World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(Research Article)

Human challenging career and future advice chatbot using RAG

Nagur Vali Shaik, Lokesh Karthik Penmetsa [*], Praisey Bathula, Spandana Salveru and Sahas Manikanta Madishetty

Department of Computer Science and Engineering (Data Science), ACE Engineering College, Telangana, India.

World Journal of Advanced Engineering Technology and Sciences, 2025, 15(02), 428-435

Publication history: Received on 25 March 2025; revised on 02 May 2025; accepted on 04 May 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.15.2.0598

Abstract

Choosing the right career path is a crucial decision for students and job seekers, often influenced by a lack of proper guidance or access to reliable information. This paper presents a Career Guidance Chatbot that leverages Retrieval-Augmented Generation (RAG) and the Zephyr language model to provide personalized career advice in a conversational manner. The system retrieves relevant data from curated documents and integrates it with the chatbot's natural language understanding to deliver meaningful suggestions. Built using Stream-Lit, the chatbot offers an interactive, user-friendly interface that simulates real-time conversations. This approach bridges the gap between static career counseling methods and dynamic AI-driven support, offering users a more engaging and accessible platform for decision-making. The chatbot not only simplifies complex career information but also tailor's guidance based on user preferences and academic background. This solution demonstrates how AI can effectively assist in shaping informed career decisions.

Keywords: Retrieval-Augmented Generation for Career Advice; Zephyr-Powered Conversational Chatbot; Interactive Stream-Lit Chat Interface

1. Introduction

The Career and Future Advice Chatbot is designed to help students make informed decisions about their future, particularly after completing their 10th and 12th grades. With so many career options and fast-changing industry demands, students often struggle to find relevant and timely guidance. This project aims to bridge that gap by providing instant, personalized support.

Traditional career counseling methods, while useful, are often not accessible to all students and may lack personalization. In many cases, especially in rural or under-resourced areas, students do not have access to professional advice. This chatbot serves as a virtual assistant that offers reliable, real-time career suggestions tailored to each user.

The chatbot uses Retrieval-Augmented Generation (RAG), a hybrid technique that first retrieves relevant documents based on the user's query and then generates context-aware responses. This method ensures that the information provided is both accurate and easy to understand.

Key technologies used in this project include Lang-Chain for managing document flows, FAISS for fast vector-based search, and Hugging-Face Sentence Transformers for generating text embeddings. The Zephyr-7B language model is responsible for producing conversational responses, ensuring clarity and relevance in each interaction.

The user interface is built with Stream-lit, making it lightweight and easy to access. Users can ask about different career paths, entrance exams, necessary skills, and study options. The system refers to documents like after-10.txt and

[*] Corresponding author: Lokesh Karthik Penmetsa

after-12.txt, which are curated with accurate and updated career information. The interface provides a chat-like experience where students can interact naturally and receive responses instantly.

Overall, this project provides a scalable and accessible solution to guide students in their career journey. By combining retrieval and generation, it offers meaningful, personalized support that adapts to evolving trends, helping students confidently plan their academic and professional futures.

2. Literature review

Career guidance has traditionally relied on manual counselling and static resources, but these methods often fail to meet the needs of today's dynamic job market. Recent research emphasizes the limitations of traditional systems, which tend to offer generalized advice rather than personalized recommendations. Studies like those by Lewis et al. (2020) highlight the importance of developing more dynamic and tailored career counselling solutions, pointing to the growing need for AI-driven approaches that can adapt to the individual needs of students and the constantly changing job landscape.

AI-powered chatbots are increasingly being used in various sectors, including career counseling. Research by Gartner (2020) demonstrates that chatbots using Natural Language Processing (NLP) can offer immediate, accessible, and personalized career advice. By enabling students to interact naturally, chatbots can provide answers to a wide range of career-related questions, from skill requirements to education paths. This technology makes career counseling more accessible, especially for students who may not have direct access to professional advisors.

The introduction of Retrieval-Augmented Generation (RAG) by Lewis et al. (2020) and Radford et al. (2019) has significantly improved the effectiveness of AI in career guidance. By combining the strengths of information retrieval with generative models, RAG enables systems to deliver more accurate and context-aware responses. This hybrid approach has been successfully applied in career recommendation systems, as demonstrated by Garg et al. (2021), who used AI to suggest career paths based on individual preferences and skill sets. AI-based systems are revolutionizing career counseling, offering personalized, real-time guidance that adapts to the evolving demands of both students and industries.

