AppGENie

GenAI Smart Solutions

RAG-Powered PDF Chatbot with Gradio and Python

Project Report

Prepared by
Team NeuralNexus

Aptech North Karachi,
Pakistan

September 23, 2024

Development of ApGENie – GenAI Smart Solutions

Contents

1 Introduction

ApGENie is an intelligent question-answer application designed to revolutionize how customer queries related to different loan types are processed. Using Generative AI (GenAI), Large Language Models (LLMs), and advanced Natural Language Processing (NLP) techniques, the application aims to automate responses to diverse user queries regarding housing, vehicle, education, and medical loans. This detailed report covers the project background, solution, scope, constraints, functional and non-functional requirements, as well as the project deliverables. Additionally, this report will address the challenges encountered and how ApGENie mitigates or overcomes these challenges.

2 Background and Necessity for the Application

The financial services industry is undergoing a paradigm shift as advancements in artificial intelligence (AI) offer new possibilities for improving customer experience and operational efficiency. Traditional customer service approaches for handling loan-related queries are often labor-intensive, prone to human error, and can result in delays that negatively impact customer satisfaction.

Customers often have diverse and complex questions about their loans, ranging from interest rates, repayment terms, documentation, and loan application status, among others. These queries can overwhelm support staff, leading to slower response times and inaccuracies.

Problem Statement: The current manual processes are inefficient, error-prone, and time-consuming, which necessitates a more automated, intelligent, and scalable solution for handling customer queries.

3 Proposed Solution

ApGENie addresses the limitations of traditional query handling methods by leveraging GenAI and LLM technologies to automate customer service. The application integrates Optical Character Recognition (OCR) for extracting data from loan-related PDFs, and NLP to interpret and respond to customer inquiries in real-time.

Key components of the solution are outlined below:

- **User Input Handling**: ApGENie accepts PDF documents containing structured and unstructured data related to housing, vehicle, education, and medical loans. Users can upload these documents, which are processed using OCR to extract relevant text.
- **Optical Character Recognition (OCR)**: OCR is utilized to extract loan-related data from PDF files. This data may include tables, clauses, and conditions that need to be accurately digitized.
- **Natural Language Processing (NLP)**: NLP technologies help analyze the extracted data. Tokenization, Named Entity Recognition (NER), and parsing allow the system to understand complex queries by identifying critical terms and concepts related to loans.
- **Text Preprocessing**: The extracted data undergoes preprocessing to remove irrelevant information, noise, and formatting issues. This step is crucial for ensuring clean, usable data for downstream processing.
- **Text Embeddings**: Using pre-trained models, such as those available from OpenAI, text chunks are converted into dense vector embeddings. This vector representation captures semantic meaning and context, enabling effective information retrieval from a large dataset.

- **Query Processing**: ApGENie processes user queries by comparing them against the loan data stored as embeddings. It understands the intent and context of the query and uses advanced search and ranking mechanisms to find the most relevant answers.
- **Response Generation**: Using GenAI, ApGENie generates accurate and contextually appropriate responses to user queries. These responses are tailored to the specific questions asked and are presented to the user via a user-friendly interface.
- **User Interface**: The application features an interactive and intuitive interface that allows users to easily upload documents and ask questions. ApGENie responds in real-time, providing clear, concise answers.

4 Scope of the Project

ApGENie aims to transform customer service within the financial industry by automating responses to queries related to housing, education, vehicle, and medical loans. The system processes PDFs, extracts relevant loan data using OCR, and responds to user queries via NLP.

Primary Features:

- Automated document scanning and text extraction using OCR.
- Data preprocessing for noise reduction and consistency.
- Text chunking and embedding generation using state-of-the-art LLMs.
- Query processing, re-ranking, and response generation.
- Seamless user interaction through an intuitive web-based interface.

ApGENie is designed to be scalable, capable of handling a growing number of users and documents, and adaptable to different loan formats and customer needs.

Challenges Overcome: 1. **Diverse Document Formats**: Loan documents come in varied formats and layouts. ApGENie's OCR and NLP pipeline is designed to handle inconsistencies in formatting, ensuring accurate text extraction.

