       patients
INNER JOIN conditions
        ON patients.Id = conditions.PATIENT
       AND conditions.DESCRIPTION LIKE '%Anemia%'
INNER JOIN claims
        ON patients.Id = claims.PATIENTID
       AND claims.SERVICEDATE >= '2024-01-01'
       AND claims.SERVICEDATE <= '2025-12-31'
INNER JOIN claims_transactions
        ON claims.Id = claims_transactions.CLAIMID
WHERE      patients.DEATHDATE IS NULL
GROUP BY   Service_Date
ORDER BY   Service_Date
LIMIT 30;
```

Result:

	Service_Date	total_claim_amount
0	2024-01-02	1,266.65
1	2024-01-03	1,372.47
2	2024-01-04	5,797.14
3	2024-01-05	1,902.54
4	2024-01-06	6,961.86
5	2024-01-07	936.9
6	2024-01-09	2,443.07
7	2024-01-11	171.1
8	2024-01-12	1,820.24
9	2024-01-14	1,653.49



RESULTS

This system serves as a supportive tool that enhances understanding of database content, facilitates trend identification, and enables early-stage insight generation, thereby streamlining healthcare decision-making. By combining AI-driven query generation with intuitive visual outputs, it allows both non-technical and technical users to efficiently extract meaningful insights. The tool has been trained to handle and execute more complex queries, enabling technically skilled users to quickly generate and test advanced queries, further improving workflow efficiency.