LLM ENHANCED AI CHATBOT

Antony Roshan R, IV year CST, SNS College of Engineering, Coimbatore. Email : roshan94425@gmail.com
Barath V, IV year CST, SNS College of Engineering, Coimbatore. Email : bharathvadivel2004@gmail.com
Deva Dharshini D, IV year CST, SNS College of Engineering, Coimbatore. Email : dharshinideva231@gmail.com
Dhilak M, IV year CST, SNS College of Engineering, Coimbatore. Email : dhushilak@gmail.com
Saraswathi R, Assistant Professor, SNS College of Engineering, Email: sarasrajagopal2@gmail.com

Abstract
The development and deployment of a Large Language Model (LLM)-based tool designed to generate human-like responses to natural language inputs in a network isolated environment presents unique technical and logistical challenges. Such a tool leverages state-of-the-art Natural Language Processing (NLP) and machine learning techniques to simulate real-time, coherent, and contextually appropriate interactions without relying on an active internet connection. This approach involves training or fine-tuning a pre-existing LLM on domain-specific datasets and configuring the model to operate efficiently on limited resources, addressing the constraints of an offline deployment. The solution provides an autonomous conversational agent capable of supports text summarization, upload CSV and PDF files. The tools used for this project Natural Language Processing(NLP), Hugging face. Key components include model optimization to reduce computational overhead, storage requirements, and latency in response generation. Future scope includes improving model robustness and incorporating ethical guardrails to ensure the responsible use of LLMs, especially in sensitive fields such as healthcare, education, and customer support.

Introduction
In recent years, Large Language Models (LLMs) have revolutionized the way machines interact with human language. These models, trained on vast amounts of textual data, can generate human-like responses to a variety of natural language inputs, offering a wide range of applications, from customer support to educational tools. However, a significant challenge arises when attempting to deploy such models in environments where internet connectivity is limited or entirely absent. This tool will provide users with the benefits of advanced natural language understanding and conversational capabilities while ensuring data privacy, faster response times, and accessibility in network-restricted or isolated environments. The project will focus on optimizing the model for local deployment, addressing hardware constraints, and delivering high-quality conversational AI experiences without relying on cloud-based resources.
This project aims to develop and deploy a Large Language Model (LLM) based tool that generates human-like responses to natural language inputs without requiring an internet connection. By leveraging advanced machine learning algorithms and pretrained language models, the system is designed to operate in an offline environment. This innovation focuses on enhancing user interactions with AI, making it accessible even in areas with limited or no internet connectivity. Additionally, the system supports functionalities like text summarization and file handling for CSV and PDF formats. It combines a PDF uploader, CSV uploader, text summarizer, URL summarizer, code summarizer, and code debugger is extensive and applicable across multiple industries and use cases. This versatility allows the project to serve users in education, software development, research, content management, and data analysis, providing valuable tools for automating workflows, improving productivity, and enhancing accuracy. Extracting text from PDFs and parsing CSV files accurately can be challenging. PDF files vary widely in formatting and may contain images, tables, and other non-text elements that make it difficult to extract clean, readable text. Summarization, both for general text and URLs, relies on advanced NLP models capable

of understanding and generating coherent summaries. Effective summarization is complex and requires balancing readability with information retention, especially for large documents or URLs with mixed content. The code summarizer and debugger components require language-specific parsers and tools to accurately interpret code 4 structures, identify functions, and locate errors.

Literature Review

A systematic literature review of Chatbot applications in education. A systematic review protocol was used to analyse 53 articles from well-known digital databases such as Scopus, IEEE, ERIC, ScienceDirect, SpringerLink, and Taylor and Francis. The findings explained the current state of knowledge on Chatbot applications in education and identified the benefits and challenges of using Chatbots in education. Furthermore, future areas of education that may benefit from this modern AI technology were identified and explained as recommended by the literature. Finally, the research findings were discussed in terms of their implications.[1]

The proposed future research directions are intended as a response to the current lack of coherence in the emerging field of chatbot research, which may in part be observed by the broad range of journals and conferences in which findings from chatbot research are presented, and also the lack of commonly agreed key constructs, models, and measurement instruments. While this may be expected in an emerging research area, future research will benefit from a greater degree of coherence and cohesiveness in the field.[2]

