AI Potential with Service Requests and Incidents

Team InfVision: Dexter Xu | Haocheng Bao | Sophie Deng | Wenjing Li
About Our Team

Sophie Deng
yuqing91@uw.edu
LinkedIn

Wenjing Li
wli25@uw.edu
LinkedIn

Dexter Xu
chenjux@uw.edu
LinkedIn

Haocheng Bao
hvb5223@uw.edu
LinkedIn
About Our Sponsor

City of Seattle

Evan Curtis
Enterprise Architect
Client Engagement Team
Background: What is the Seattle IT Service Hub?

The Service Hub provides clients with a one-stop shop service for requesting services, reporting issues and tracking orders, making it easier and faster to engage with Seattle IT.
Problem

😔 Hard to retrieve valuable information from large dataset
😔 Analytical difficulties arising from the complexity of the dataset.

Limitations

Pivot Tables
Potential Solution?

Data is not the thing I lack, but information is what will help me to be successful.

What if the information could be accessed through conversation?

What if have a modern interface to get information at the speed of thought?
Objective: Transforming Organizational Management

- Which team owns the most requests?
- What are the most common subject in the requests?
- What department has the most requests?

- How many requests were assigned to me this week?
- What is the oldest request?

- How do I change the team for a task assigned to my team?
Opportunity: Large Language Model Applications

Large language models (LLMs) are models trained on immense amounts of data, making them capable of understanding and generating natural language and other types of content to perform a wide range of tasks.
Action: Ollama + Langchain + Streamlit = Local Chatbot

- Enhanced data privacy
- Customization and Control
- Compatible with existing frameworks
What is Langchain?
How does Langchain work?

Document Loading

Splitting

Storage

Retrieval

Output

URLs

PDFs

Documents

Splits

Vectorstore

Relevant Splits

Query

Prompt

LLM

< Answer >
How does Evan 9000 work?

Step 1: Ask a Question
Step 2: Process Question & Generate a SQL
Step 3: Extract Data through SQL
Step 4: Process data and generate answer with good format
Prototype

Give me the oldest request?

The oldest request is:


Which client has made the most requests?

The client who has made the most requests is [blank] with 9 requests.
Performance of Evan 9000

“Straight Forward” Questions:
How many requests from VIP clients were breached?

Accuracy Above 90% ★

Complex Questions:
Which team has the most workload?

Accuracy around 20%
Challenges & Solutions in Techs

How to face the Challenges and improve performance?

- Lack of background information
  - Hallucination Problem
- Limitations of LLM
  - Incorrect SQL

Solutions:
- Fine-Tune Model
- Better Performance Model
- Prompt Engineering
Prompt Engineering: Elevating ability in generating complex SQL

Provide background information in prompt to optimize AI model performance

**Before prompt engineering: incorrect SQL**

```
ProgrammingError: (psycopg2.errors.SyntaxError) syntax error at or near "\"" LINE 1: SELECT subject, count(*) as num\_breaches ^ [SQL: SELECT subject, count(*) as num\_breaches FROM requestinfo WHERE isbreached = true GROUP BY subject ORDER BY num\_breaches DESC;] (Background on this error at: https://sqlalche.me/e/14/f405)
```

**After prompt engineering**

- What kind of requests are the most likely to be breached?

- Based on the given table schema, question, SQL query, and SQL response, here is a natural language answer:

  "Of the requests that have been breached, the most common types are:

  1. Purchase Hardware or Software (3 requests)
  2. Consulting for IT Services (2 requests)
  3. New or Update Service Hub Request Offering (1 request)
  4. Mobile App Intake (1 request)
  5. Software/Application Access and License Management (1 request)"
Our project is an experiment of the industry/organizational level applications of large language models (LLMs), focusing on exploring promising use cases.
Smart Agents

Automate
Leverage LLM technologies to streamline repetitive tasks and boost productivity.

Tech-Savvy Tasks
Enable employees to leverage cutting-edge technologies with ease.

Empower Employees
Free employees to concentrate on high-impact, strategic initiatives.
Upgraded Workflow

Seamless Experience
Provide a user-friendly, intuitive interface that simplifies complex tasks.

Accessible Data
Enhance data retrieval, empowering employees with independent analysis.

Increased Efficiency
Streamline workflows and boost productivity through smart automation.
Balancing LLM Strengths and Weaknesses?

Using AI? ↔ Collaborating with AI!
Enterprise Application: Use Case

What department submits the most requests?

Which department submits most requests?

Based on the table schema provided, here is the SQL query that would answer the user’s question:

```
SELECT client_department FROM tbrequestinfo GROUP BY client_department ORDER BY COUNT(*) DESC;
```

And here is the natural language answer to the user’s question:

"Based on the requests in the table, the department that submits the most requests is 'Finance and Administrative Services', followed by 'Seattle Information Technology', 'Seattle Department of Transportation', 'Seattle Department of Construction and Inspections', and 'Human Services Department'."
Enterprise Application: Customized Business Context

Vector DB → Embedding model → Retrieved contexts → LLM Agent #1, LLM Agent #2, …, LLM Agent #N

Pre-Defined context (SQL) → Query → Response
Industry Commercial Efforts

City of New York

Snowflake
Customization is All You Need

- Data Privacy
- Cost
- Small Business
- Architecture Platform
The limits of my languages mean the limits of my world.

— Wittgenstein, 1922
Thank you!

Q & A

Contact: infvisiongroup@gmail.com

Visit Project Website