3. Existing System

Traditional career guidance systems are largely manual and dependent on human counsellors. These systems often provide generic advice based on the counsellor's experience, which can lead to inconsistent and outdated recommendations. They are also limited in scalability, as one counsellor can only assist a small number of users at a time. Moreover, traditional counselling struggles with personalization, real-time updates, and often requires users to schedule in-person appointments, which adds to the inconvenience.

On the other hand, automated career guidance models, like rule-based chatbots and static career quizzes, attempt to streamline the process but fall short in key areas. These systems often lack context awareness, providing generic and static recommendations that fail to address individual user profiles. Additionally, they are unable to learn from user interactions or fetch real-time data, making them rigid and disconnected from evolving job market trends.

To overcome these limitations, a more adaptive and intelligent system is needed. By leveraging Retrieval-Augmented Generation (RAG), a dynamic approach can be developed that fetches the latest, personalized career recommendations based on up-to-date data. This model ensures real-time relevance, improves accuracy, and creates a more engaging and responsive experience for users, helping them navigate their career paths in a rapidly changing world.

4. Proposed System

The proposed AI-powered Career Advice Chatbot, built using Retrieval-Augmented Generation (RAG), offers a smarter, more personalized alternative to traditional career counselling. Unlike outdated systems that give generic suggestions or rely heavily on static quizzes, this chatbot uses real-time data, user profiling, and contextual understanding to deliver tailored guidance. It considers a user's academic background, skills, interests, and goals to recommend career paths, skill development plans, and job roles that align with their journey—whether they're just starting out or ready to enter the workforce. Through multi-turn conversations, it also adapts dynamically to follow-up questions and provides meaningful, stage-specific advice.

Powered by technologies like vector-based search (FAISS), language models (GPT or LLaMA), LangChain, and custom databases, the system is capable of fetching the most relevant content and responding in a conversational, human-like manner.

It's designed for 24/7 access through a web or chatbot interface and supports rich interaction features like comparing careers, follow-up support, and linking to external resources such as resume builders, certification platforms, and job portals. Future enhancements like emotion detection and visual elements will further make the experience more intuitive and engaging. Ultimately, the chatbot acts not just as a tool, but as a companion in the user's career journey—evolving with them, offering timely insights, and making guidance more accessible, interactive, and impactful.

5. Methodology

The methodology section outlines the systematic approach used in designing, developing, and implementing the Multi-modal RAG Chatbot. This project adopts a Retrieval-Augmented Generation (RAG) approach combined with the Zephyr language model to build an intelligent career guidance chatbot. It retrieves relevant information from a vector-based knowledge base and generates personalized, context-aware responses. The system is deployed using a Stream-Lit interface, enabling real-time, multi-turn user interactions for dynamic and adaptive career advice.