2. **Complex Queries**: Customer queries are often multifaceted, involving several loan-related factors. ApGENie's chunking and embedding techniques ensure that complex queries are broken down and matched with the most relevant loan data.

5 Constraints

The development and deployment of ApGENie face several constraints:

- **Handling Multiple Document Formats**: The application must handle different document layouts, which can impact the accuracy of OCR. ApGENie has been designed to cope with inconsistent document structures through advanced preprocessing techniques, but errors in OCR can still occur.
- **Security and Privacy**: As ApGENie processes sensitive loan-related documents, security and privacy are major concerns. The application must comply with data protection regulations such as GDPR to prevent unauthorized access to personal data.
- **Accuracy of OCR and NLP**: The accuracy of OCR and NLP processes is vital for ensuring correct data extraction and query resolution. OCR errors can lead to inaccurate data extraction, while NLP misinterpretations can result in incorrect responses.

- **Processing Speed**: ApGENie must provide real-time responses even when handling large PDF documents. The system's performance depends on the efficiency of the OCR, NLP, and search algorithms.

6 Challenges Overcome

1. **OCR Accuracy**: ApGENie enhances OCR accuracy through preprocessing and error correction techniques that minimize noise and formatting inconsistencies.
2. **Data Security**: ApGENie implements encryption and secure storage mechanisms to protect sensitive loan-related data from unauthorized access.
3. **Speed Optimization**: By employing advanced vector-based search techniques, ApGENie optimizes query matching, ensuring fast response times even with large datasets.

7 Functional Requirements

ApGENie's functional requirements ensure that the system meets the needs of users while performing its tasks efficiently:

(a) **PDF Input Handling**: Users can upload PDF documents that contain loan-related data.

(b) **Text Preprocessing**: The uploaded text is cleaned and structured for better processing.

(c) **Text Chunking and Embedding**: Extracted data is broken into manageable chunks, which are converted into vector embeddings using pre-trained models.

(d) **Query Processing**: User queries are processed by the system, with the most relevant data being retrieved and used to generate a response.

(e) **Re-ranking and Search**: The system re-ranks retrieved chunks based on their relevance to the query and generates a coherent response.

(f) **Response Delivery**: The final response is displayed to the user through an interactive interface.

8 Non-Functional Requirements

The following non-functional requirements ensure the quality and usability of ApGENie:

- **Scalability**: ApGENie must be able to handle increasing numbers of users and documents without significant degradation in performance.
- **Security**: The system must prevent unauthorized users from accessing sensitive loan data.
- **Performance**: ApGENie must process and respond to queries quickly, even when dealing with large PDF documents.
- **Usability**: The interface must be intuitive and accessible, making it easy for users of all technical backgrounds to interact with the system.
- **Accuracy**: The application must maintain high levels of accuracy in OCR and NLP to ensure the correct handling of loan data and customer queries.

Challenges Overcome: 1. **Scalability**: ApGENie leverages distributed computing resources to ensure that it can scale with increasing demands. 2. **Usability**: By designing an intuitive and simple user interface, ApGENie minimizes the learning curve for users, allowing for seamless interaction with the system.

9 Interface Requirements

9.1 Hardware

ApGENie requires the following hardware for optimal performance:

- Intel Core i5/i7 processor or higher.
- 8 GB RAM or higher.
- 500 GB hard disk space.
- SVGA color display.
- Standard keyboard and mouse.

9.2 Software

- **Frontend**: Gradio and Hugging Face
- **Backend**: Python and Google Colab.
- **Libraries**: torch, langchain, transformers, sentence-transformers, langchain-community , tqdm, accelerate , pypdf and faiss-gpu

10 Project Deliverables

The deliverables for ApGENie include:

- **Problem Definition**: The financial services industry is evolving rapidly, with increasing demands for faster, more accurate, and personalized customer service. Traditional methods of handling and processing customer queries related to loans—such as housing loans, education loans, vehicle loans, and medical loans—are labor-intensive, time-consuming, and prone to errors. These manual processes often lead to customer dissatisfaction due to delays and inaccuracies, resulting in operational inefficiencies for financial institutions.

  Moreover, the complexity and diversity of loan-related queries, coupled with the growing volume of customer inquiries, place an additional strain on customer service teams. Financial institutions need a solution that can handle large volumes of data, ensure accurate and timely responses, and deliver personalized customer service at scale.

