**Report on Implementation of RAG Chatbot**

Yue Wang

[Yueyuelala19@gmail.com](mailto:Yueyuelala19@gmail.com)

**Overview:**

This project involves building a RAG (Retrieval-Augmented Generation) chatbot designed to provide wedding music recommendation.

**Tech Stack:**

* **Programming Language:** Python
* **Vector Database:** Pinecone
* **Language Model:** OpenAI GPT-3.5-turbo
* **Workflow Orchestration:** LangChain

**Challenges:**

* Initial uncertainty about how to begin the project
* Confusion regarding data capture and storage in the vector database
* Numerous failures in setting up Milvus/Pinecone/Weaviate
* Issues with setting up the virtual environment
* Module import errors related to pinecone-client
* API key errors

**Overcoming Challenges:**

* Experimented with various prompts to seek help from AI
* Ask assistant from TAs
* Followed tutorials and tested with small datasets
* Consulted Pinecone documentation and community forums for solutions
* Ran small blocks of code incrementally, addressing errors as they arose to keep them manageable