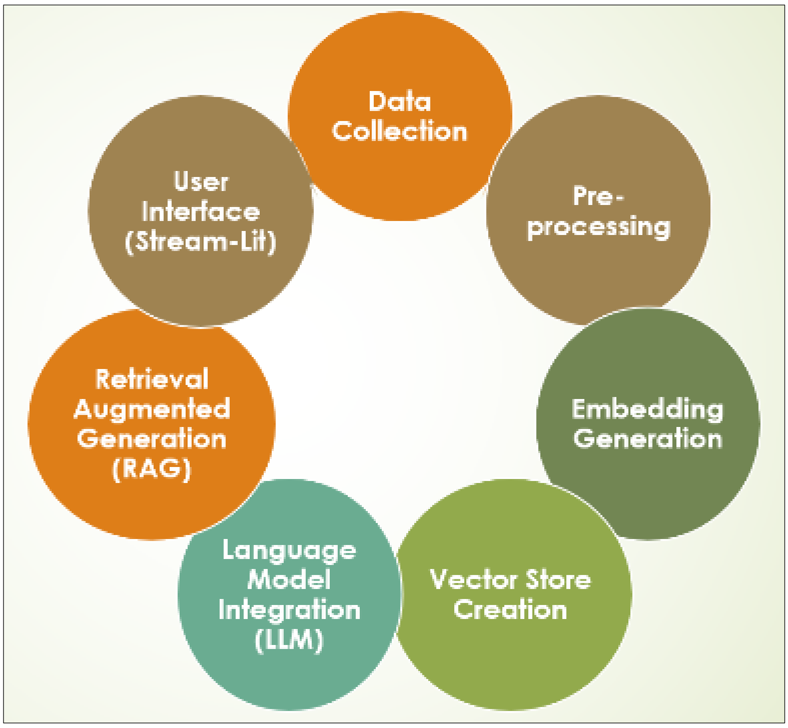

Data Collection
Pre-processing
Embedding Generation
Vector Store Creation
Language Model Integration (LLM)
Retrieval Augmented Generation (RAG)
User Interface (Stream-Lit)

Figure 1 Methodology

5.1. System Architecture

The architecture of the Multi-modal RAG Chatbot is designed to process and understand various types of data—text, images, and video transcripts—through a unified pipeline. The system employs a modular design with independent processors for each data type, all connected to a centralized retrieval-augmented generation engine powered by Google's Gemini model. This ensures scalable, extensible, and intelligent information extraction and interaction.

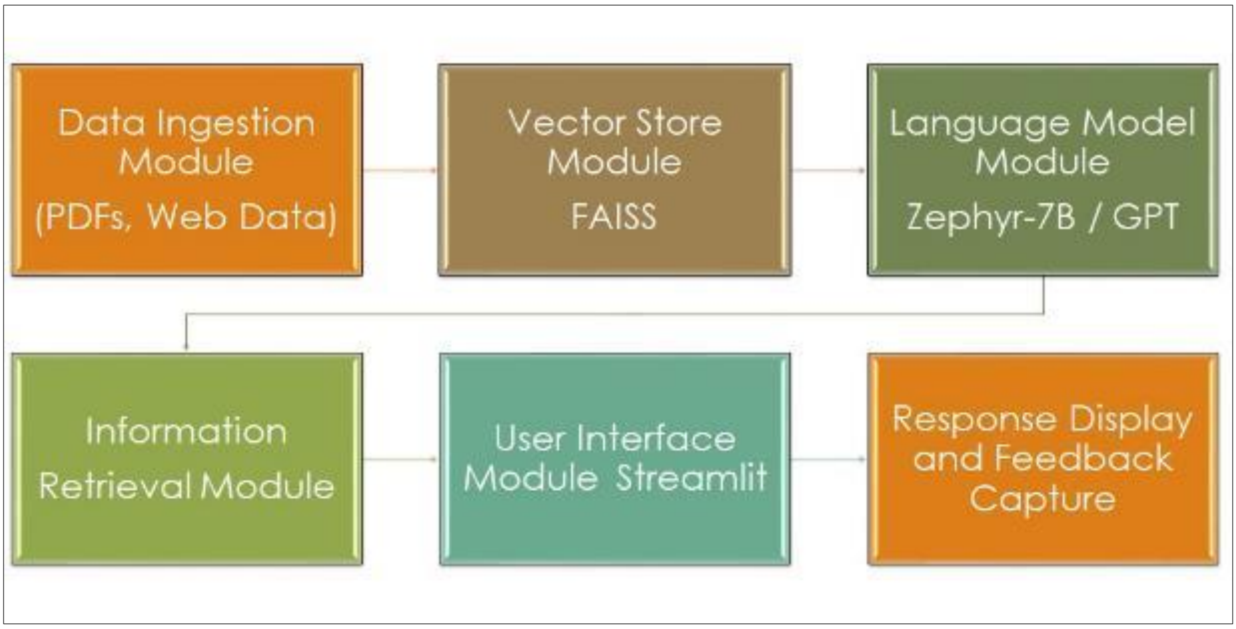


Figure 2 System Architecture

5.1.1. Data Ingestion Module

This module serves as the entry point for all external data sources. It is responsible for gathering structured and unstructured data, including PDFs, web pages, and other digital content. The primary goal is to ensure a rich and diverse knowledge base for downstream processing.

5.1.2. Vector Store Module

Once the data is ingested and processed, it is converted into numerical embeddings and stored in a vector database using FAISS (Facebook AI Similarity Search). This module allows for fast and efficient similarity searches, enabling the system to retrieve contextually relevant data segments during user interaction.