A chatbot is one of the innovative ways to exchange information from a machine or system because individuals might ask questions and get answers without being required to think of the appropriate search queries or navigate through with a number of websites. This paper has provided details on the creation and maintenance of the chatbot. From the review, it can be concluded that the growth and improvement of chatbot design expand at a variable rate since a chatbot may be designed using a number of styles and methodologies. A chatbot is an excellent tool for conducting speedy business with stoners. They assist us by providing amusement, saving time, and providing answers to issues that are difficult to locate.[3]

As a result, it's critical to examine and analyze the data used to train the various models. This type of study provides for a more accurate comparison of different models and their results. In fact, the distinction between chatbots' applications and social or companion chatbots appears to be hazy. Chatbot modeling is a fascinating challenge that mixes Deep Learning and Natural Language Processing. Despite the fact that the first chatbots were created sixty years ago, the area has continued to grow and provide new and exciting problems. To bridge these gaps, smaller, flexible, less domain dependent models would be beneficial. Improved, scalable, and flexible language models for industry specific applications, more human-like model architectures, and improved evaluation frameworks would surely represent great steps forward in the 10 field.[5][7]

Existing Approach

The existing systems available for this project leverage advanced Natural Language Processing (NLP) and machine learning techniques to offer robust functionalities such as text summarization, link summarization, file processing, and programming assistance. Text summarization systems use pre-trained models like OpenAI GPT or BART to generate concise and accurate summaries of long texts, enabling quick comprehension.

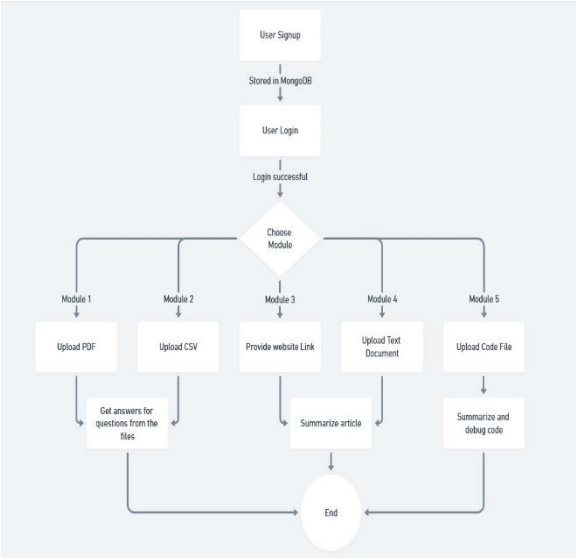
Link summarizers extract key information from shared URLs, providing brief overviews without requiring users to open the links. File processing systems utilize tools like PyPDF2 and pandas to extract, analyze, and summarize data from PDF and CSV files, catering to diverse use cases like data extraction and query-based retrieval. Additionally, code summarization and debugging solutions powered by models like OpenAI Codex and CodeT5 analyze code snippets to provide functional summaries and error identification. These systems integrate seamlessly with user interfaces, offering interactive and user-friendly experiences for real-time assistance and data processing.

Proposed Approach

The proposed project aims to develop an intelligent chatbot system that combines advanced Natural Language Processing (NLP) and machine learning capabilities to enhance user interaction and data processing. The chatbot will support multiple functionalities, including text summarization, link summarization, and data analysis through file uploads (PDFs and CSVs). It will provide users with concise, accurate summaries of long texts and shared links, helping them quickly grasp key information.

Additionally, users can upload files for processing, with the chatbot extracting and analyzing the data for tasks such as answering queries, generating summaries, or highlighting key insights. The system will also include a code summarizer and debugger to assist programmers in understanding code logic and identifying errors in real time. Built with scalable infrastructure using MongoDB for data storage and streamlit ui for hosting, the chatbot will provide a versatile and user-friendly solution, making it applicable across various domains, such as education, research, and business intelligence.

Flow diagram



List of modules and its working:

1. Login Module

The sign-in and sign-up modules enable new users to register on the platform. The login and logout functionality allows users to securely access their accounts.

2. Text Summarizer

The chatbot can generate concise and accurate summaries of long texts, helping users quickly understand large volumes of information. The text summarizer can better understand the user's needs and respond concisely.

3. Link summarizer

Link summarization involves extracting the main points from the content of a shared link to present a brief, relevant summary. When a user shares a URL, the chatbot can pull key details like article highlights, core themes, or main finding is allowing users to get an overview without leaving the chat.

4. PDF File Uploader

Users can upload PDF documents and the chatbot will process and analyze the data, making it versatile for various use cases such as data extraction, analysis, and summarization.