  In this context, leveraging Generative AI (GenAI) and Large Language Models (LLMs) presents a transformative opportunity. By automating the processing of customer queries using advanced AI technologies, financial institutions can streamline operations and significantly improve the customer experience.
- **System Architecture** ApGENie employs a modular architecture, composed of several interconnected components designed to handle customer queries related to loans efficiently and accurately. The key components are:

  - **User Interface (UI)**: A web-based platform that allows users to interact with the system, upload documents, and submit queries.
  - **Backend Server**: Manages the business logic, handles queries, and interacts with the database. It uses frameworks such as Flask or Django.
  - **Database**: Stores extracted data from the loan documents and user-related information. A relational database like PostgreSQL is recommended.
  - **NLP Module**: Analyzes and processes the user's natural language queries.
  - **Embedding Module**: Converts text data into vector representations using pre-trained models.
  - **Query Processing Engine**: Matches user queries with relevant data and generates responses.
- **User Flow Diagrams**: Visual representations of how users will interact with ApGENie.

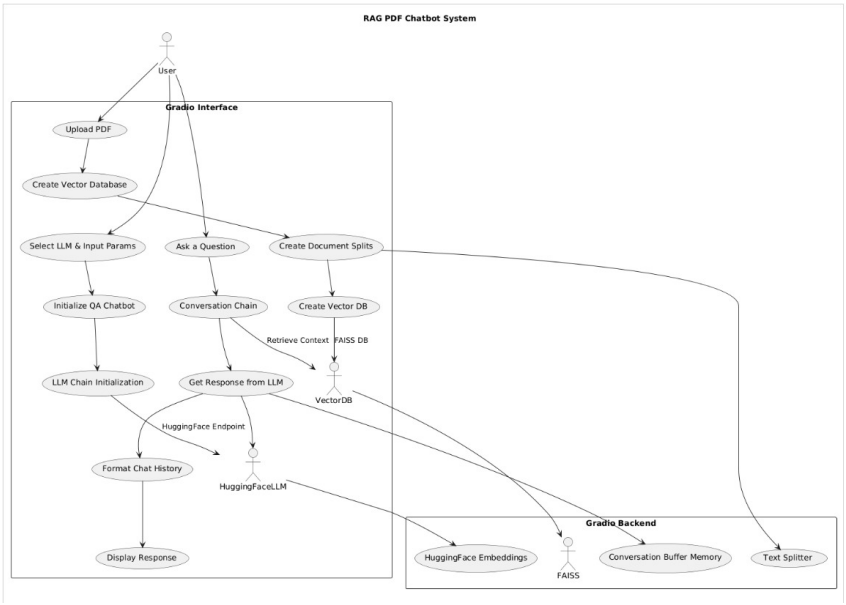

Figure 1: Flow Chart

- **Execution Steps**: Instructions for running the application.
- **Installation Instructions**: Detailed installation steps.
- **GitHub Repository**: The complete source code hosted on GitHub, with a ReadMe file.
- **Blog Post**: Here is the link where yo can read full blog: https://viroshieldinsights.blogspot.com/2024/09/a-rag-powered-pdf-chatbot-with-gradio.html
- **Demo Video**: A video demonstrating the application's functionality.

11 Conclusion

ApGENie addresses the limitations of traditional loan-related customer service by providing a scalable, accurate, and efficient AI-powered solution. Using advanced OCR, NLP, and GenAI technologies, it automates query handling, significantly improving customer satisfaction and operational efficiency. By overcoming challenges such as document format variability, query complexity, and processing speed, ApGENie sets a new standard for intelligent customer service in the financial industry.

