Project Title:

Digital Egypt Pioneers

Chat With Your Database

AI Database Assistant for Non-Technical Users

Team Members

1. Omer Alaa El Deen
2. Abdelmageed Fathy Abdelmageed
3. Ahmed Eldairy
4. Hanan Khaled
5. Mohamed Hussein

Under Supervision: Eng/Mohammed Agoor

Instructor of Data Science & AI in Digital Egypt Pioneers Initiative

Introduction

Digital Egypt Pioneers

- In today's data-driven world, many individuals and businesses need to interact with databases but lack the technical expertise to write SQL queries.
- This project introduces an AI-powered system that bridges the gap between non-technical users and databases by enabling them to retrieve data through simple natural language questions.

Project Objectives

- Build a system that connects to any database based on user-provided credentials.
- Fetch database schema automatically
- Use a fine-tuned LLM (Large Language Model) to generate SQL queries from user input based on the schema.
- Execute generated SQL queries and return results to the user in a user-friendly format.

System Architecture

Digital Egypt Pioneers

Main Components:

1. User Interface:

- Collects database connection info (type, host, port, username, password).
- Collects user's natural language questions.

2. Database Connector:

- Establishes secure connection to the specified database.
- Retrieves full schema information (tables, columns, relationships).

3. LLM (Large Language Model):

- Fine-tuned using LLaMA-Factory.
- Input: Schema + User question (prompt).
- Output: SQL Query.

4. Query Executor:

- Executes generated SQL against the connected database.
- Retrieves and formats the results.

5. Response Layer:

- Presents the result back to the user.

Technologies Used

Digital Egypt Pioneers

Component	Technology
Backend Development	Python (FastAPI)
Database Connections	SQLite / PostgreSQL / MySQL / MariaDB
LLM Fine-Tuning	LLaMA-Factory
Model Base	Gemma3
Training Dataset	Custom dataset (105,000+ examples covering 100 domains)

Dataset Details

- Size: Over 105,000 examples.
- Domains: 100+ industries (Healthcare, Aviation, Sports, etc.).
- SQL Complexity: Includes simple to complex queries (aggregations, joins, window functions).
- Contextual Data: Provided CREATE TABLE scripts + natural language query explanations.
- Validation: Surpassed Spider Benchmark in SQL generation quality (validated with GPT-4).

Model Training Summary

- **Framework:**

  - LLaMA-Factory

- **Fine-Tuning Method:**

  - Supervised fine-tuning (SFT)

- **Input Format:**

  - Schema information.