5. CSV File uploader

Users can upload CSV files and the chatbot will process and analyze the data, making it versatile for various use cases such as data extraction, analysis, and summarization.

6. Code Summarizer & Debugger

A code summarizer in a chatbot provides concise explanations of programming code shared by users. Using NLP and machine learning models, the chatbot can analyze the code's structure, functions, and logic to generate a brief, understandable summary. This helps users quickly grasp the purpose and functionality of the code without going through each line in detail. A code debugger in a chatbot helps identify and explain errors or bugs in code snippets shared by users. By analyzing the syntax and logic of the code, the chatbot can detect common issues, such as syntax errors, undefined variables, or logic flaws, and provide recommendations for fixes. This is particularly valuable in programming assistance, as it offers real-time debugging support and guidance.

Result

This LLM Enhanced AI Chatbot aims to encompassing a PDF uploader, CSV uploader, text summarizer, URL summarizer, code summarizer, and code debugger, demonstrates remarkable versatility and utility across various industries. By integrating these functionalities, it can streamline workflows, improve productivity, and boost accuracy for users in diverse sectors such as education, software development, research, content management, and data analysis. In education, the tool can assist students and educators by enabling easy access to summarized academic papers, research articles, and lecture notes. For software development, the project's code debugger and code summarizer provide critical assistance. Developers can not only identify and resolve code issues but also get concise summaries of code logic, facilitating better code comprehension and documentation. In the research field, the tool aids researchers by simplifying the process of handling large datasets and text-heavy content. The PDF and CSV uploaders support data-driven projects, allowing researchers to analyze, summarize, and interpret data or documents more effectively. the project's ability to automate various tasks and integrate multiple functionalities enhances its adaptability and usefulness..

Conclusion

In conclusion, It helps in education, the tool can assist students and educators by enabling easy access to summarized academic papers, research articles, and lecture notes. For software development, the project's code debugger and code summarizer provide critical assistance. Developers can not only identify and resolve code issues but also get concise summaries of code logic, facilitating better code comprehension and documentation. In the research field, the tool aids researchers by simplifying the process of handling large datasets and text-heavy content. The PDF and CSV uploaders support data-driven projects, allowing researchers to analyze, summarize, and interpret data or documents more effectively. the project's ability to automate various tasks and integrate multiple functionalities enhances its adaptability and usefulness.

Future Work

Future work of this Chatbot could include developing a language translation feature to chatbots opens up significant possibilities, making them more inclusive, accessible, and versatile in a globalized environment. Enable chatbots to work offline with a language translation feature, they would need specialized design considerations, particularly in storage, processing power, and efficiency.

References

[1] Chinedu Wilfred Okonkwo, "Chatbots applications in education", 2023

[2] Asbjorn Folstad and Theo Araujo, "Future Directions for Chatbot Research", 2021

[3] Ajay Sahu, Suyash Pandey, Mahak Agarwal, and Shivam Chauhan,"Offline Virtual Chat Bot by Using Natural Language Processing", 2023

[4] Filippo Chiarello, Vito Giordano, and Irene Spada, "Future applications of generative large language models", 2024

[5] Guendalina Caldarini, Sardar Jaf, and Kenneth McGarry, "A Literature Survey of Recent Advances in Chatbots", 2022

[6] Jan Clusmann, Fiona R. Kolbinger, Hannah Sophie Muti, and Zunamys I Carrero "The Future Landscape of Large Language Models in Medicine", 2023

[7] Chiara Valentina Misischia, Flora Poecze, Christine Strauss, "Chatbots in customer service", 2022

[8] Hyojin Chin, Hyeonho Song, Gumhee Baek, Mingi Shin "The Potential of Chatbots for Emotional Support and Promoting Mental Well-Being", 2023

[9] Chris Zielinski , Margaret A. Winker , Rakesh Aggarwal "Chatbots, generative AI, and scholarly manuscripts", 2023

[10] Rona Joy Corpuz Arangote , Teresa Abbygail Manigo Dela Paz , Princess Marie Gera Vinarao "Offline-Online Submission and Assessment System (OSAS) chatbot", 2022

[11] Ajay Sahu, Suyash Pandey, Mahak Agarwal "Offline Virtual Chat Bot by Using Natural Language Processing", 2023

[12] Marcello M. Mariani, Novin Hashemi, Jochen Wirtz, "Artificial intelligence empowered conversational agents", 2023