12 Project Working

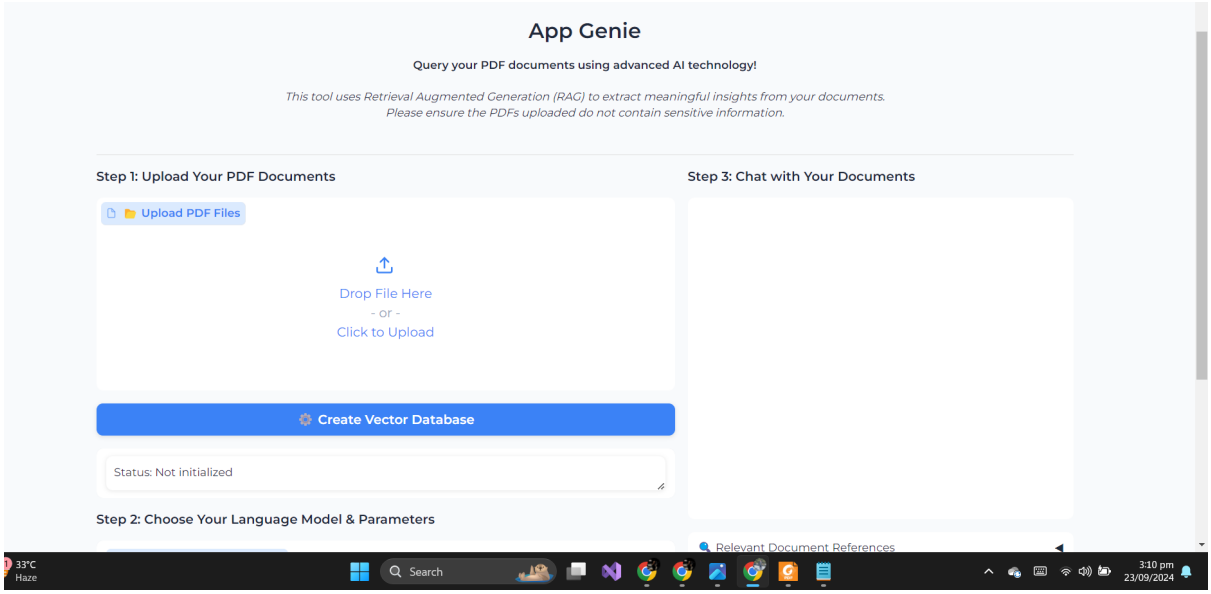

Figure 2: Document loading

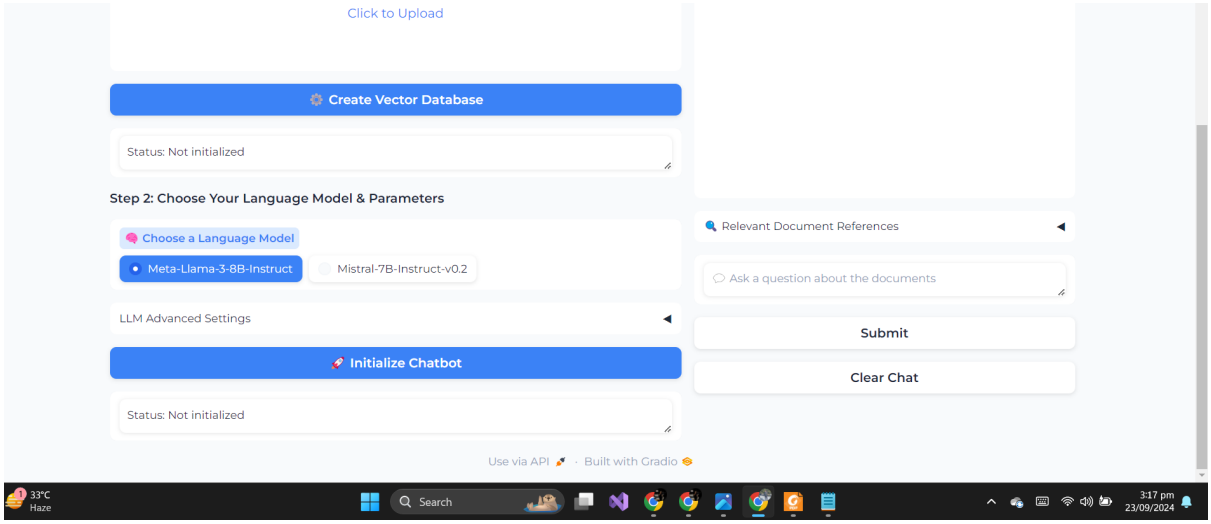

Figure 3: Creating Vector Db and selecting model