

Intelliconverse: An Adaptive LLM-Based Q&A Assistant

Abstract

An innovative system named an Adaptive LLM-Based Q&A Assistant is designed to provide intelligent and contextually aware responses to user queries by leveraging the Gemini large language model (LLM). Intelliconverse leverages Flask and Python for backend processing and secure API key to integrate Gemini's generative capabilities for accurate and dynamic question-answering features. The frontend leverages HTML and JavaScript to create an intuitive user interface that makes it easier for users to interact through general chat and document-specific queries. Users can upload PDF documents, from which Intelliconverse extracts and parses text using PyPDF2 so that, based on the content of the document, well-informed responses are possible. The robust design of the system ensures a stable and adaptable user experience while enabling real-time communication, history management, and error handling.

Introduction

Over the past few years, artificial intelligence (AI) has grown exponentially, particularly in the field of natural language processing (NLP). The capacity of machines to comprehend and produce human language has created new opportunities in a number of industries, including content creation and customer service. Significant progress has been made in this field thanks to generative models, like Google's Gemini and OpenAI's GPT, which enable the development of extremely intelligent conversational agents that can produce language that resembles that of a human depending on the input.

This work presents Intelliconverse, a cutting-edge online application that uses Google's Gemini language model to power two main features: a document-based question-answering system and a conversational chatbot. The purpose of the application is to enable smooth interactions by letting users ask questions regarding specific content in uploaded documents and have talks in natural language. Intelliconverse integrates sophisticated natural language processing (NLP) methods with intuitive user interfaces to provide a flexible tool for general-purpose conversations as well as specialized document analysis.

Background and Motivation

Conversational agents powered by artificial intelligence (AI) are in high demand as people and organizations look for more effective ways to manage massive amounts of textual data, undertake repetitive jobs, and offer immediate customer service. Even though they are helpful, traditional chatbots frequently have trouble comprehending context or answering complicated questions, especially ones that call for the reference of other sources. More sophisticated systems that can process and comprehend large amounts of text from documents in addition to holding conversations are required as a result of this constraint.

Users can upload PDFs and ask queries that are specifically connected to the content of the documents by integrating document analysis capabilities into the chatbot interface provided by Intelliconverse. When customers require rapid and precise responses from vast amounts of text, such as in corporate intelligence, academic research, and legal document assessment, this dual feature comes in handy.

Objectives

The following are the main goals of Intelliconverse:

1. **Conversational AI:** The aim is to develop a conversational agent that is both sensitive and aware of context, enabling users to engage in natural and intuitive interactions.
2. **Document-based Query Handling:** To allow users to upload documents and pose inquiries about the information contained in them, with the system processing and obtaining pertinent data from the text.
3. **User Experience:** to create an interface that is simple to use so that users may manage their interaction history, switch between different functions, and communicate with the AI.

Significance and Contributions

The integration of conversational AI and document analysis has advanced significantly with Intelliconverse. The system provides a distinctive solution for users that require conversational interaction and information extraction from documents by merging these two features. In order to increase textual information handling efficiency and accessibility, Intelliconverse can be used in a variety of real-world circumstances, as this article illustrates through its architecture, implementation, and possible uses.

In conclusion, this study introduces Intelliconverse as a powerful and adaptable tool that increases user interaction with both general conversational queries and document-specific inquiries by utilizing the most recent developments in generative AI. The approach creates new opportunities for AI applications in several sectors by bridging the gap between document processing and chat-based interfaces.

Literature Survey

S. K., Wang et al. [1] LLM-based Conversational Assistants (CAs) outperform intent-based equivalents in Knowledge Management (KM) duties due to their more effective and flexible interactions. Unlike intent-based CAs, which frequently struggle to remain on course, LLM-based CAs thrive at natural language processing, resulting in more successful fulfillment of knowledge management activities such as information retrieval, rating, and sharing. Dam, S. K., Hong et al. [2] This review looks at the progress of LLM-based chatbots, including their

design, applications, and problems. It covers developments in efficiency and ethics, with the goal of guiding conversations about the role of LLMs in the route to AGI.

Park, J. I., Abbasian et al. [3] Advances in LLMs have increased the usage of mental health chatbots, but a comprehensive safety assessment is needed. To address this, 100 benchmark questions and 5 guideline questions were created, coupled with LLM-based tools for expedited evaluation, with the Agent and Embedding techniques demonstrating the best accuracy. Roffo, G. (2024) [4] Important developments in LLMs, including RAG, PAL, and LangChain, are covered in this lesson along with issues including out-of-date information and computational constraints. It focuses on integration frameworks and fine-tuning techniques to increase LLM usability and accuracy.

Buscemi, A., et al. [5] In this study, we evaluated the cross-language emotional nuance handling capabilities of ChatGPT, Gemini Pro, and LLaMA2. These models are good at predicting sentiment in general, but they have biases and have trouble understanding irony and sarcasm. The results highlight the need for better techniques to deal with these issues in practical applications.

References

1. Freire, S. K., Wang, C., & Niforatos, E. (2024). Chatbots in knowledge-intensive contexts: Comparing intent and llm-based systems. *arXiv preprint arXiv:2402.04955*.
2. Dam, S. K., Hong, C. S., Qiao, Y., & Zhang, C. (2024). A Complete Survey on LLM-based AI Chatbots. *arXiv preprint arXiv:2406.16937*.
3. Park, J. I., Abbasian, M., Azimi, I., Bounds, D., Jun, A., Han, J., ... & Rahmani, A. (2024). Building Trust in Mental Health Chatbots: Safety Metrics and LLM-Based Evaluation Tools. *arXiv preprint arXiv:2408.04650*.
4. Roffo, G. (2024). Exploring Advanced Large Language Models with LLMsuite. *arXiv preprint arXiv:2407.12036*.
5. Buscemi, A., & Proverbio, D. (2024). Chatgpt vs gemini vs llama on multilingual sentiment analysis. *arXiv preprint arXiv:2402.01715*.