

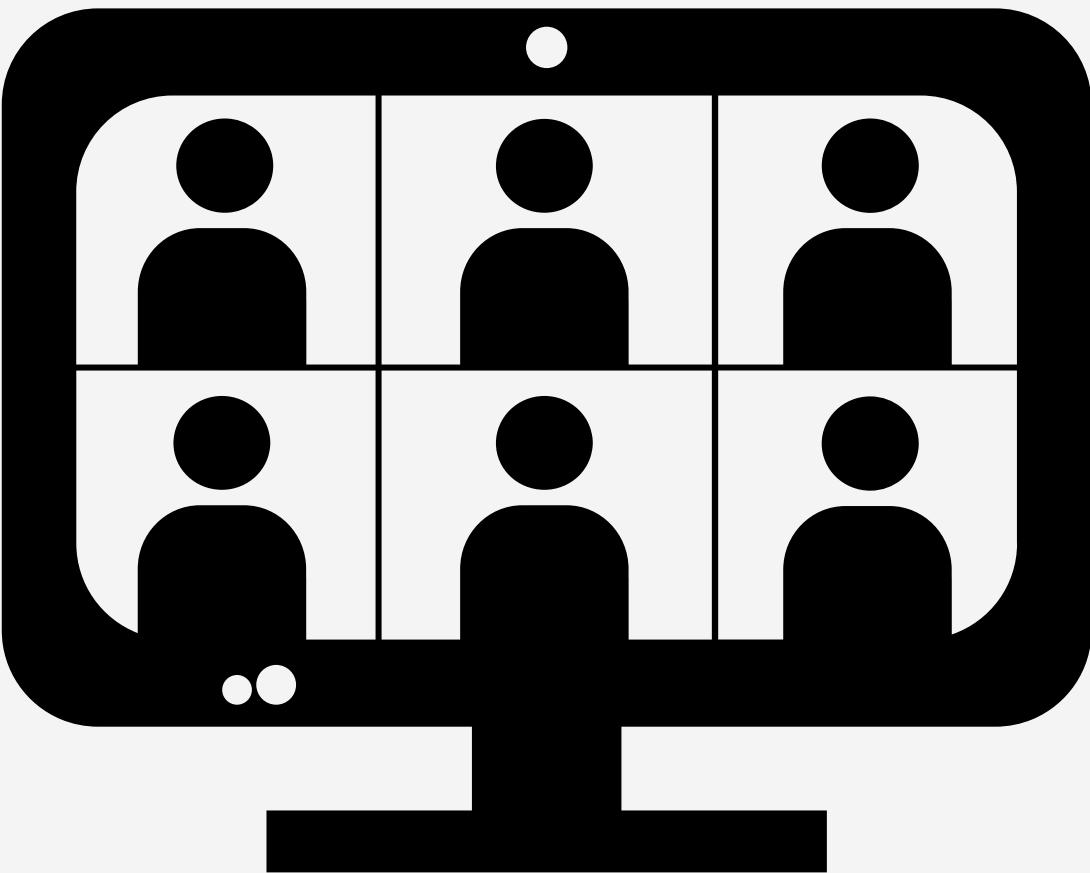
HPE Project Presentation

Text-2-SQL Code Generation And Execution

Mentor: Mr. Karthikeyan Soundararajan Sir

Agenda Of Meeting

Objective	3
Roadmap of Project	4
Development Cycle	5
Architecture	6
Outcome Of Project	7
Initial Prototype	8
References	9



Objective

Use machine Learning models to generate SQL code from Natural Language



Goal # 1

Create a model that accurately translates natural language queries into SQL statements.



Goal # 2

Integrating Large Language Models (LLMs) and Generative AI into a Text-to-SQL machine learning project



Goal # 3

- Integrating schema retrieval into the Text-2-SQL pipeline.
- Optimization and Deployment.

Roadmap

Week 1: Fundamentals of NLP & SQL

- Text preprocessing (Tokenization, Stemming, Lemmatization)
- Named Entity Recognition (NER) and Part of Speech (POS) tagging.
- Word Embeddings (Word2Vec, TF-IDF).
- DML (Select, Insert, Update, Delete), Joins, sub queries, aggregate functions and normalization.



Week 2: Introduction to LLMs and Neural Learning

- Basics of transformers (encoder-decoder architecture).
- Popular models for Text-to-SQL (T5, BERT, GPT, BART).
- Fine-tuning pretrained models for SQL generation.
- Hugging Face's transformers library.



