Intelligent Data Solutions

Enterprise Agentic Text-to-SQL System

Secure Data Democratization via Self-Correcting AI

Presenter - Anubhav Agrawal

January 16, 2026

Introduction: The Data Bottleneck

Problem Statement

Data Overload

Massive enterprise data stored in relational databases creates significant challenges for efficient data retrieval and analysis.

User Barriers

Non-technical users often struggle to query effectively, leading to frustration and a lack of actionable insights.

Analyst Dependency

Organizations become dependent on data analysts for queries, causing delays in accessing crucial information and insights.

Latency

Faster insights lead to better decisions and improved ROI, making efficient data access essential for business success.

Scope and Business ROI

Project Scope

Targeting enterprise relational databases (Sales, CRM, Product Data). The goal is to allow non - technical staff to query this data using natural English language

Operational ROI

Reduces “Time - to - Insight” from days to seconds. Frees up expensive data engineering resources for strategic tasks rather than ad-hoc queries.

Productivity Impact

Empowers a self - service culture. Managers can instantly verify hypothesis without waiting for weekly reports.

Literature Review: The Paradigm Shift

Traditional Approaches

Static Dashboards : Rigid views. Cannot answer “Why” or ad-hoc questions.

Basic Chatbots: Standard LLMs hallucinate data and fails on complex joins.

Major Gap : Lack of privacy. Sending raw SQL data to the cloud is compliance violation.

Our GenAI Approach

Agentic Reasoning : Uses a “Self-Correction Loop” to fix its own SQL errors.

Zero-Trust Security: Redacts PII locally before the LLM ever sees it.

Context Awareness: Injects DB Schema dynamically for high precision and efficiency.

We Aim to Solve

Inflexibility of Traditional Dashboards

Limitations of Standard LLMs Explained

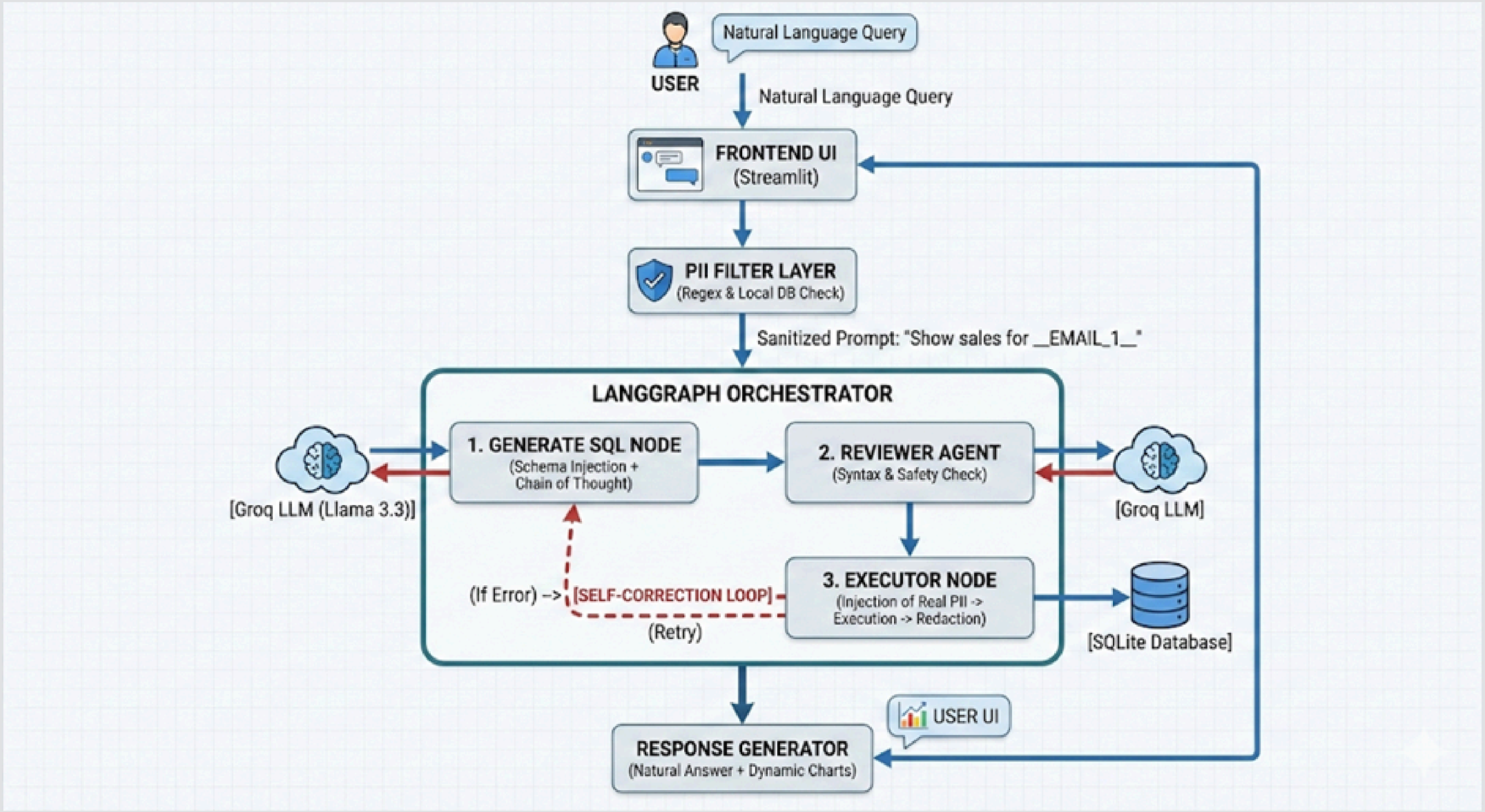
Address Key Gaps in Security

System Architecture

A secure, multi-stage pipeline designed for Enterprise Compliance.