5.1.3. Language Model Module

The language understanding and generation tasks are handled by a powerful transformer-based language model such as Zephyr-7B or GPT. This module interprets user queries, generates coherent responses, and integrates retrieved knowledge to provide context-aware answers.

5.1.4. Information Retrieval Module

Acting as the bridge between user queries and stored data, this component queries the FAISS vector store to retrieve the most relevant chunks of information. These are then passed to the language model to enhance the relevance and accuracy of responses.

5.1.5. User Interface Module

The front-end of the system is developed using Stream-it, providing a lightweight and interactive interface. It allows users to input their queries, view generated answers, and interact with the system seamlessly.

5.1.6. Response Display and Feedback Capture

Finally, the system presents the generated output and provides an option for users to give feedback. This feedback is crucial for improving model performance and refining future iterations of the system.

The proposed methodology efficiently integrates multi-modal data processing through a unified RAG framework, enabling accurate information extraction from documents, images, and videos. With FAISS-based semantic retrieval and Gemini-powered generation, the system delivers context-aware, intelligent responses. It supports multilingual outputs and session management, enhancing user experience. This structured approach provides a scalable solution for educational tools, content summarization, and real-world AI applications that require reasoning across diverse data formats.

6. Results and Discussion

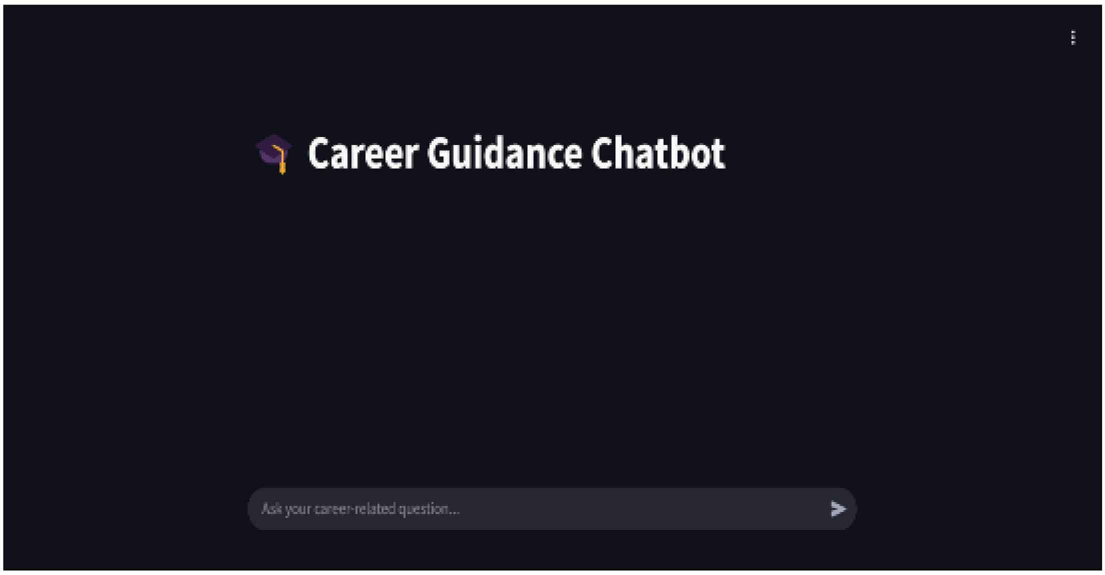
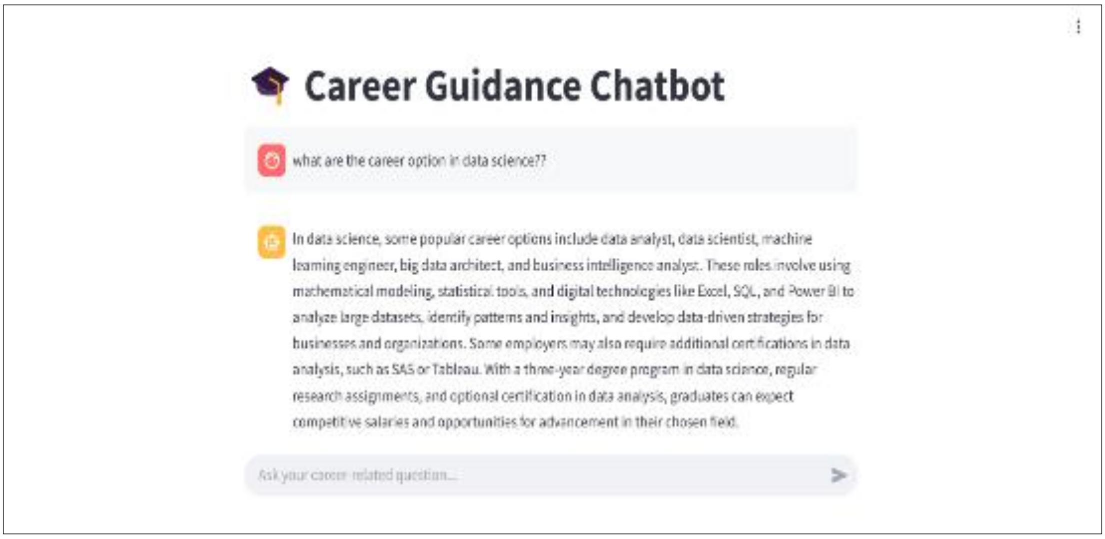