Week 4: Model Fine-tuning and Evaluation

- Dataset:
- Avoiding overfitting.
- Error analysis (why the model generates incorrect SQL).



Week 3: Retrieval-Augmented Generation (RAG)

- Integrating schema retrieval into the Text-2-SQL pipeline.
- Vector databases (FAISS, Pinecone).

Week 5: Building the End-to-End System

- Flask/ Fast API for the API layer.
- Integrating with MySQL/Postgres SQL to retrieve schema dynamically.
- Simple UI to accept natural language queries.
- Display generated SQL and query results.



Week 6: Optimization and Testing

- Peephole optimization for SQL queries.
- Caching frequently used queries.
- Unit tests for model and API.
- Load testing for multiple queries.



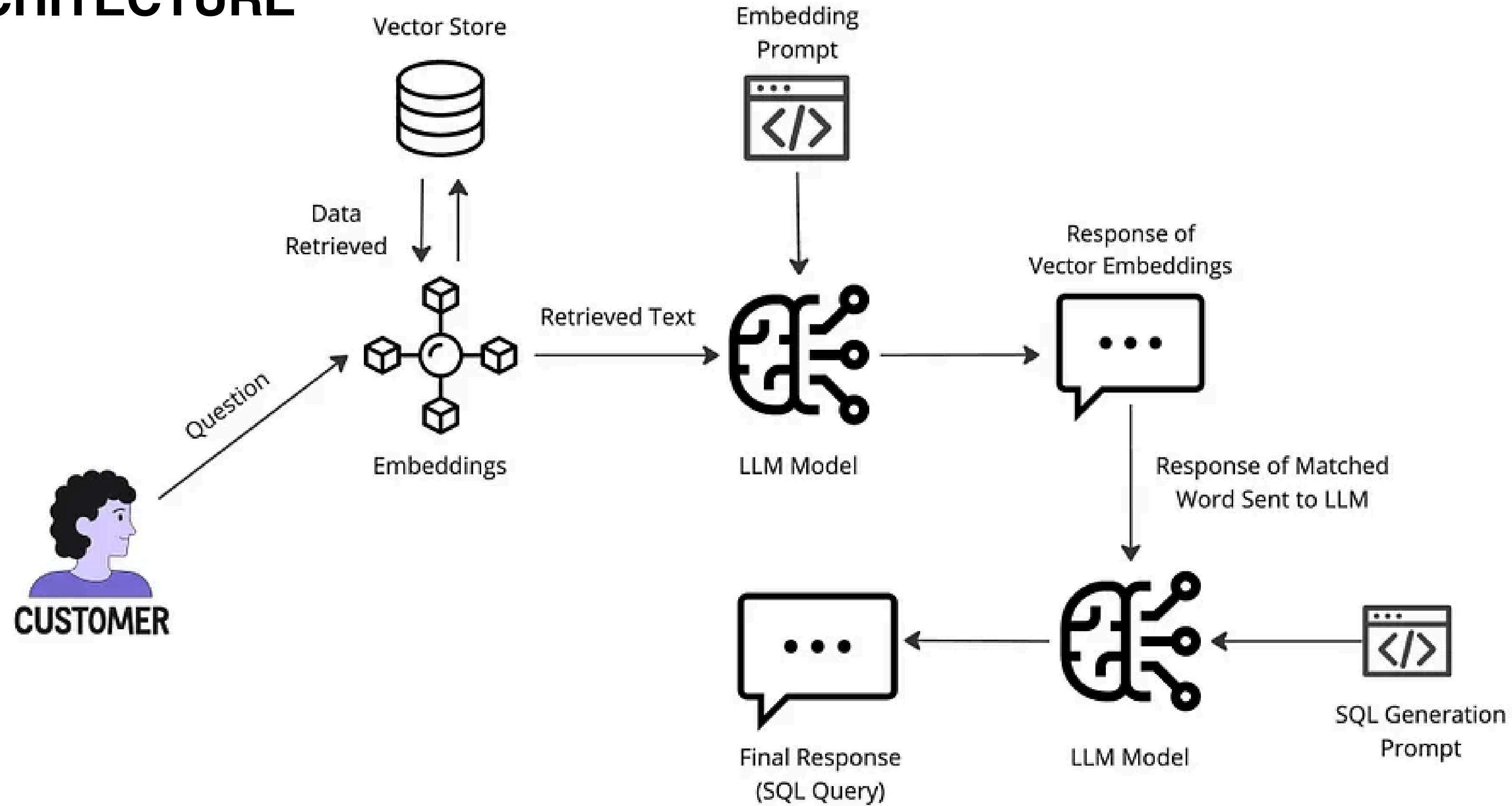
Week 7–8: Final Project Polish & Deployment

- Dockerize the app.
- Deploy on cloud (Heroku, AWS, or GCP).
- Model performance tuning.
- Create a project report and presentation.

