IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 1, May 2024

Enterprise-Grade Conversational Intelligence: A Domain-Aware Chatbot Framework using Gpt-3.5, Langchain, And Rag with Local Vector Indexing

Dheerendra Yaganti
Software Developer,
Astir Services LLC, Frisco, Texas.
dheerendra.ygt@gmail.com

Abstract: *This thesis presents a scalable and domain-aware chatbot framework that integrates GPT-3.5 with LangChain and Retrieval-Augmented Generation (RAG) to deliver context-sensitive responses grounded in enterprise-specific knowledge. The proposed architecture leverages local vector databases, including FAISS and Chroma, to perform efficient semantic retrieval from proprietary document repositories. By embedding domain documents into high-dimensional vector space and linking them with transformer-based query models, the system retrieves relevant context passages in real time, enhancing the language model's relevance and accuracy. LangChain orchestrates the interaction between the language model and retrieval components, enabling modular and extensible prompt chains tailored to organizational needs. The framework supports document ingestion in varied formats, including PDFs, Word documents, and structured CSV files, converting them into persistent embeddings for rapid querying. Security and data privacy are maintained through localized storage, ensuring compliance with enterprise governance standards. Experimental evaluations demonstrate significant improvements in factual consistency and contextual relevance across test scenarios in finance, legal, and customer support domains. This work underscores the potential of combining generative AI with vector-based retrieval to build intelligent, responsive assistants for domain-specific enterprise applications.*

Keywords: GPT-3.5, Retrieval-Augmented Generation (RAG), LangChain, Vector Embeddings, Semantic Search, FAISS, Chroma DB, Natural Language Processing (NLP), Information Retrieval

DOI: 10.48175/IJARSCT-18099