User → PII Filter → LangGraph Orchestrator → SQLite → Visualizer

Architecture Flow



Technology Stack

Groq (Llama 3.3)

Inference Engine

Selected for ultra-low latency SQL generation.

LangGraph

Orchestrator

Manages state, retry loops, and agent decisions.

Streamlit

Frontend

Provides the chat interface and dynamic visualizations.

How It Works: Data Flow

1

Sanitize

User input is intercepted. Emails are masked (e.g. “_EMAIL_!_”) locally.

2

Generate

Agent receives masked prompt + Schema. Writes SQL Code.

3

Validate

Reviewer Agent checks syntax. If error, it loops back to fix it.

4

Execute

Real data is swapped back in for execution. Results are redacted and secure.

Live Demo

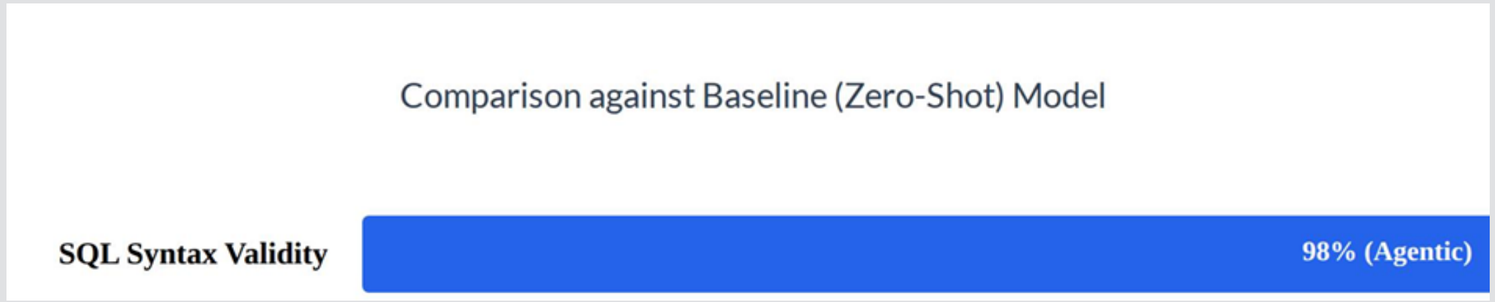
Showcasing : Chat Interface | PII Redaction | Dynamic Charts

Evaluation and Performance Metrics

```
==================================
📊 EVALUATION REPORT
==================================
Total Tests:       10
Execution Accuracy: 100.00%  <-- Your Main Metric
Valid SQL Rate:     100.00%
Avg Latency:        1.60s
==================================
```

Method Used:

Used Execution Accuracy to evaluate the pre-set SQL queries with the one that LLM generated comparing the output produced.

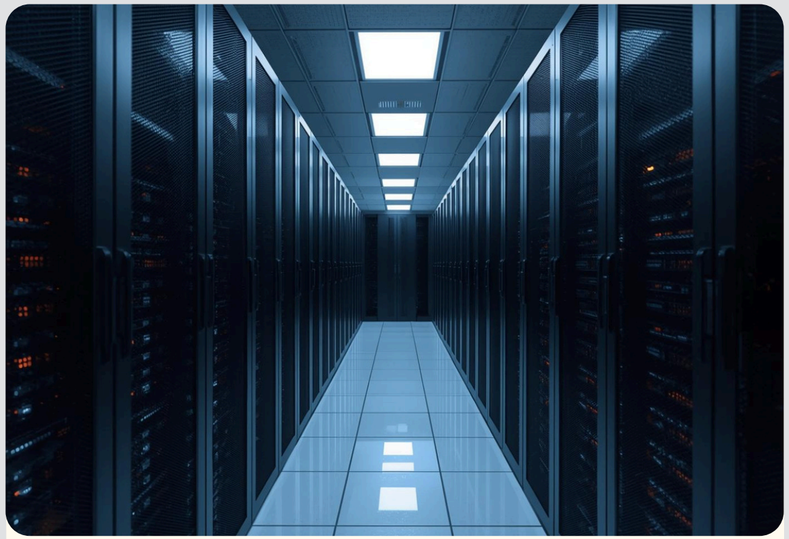


Used Syntax Validator to evaluate the string matching (Fuzzy Match) with the pre-set SQL queries with the LLM generated syntax.

Unique Capabilities

Boosting accuracy with self-correction loop technology

Zero Trust PII Filtering Using Placeholder Injection



Ensuring compliance through local execution strategies

Achieving minimal latency for enterprise applications

Challenges and Roadmap

Current Challenges

Ambiguity : Vague questions (“Show me the good products”) still require user clarification.

Latency: The “Check → Fix” loop increases the response time.

Future Roadmap

Voice Interface: Hands - free SQL querying for field agents.

Multi-DB Connectors: Support for multiple DB including PostGres and SnowFlake.

Conclusion

“This project moves beyond simple chatbots to create a secure, reasoning Agent. By solving the “Data Access” problem, we empower every employees to play with data.”

Thank You !!

Questions and Feedback