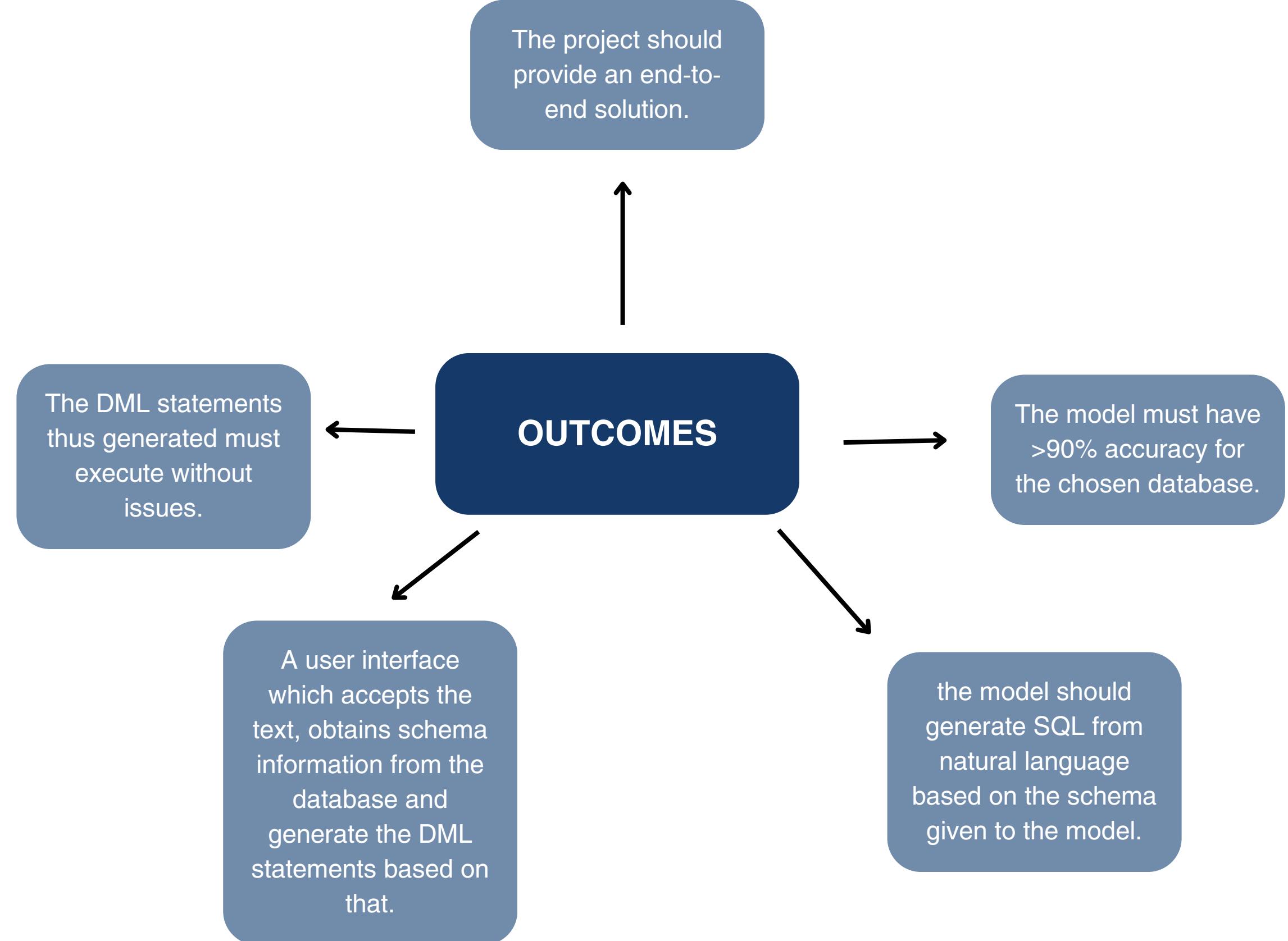
Development Cycle

- ◆ Problem Definition & Requirement Gathering
- ◆ Data Collection & Preprocessing
- ◆ Model Selection & Training
- ◆ Database Connectivity & Schema Understanding
- ◆ Query Generation & Optimization
- ◆ Data Collection & Preprocessing
- ◆ Deployment & Integration
- ◆ Monitoring & Continuous Improvement