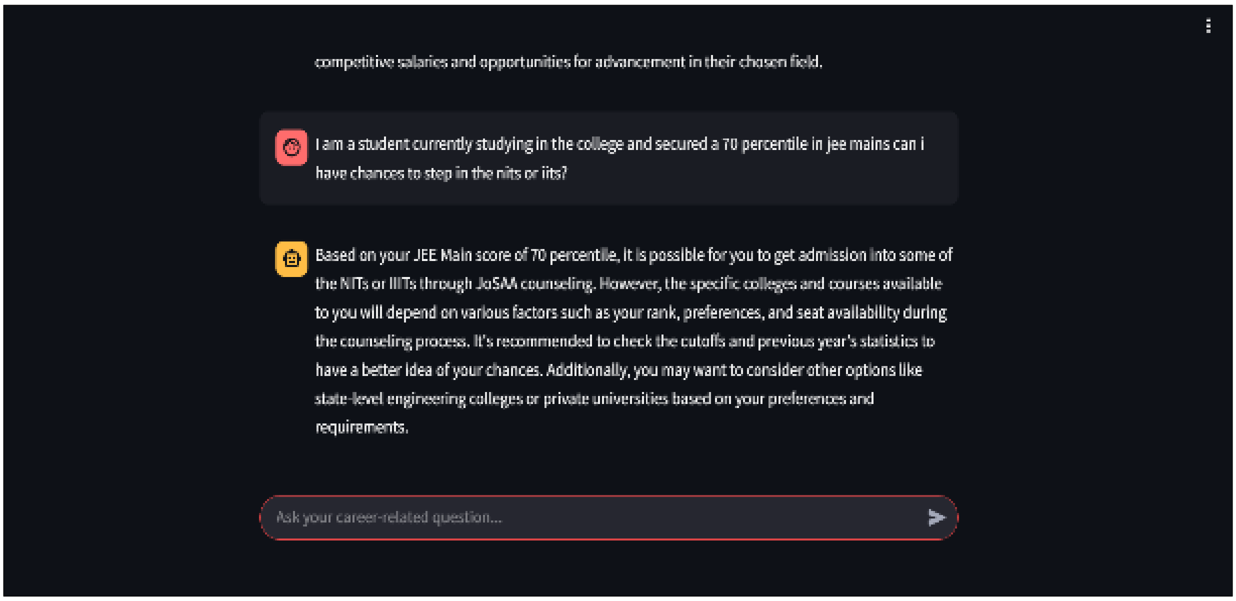
Figure 3 User Interface



what are the career option in data science??

In data science, some popular career options include data analyst, data scientist, machine leaming engineer, big data architect, and business intelligence analyst. These roles involve using mathematical modeling, statistical tools, and digital technologies like Excel, SQL, and Power BI to analyze large datasets, identify patterns and insights, and develop data-driven strategies for businesses and organizations. Some employers may also require additional certifications in data analysis, such as SAS or Tableau. With a three-year degree program in data science, regular research assignments, and optional certification in data analysis, graduates can expect competitive salaries and opportunities for advancement in their chosen field.

