**University of North Texas**

**ADTA 5760-501**

**Natural Language Processing with Artificial Neural Networks**

**Financial Intelligence: A Generative AI-Powered Q&A System for Corporate Financial Analysis**

**Group Members** – Biniam Abebe, Srilekha Aduvala ,Nithin Marpu, Joshua Terrazas

**PART: V Generative AI A&Q-Search System: System Analysis**

1. **Introduction**

The project is based on the development of a generative AI system for a financial institution. The main goal of our project is to assist employees and clients by answering queries, providing personalized advice, and also in generation of the financial summaries. This system leverages advanced LLMs like Gemini 2.0 via the Google Cloud Platform to improve workflow efficiency and also customer service as well.

1. **Problem Statement**

Financial institutions often face challenges in the sectors of customer inquiry management, generation of financial reports, also in providing real-time advisory services. These all tasks are time-consuming and also resource-intensive. Our system will address these problems by integrating a gen AI that will automate the responses, which will assist in reporting and also that will personalize the user interactions as well. Everything was done by maintaining data security and compliance.

1. **System Requirements Analysis (HW 4)**

Our team is developing a generative AI system for a financial company that is to be used to streamline workflow, improve customer satisfaction scores, and support data-driven business decisions. The organization we are working with is interested in the end results that AI can drive, from saving employee time, responding to customer’s queries, offering financial advice, and generating financial reports.  Implementing this generative AI system into the company's system will allow their employees to focus on more difficult tasks while AI manages the background noise that usually gets in the way. The company hopes to give customers a more personalized experience and faster service. It also helps the company stay up to date with technology and modernize its financial services by implementing this AI system.

The business goals for implementing this LLM-based generative AI model include increasing customer interactions, automating workflow, providing financial advice, and generating up-to-date financial reports for internal use. The system is also expected to close the language gap between clients and employees of all linguistic backgrounds. The goal is to seamlessly integrate this AI system into the company’s system without adding to the daily workload.  Instead, the system should alleviate the staff's repetitive workload and allow them to hyper-focus on providing personalized experiences to their clientele. The system will allow financial advisors and staff to interact with the AI system to refine and recommend relevant content.

The group has decided to use Google Cloud Platform (GCP), leveraging Vertex AI to manage, deploy, and fine-tune machine learning workflows. This will be our AI systems platform used to develop the generative AI system. Gemini 2.0 experimental is the large language model (LLM) chosen for this generative AI system.  Gemini 2.0 was selected due to its high performance in relaying useful relevant financial information and generating clear human-like responses for any user. We understand that the AI system must be able to support and provide real-time APIs for the customers to interact, offer secure data, and offer tracking for audit purposes. Additionally, the system should be able to manage natural language understanding and keep updated with real-time data to support business improvements. These features will ensure that the AI system is efficient and delivers high-quality data-driven business results that align with the financial services goals of the company.

* AI platform: Google Cloud Platform (GCP) using Vertex AI
* Large Language Model (LLM): Gemini 2.0 experimental
* Business Requirements:
* Automate daily tasks such as financial reporting/ customer queries
* Provide financial advice/ tips
* Integrate AI system seamlessly into workflow
* Enhance productivity by alleviating staff workload
* Technical Requirements:
* Real-time APIs for customer interaction
* Data security
* Natural Language Processing (NLP) capabilities
* Real-time data for business improvement
* Data Requirements:
* Use of Google Cloud Storage (adta5770-docs-folder98)
* Subfolders for document management (PDFs)
* PDF documents; These are used for vector embedding and semantic search as well.

1. **Feasibility Analysis**

**4.1 Technical Feasibility:**

**i. Can we complete the project successfully as required?**

Yes. The project can be successfully completed on time. With GCP’s Vertex AI usage, Gemini 2.0, and Langchain’s embedding models.

**ii. Discuss any technical risks while working on the project.**

Initial usage of GCP and Lang chains libraries.

**4.2 Business Feasibility:**

**i. Will the project provide good business value after its completion?**

Streamline daily workflow, Reduce employee workload, Improve customer satisfaction scorecards, Personalize customer experience, and also modernize the operations.

**ii. Discuss any financial risks while working on the project, e.g., running out of funding.**

Nope.

**4.3 Operational Feasibility:**

* + 1. **If we build the system, will it be used by the organization as expected?**

Yes. The organization can use this and will benefit in streamlining daily workflow, Reducing employee workload, Improving customer satisfaction scorecards, Personalizing the customer experience, and also modernize the operations.

* + 1. **Discuss any risks that may hinder the system’s deployment after its completion**

User resistance, and system downtime.

1. **Project Management**

**i. Timeline and Phases:**

**1/13 to 2/09:** System Design and Setup (GCP buckets, access roles)

**2/09 to 2/23:** Document Collection and Preprocessing

**2/23 to 3/30:** Embedding Generation, Search, and Rag Pipe Line Integration

**3/31 to 4/13:** Testing and Debugging

**4/14 to 4/28:** Deployment and Evaluation

**ii. Human Resources:**

4 people were assigned to do the project. The tasks are split among us as follows:

Biniam Abebe: System Design and Setup, Document Collection and Preprocessing, Testing and Debugging, Deployment and Evaluation

Srilekha Aduvala: Search, and Rag Pipe Line Integration, Testing and Debugging, Deployment and Evaluation

Nithin Marpu: Search, and Rag Pipe Line Integration, Testing and Debugging, Deployment and Evaluation

Joshua Terrazas:System Design and Setup,Embedding Generation, Testing and Debugging, Deployment and Evaluation

1. **Conclusion**

This project is ambitious, but by leveraging the GCP and Gemini 2.0 with dedicated team effort we believe that it can significantly improve the experience of client servicing and also the employee productivity within the finance industry.