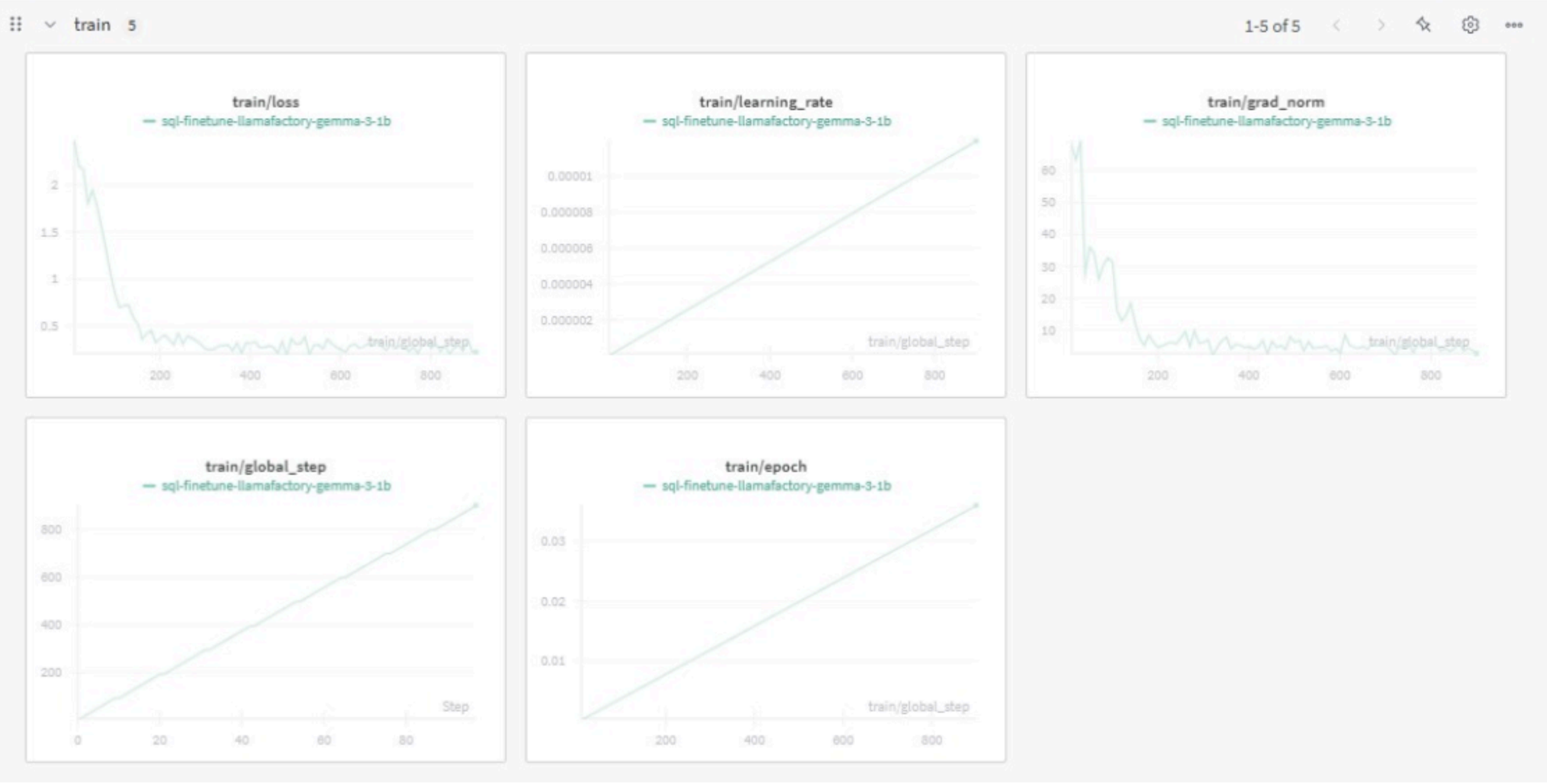
  - User question.

- **Output Format:**

  - Valid and optimized SQL query.