Figure 4 Chatbot Response Generation-1

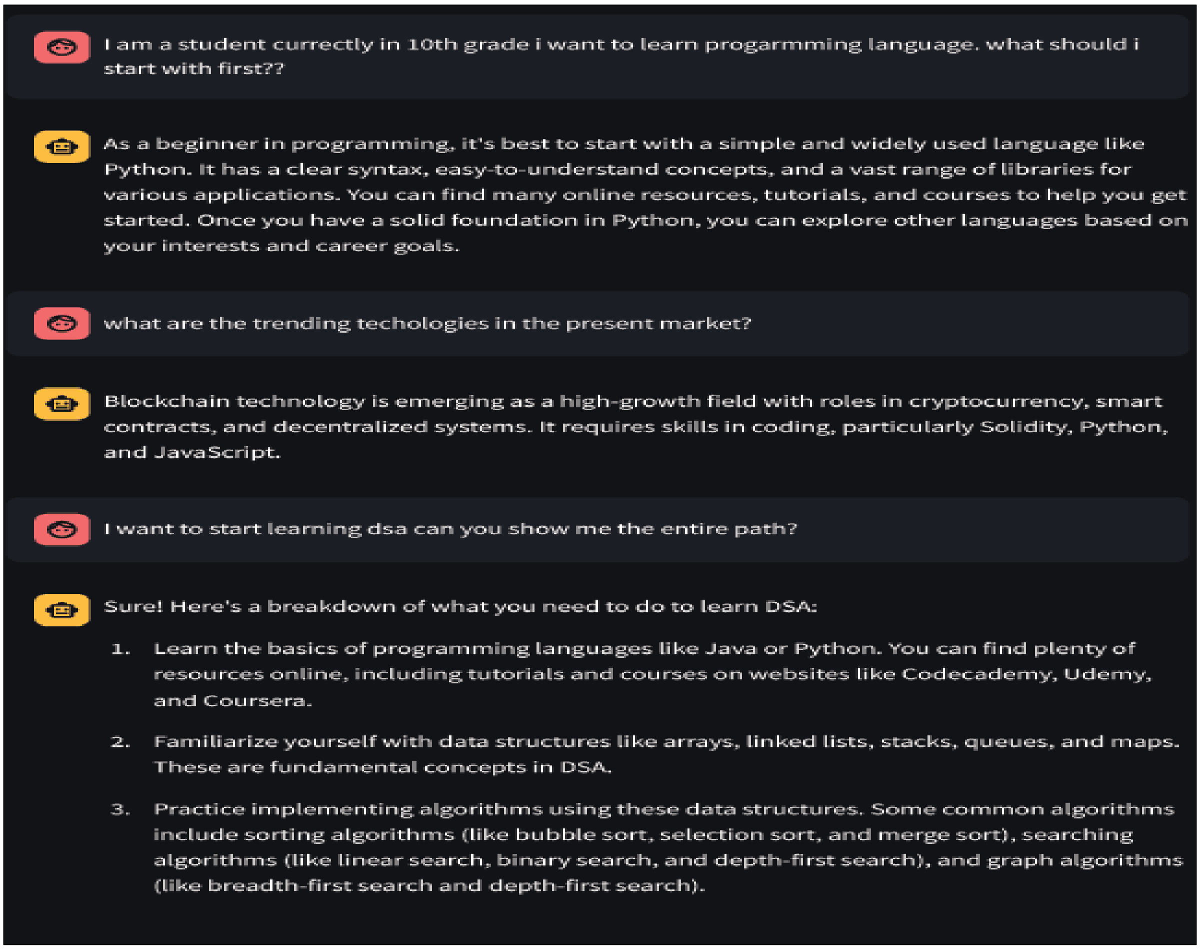


competitive salaries and opportunities for advancement in their chosen field.

I am a student currently studying in the college and secured a 70 percentile in jee mains can i have chances to step in the nits or iits?

Based on your JEE Main score of 70 percentile, it is possible for you to get admission into some of the NITs or IIITs through JoSAA counseling. However, the specific colleges and courses available to you will depend on various factors such as your rank, preferences, and seat availability during the counseling process. It's recommended to check the cutoffs and previous year's statistics to have a better idea of your chances. Additionally, you may want to consider other options like state-level engineering colleges or private universities based on your preferences and requirements.