ARCHITECTURE



OUTCOME OF PROJECT



Initial Prototype



Convert Natural Language to SQL Queries

Enter your question:

find all section of employee

Ask the Question

Generated SQL Query:

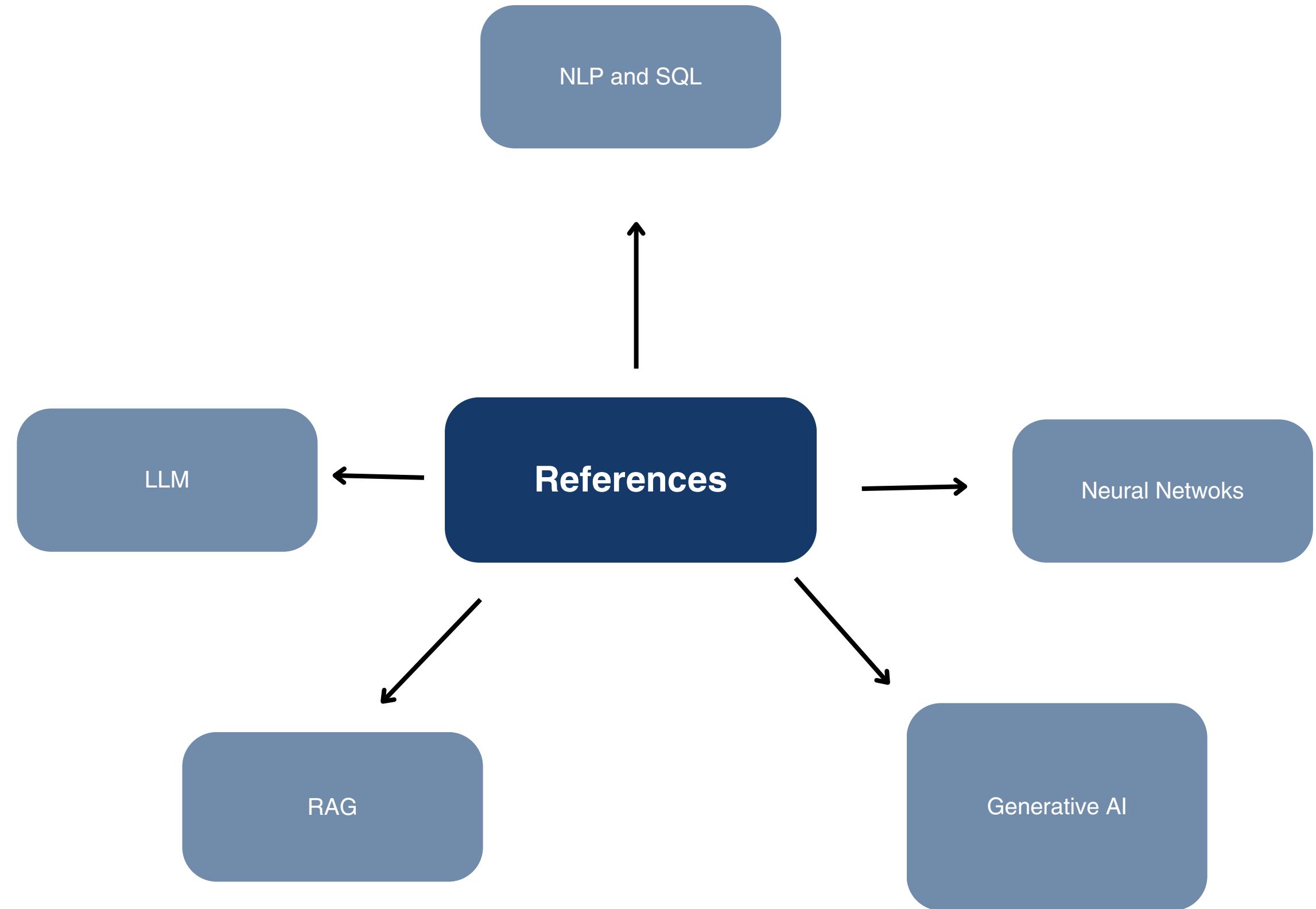
```
SELECT DISTINCT SECTION FROM employee;
```

Query Results:

('Systems Software',)

('Intern',)

REFERENCES



THANK YOU

Team Members:

- Raunak Jain
- Anshul Suwalka
- Vidhi Sharma
- Vishesh Jain
- Priyanshi Sharma