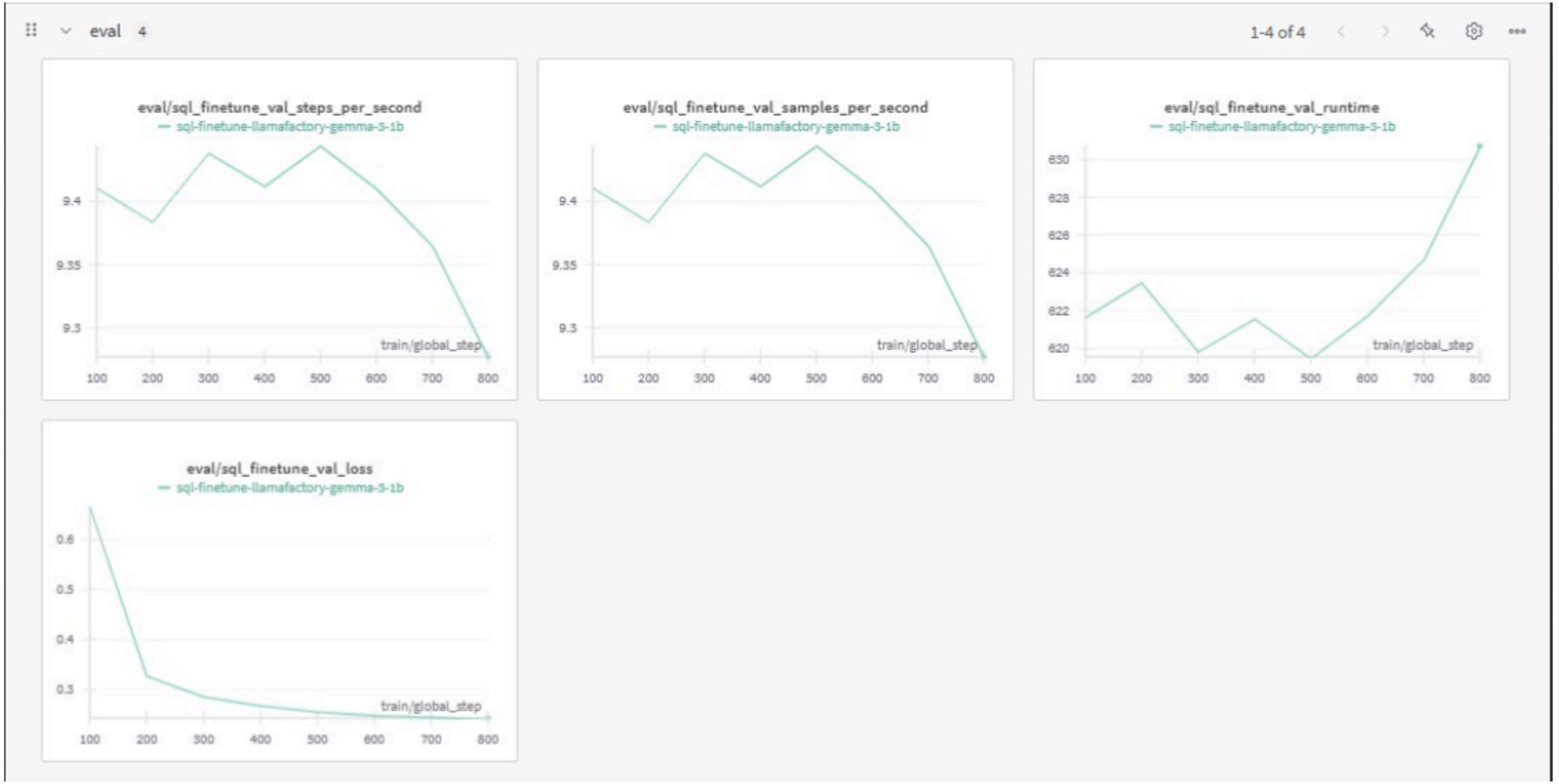
Model Training Summary

Digital Egypt Pioneers