Figure 5 Chatbot Response Generation-2



I am a student currectly in 10th grade i want to learn progarmming language. what should i start with first??

As a beginner in programming, it's best to start with a simple and widely used language like Python. It has a clear syntax, easy-to-understand concepts, and a vast range of libraries for various applications. You can find many online resources, tutorials, and courses to help you get started. Once you have a solid foundation in Python, you can explore other languages based on your interests and career goals.

what are the trending techologies in the present market?

Blockchain technology is emerging as a high-growth field with roles in cryptocurrency, smart contracts, and decentralized systems. It requires skills in coding, particularly Solidity, Python, and JavaScript.

I want to start learning dsa can you show me the entire path?

Sure! Here's a breakdown of what you need to do to learn DSA:

1. Learn the basics of programming languages like Java or Python. You can find plenty of resources online, including tutorials and courses on websites like Codecademy, Udemy, and Coursera.
2. Familiarize yourself with data structures like arrays, linked lists, stacks, queues, and maps. These are fundamental concepts in DSA.
3. Practice implementing algorithms using these data structures. Some common algorithms include sorting algorithms (like bubble sort, selection sort, and merge sort), searching algorithms (like linear search, binary search, and depth-first search), and graph algorithms (like breadth-first search and depth-first search).

Figure 6 User Interaction

7. Conclusion

Career & Future Advice Chatbot using RAG is an AI-based solution that gives personalized career guidance. Unlike traditional methods, it is available 24/7 and uses real-time data to help users with career paths, skills, resume building, and interview tips.

Built using Stream-Lit, Lang-Chain, FAISS, Sentence Transformers, and the Zephyr-7B model, the chatbot retrieves accurate information and responds like a real mentor. It's simple to use, supports students after 10th and 12th grades, and suggests streams, exams, and job options.

Overall, this chatbot bridges the gap between career aspirants and market trends, making career planning easier and more efficient. It not only addresses an immediate challenge but also lays a strong foundation for the future of AI in education and career mentoring.

Compliance with ethical standards

Disclosure of conflict of interest

There is no conflict of interest.

References

[1] Lewis et al. (2020) - Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks (Core RAG model concept).

[2] What is Retrieval-Augmented Generation (RAG)? https://cloud.google.com/use-cases/retrieval-augmented-generation

[3] Lang-Chain - https://python.langchain.com/ (Framework for RAG-based chatbot development).

[4] Radford et al. (2019) - Language Models are Few-Shot Learners (GPT-based models for career chatbots).

[5] Reference for the explanation about the rag and use of large language model's https://medium.com/@s.subodh7976/career-counselling-chatbot-using-rag

[6] Hugging Face - https://huggingface.co/ (Pre-trained NLP models).

Author's short biography

Mr. Shaik. Nagur Vali I'm Mr. Shaik. Nagur Vali, an Assistant Professor in Computer Science and Engineering (Data Science) with 5.2 years of Industry Experience in domains like DevOps, AWS Cloud, and Data Science and 1.2 years Teaching Experience. Holding a B.Tech and M.Tech in Computer Science and Engineering. I inspire students and contribute to advancements in technology through my work.	
P. Lokesh Karthik Varma I am P. Lokesh Karthik Varma, a Final-Year B.Tech Student at ACE Engineering College, specializing in CSE (Data Science). I am passionate about coding and problem-solving in Java and Python. I strive to improve myself continuously and innovate new things in the tech world. My goal is to explore industry work cultures and contribute to impactful technological advancements.	
S. Spandana I am S. Spandana, A Final-Year B.Tech Student at ACE Engineering College, Specializing in CSE(Data Science). I am Passionate about coding and problem-solving in java and python, I Strive to improve myself and enjoy exploring new things. My goal is to explore various industry works and contribute to new technologies.	

B. Praisey I am B. Praisey a final-year student specialising in Data Science at ACE Engineering College. I have a strong interest in ALML and am passionate about Data Analytics. I aspire to bridge technology with real-world challenges, focusing on data-driven innovations to solve real world problems and make a meaningful impact.	
M. Sahas Manikanta I am M. Sahas Manikanta, A Final-Year B.Tech Student at ACE Engineering College, Specializing in CSE(Data Science). I have a strong interest in Python and enjoy working on UI/UX development. I love creating interesting websites to enhance my tech skills and explore new design possibilities	