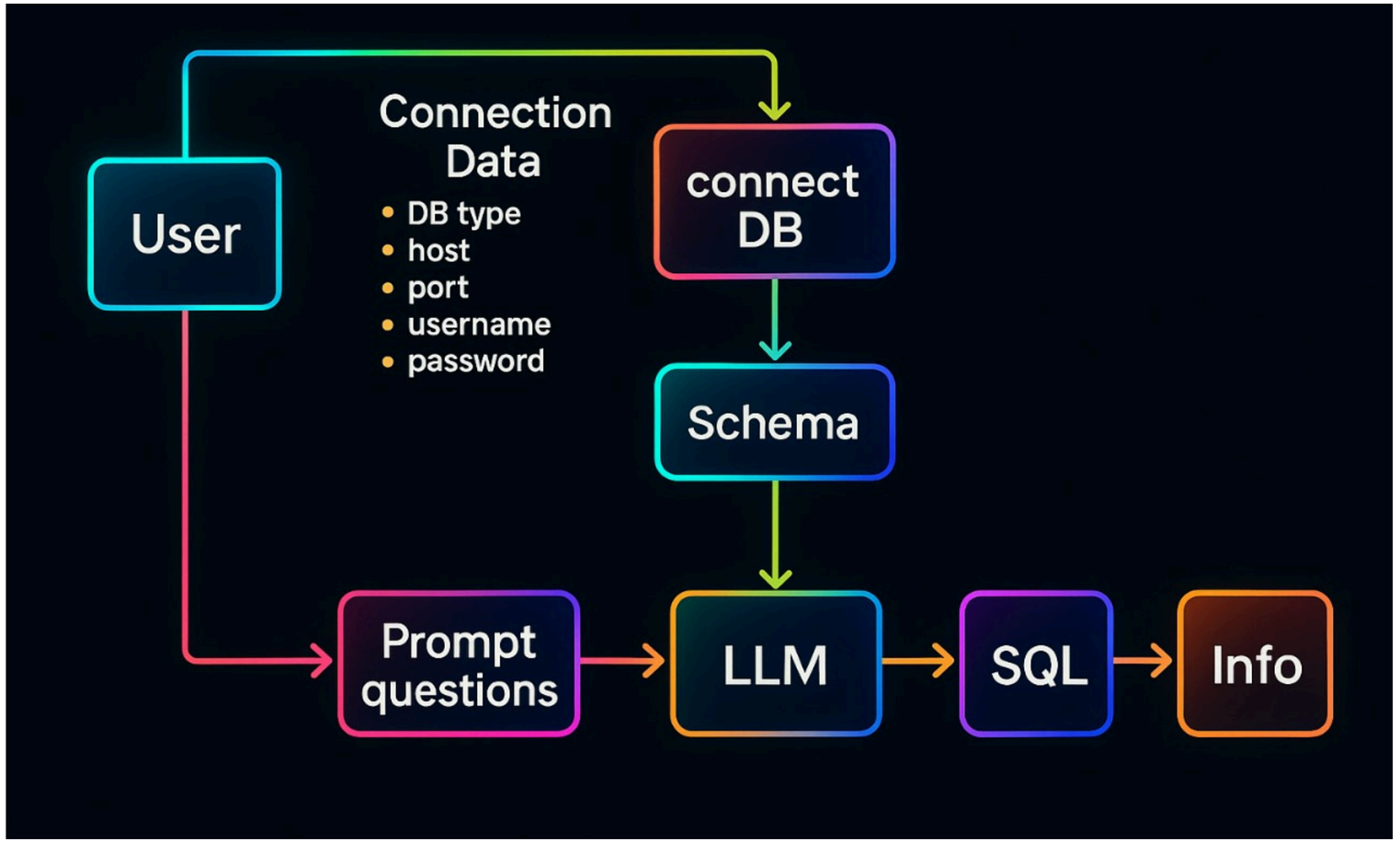
Model Validation Summary

Digital Egypt Pioneers



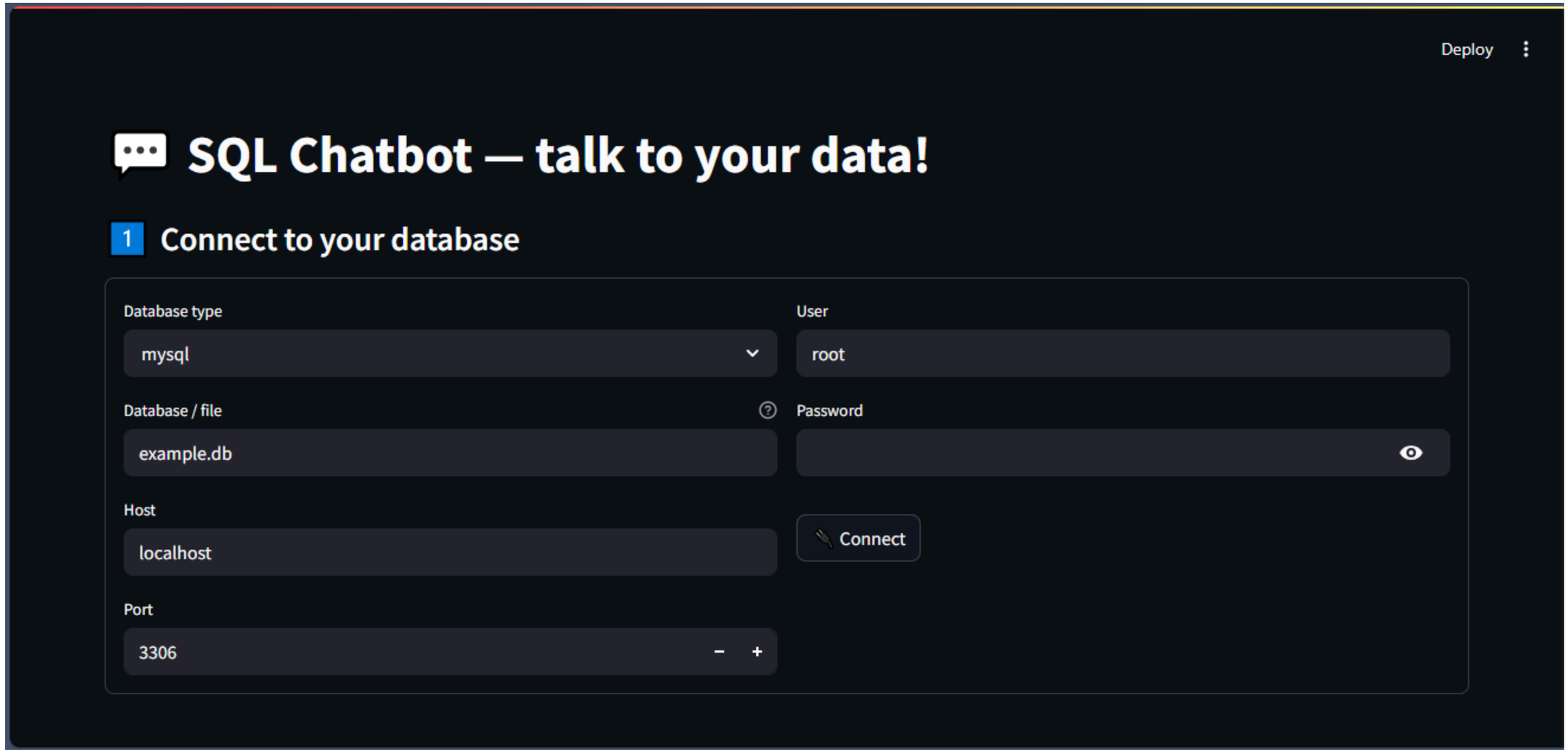
System Workflow (Step-by-Step)

Digital Egypt Pioneers



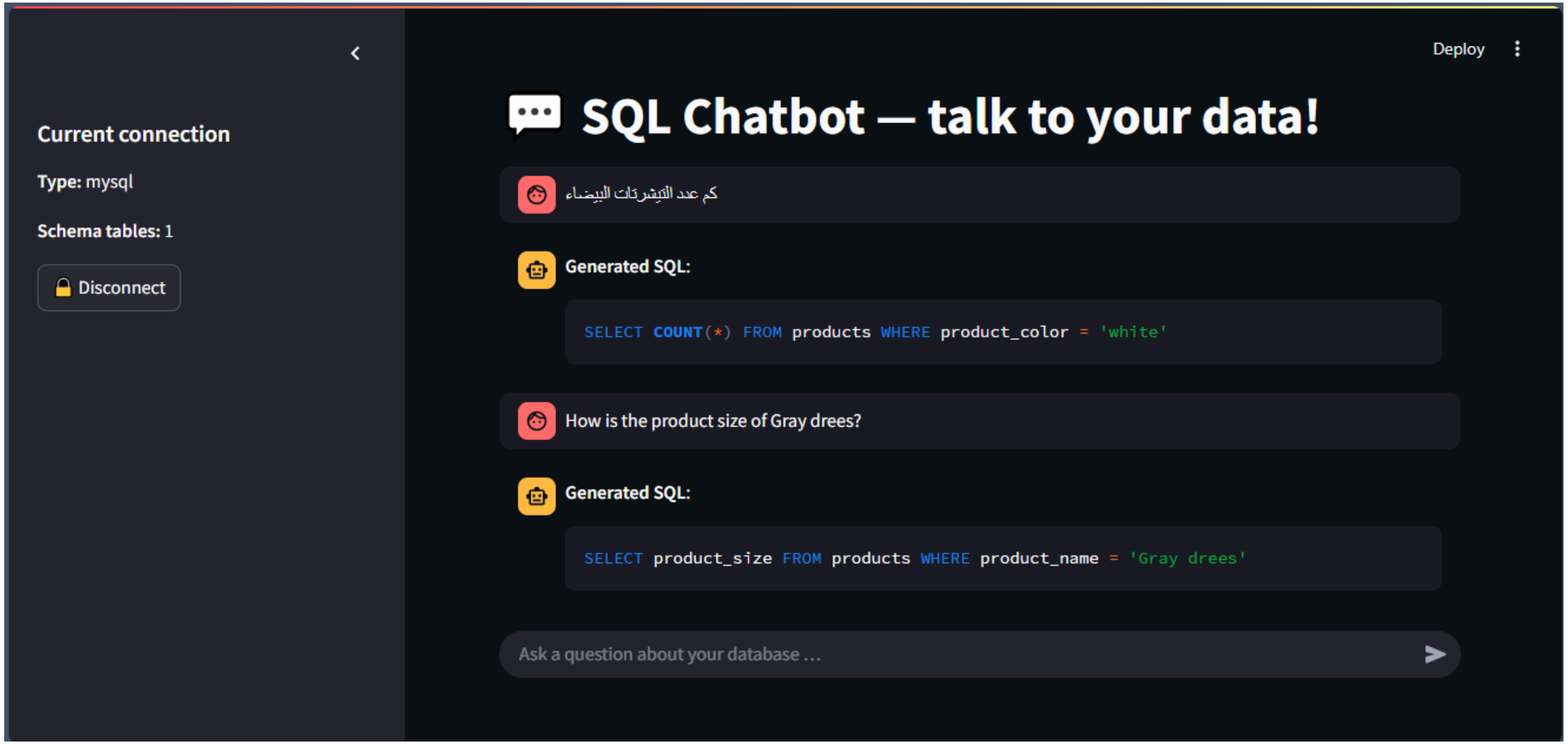
System Workflow (Step-by-Step)

Digital Egypt Pioneers



System Workflow (Step-by-Step)

Digital Egypt Pioneers



1- User inputs:

- Database type (MySQL, PostgreSQL, etc.)
- Host, Port, Username, Password

2- System connects to the database and extracts schema metadata.

3- User submits a natural language query.

4- System sends (schema + question) to the LLM.

5- LLM responds with a valid SQL Query.

6- System executes the SQL on the live database.

7- System returns the result to the user.

Challenges Faced

- Handling different database engines and schema formats.
- Ensuring SQL generation robustness across domains.
- Managing SQL execution safely (avoiding harmful queries).
- Reducing latency in response time for real-time interaction.

Conclusion

Digital Egypt Pioneers

- This project demonstrates the power of AI in making complex technologies like databases more accessible to everyone.

- Through fine-tuning a powerful LLM and designing a seamless system architecture, we successfully built a tool that democratizes data interaction, making it as simple as asking a question in natural language.

LINK : https://github.com/OmarAladi/chat-with-your-database-