ISSN: 2583-6749 (Online)

DTC Journal of Computational Intelligence

DESIGNING AND DEVELOPING AI-POWERED JOB PORTAL CHATBOT: LEVERAGING MACHINE LEARNING FOR PERSONALIZED CAREER ASSISTANCE

Rachana Sharma[1], *Tushti Joshi[2], Jai Kumar Mishra[3], Vidhushi Sain[4], Sahil Aggarwal[5]*
[1]Assistant Professor, Delhi Technical Campus, Greater Noida, India
[2]Student, Delhi Technical Campus, Greater Noida, India
[3]Student, Delhi Technical Campus, Greater Noida, India
[4]Student, Delhi Technical Campus, Greater Noida, India
[5]Student, Delhi Technical Campus, Greater Noida, India

Abstract

The rapid digitalization of industries has led to an increasing reliance on AI-driven tools for personalized solutions, such as in job assistance and career guidance. This research introduces an AI-powered job portal chatbot that utilizes machine learning and web scraping to offer real-time, tailored career assistance. By leveraging advanced natural language processing (NLP) models and a retrieval-augmented generation (RAG) pipeline, the system provides context-aware interactions for job seekers. Web scraping is employed to gather diverse and up-to-date job market data, which is analyzed using pre-trained embeddings and machine learning algorithms to match user profiles with relevant job opportunities. Inspired by conversational analysis systems, the chatbot integrates tools like LangChain for managing complex queries, Hugging Face embeddings for semantic understanding, and FAISS-based vector storage for efficient information retrieval. The system offers an interactive user experience through a Streamlit-based interface, enabling users to upload documents, process job-related queries, and receive precise recommendations. With a scalable and modular architecture, the chatbot maintains multi-turn conversations, ensuring contextual relevance throughout interactions. This work emphasizes the transformative potential of AI in enhancing job search efficiency and personalization, paving the way for further advancements in AI-driven career solutions.

Keywords: *AI-Powered Chatbot, Job Portal, Natural Language Processing (NLP), Machine Learning, Web Scraping, Llama 3.1 Model, Large Language Models (Llms), Retrieval-Augmented Generation (Rag), Faiss Vector Storage*

1. INTRODUCTION

The rapid advancements in artificial intelligence (AI) and machine learning (ML) have opened new horizons for innovative applications in various domains, including career assistance. In today's competitive job market, individuals often encounter significant challenges in identifying suitable opportunities, enhancing their resumes, and preparing for interviews. The ever-growing volume of job listings, coupled with dynamically evolving industry standards, demands intelligent solutions capable of providing personalized and real-time guidance. This research proposes the development of an AI-powered job portal chatbot, leveraging Natural Language Processing (NLP) and Retrieval-Augmented Generation (RAG), to address these challenges and streamline the job search process. The proposed chatbot is designed to serve as an intelligent career assistant, offering tailored job recommendations, actionable resume feedback, and mock interview questions based on user profiles. Users can upload their resumes, which are then parsed using NLP techniques to extract key skills, qualifications, and experiences. By integrating real-time web scraping techniques with RAG, the system retrieves current job listings from platforms like LinkedIn and Indeed. This ensures that users receive up-to-date and personalized job suggestions that align with their career goals and qualifications. Additionally, the chatbot provides resume improvement suggestions by comparing user profiles against industry standards and generates role-specific mock interview questions to enhance interview preparedness.

Web scraping plays a pivotal role in the functionality of the chatbot, enabling the dynamic extraction of job market data from multiple online platforms. By automating data collection, the chatbot can provide users with the latest job opportunities, industry trends, and skill requirements, significantly reducing the manual effort involved in job hunting. This integration of web scraping with advanced AI techniques ensures a seamless and efficient user experience, making the job search process more targeted and effective.

By harnessing state-of-the-art machine learning technologies, including NLP and RAG, this project aims to create a comprehensive tool that transforms the way individuals navigate the complexities of the modern job market. The following sections detail the methodology, architecture, and potential applications of this system, highlighting its significance in empowering job seekers with personalized career assistance assessing skill development outcomes. The study also anticipates future trends and innovations in digital learning, including the impact of emerging technologies like artificial intelligence and virtual reality on employee training.

2. LITERATURE REVIEW

Conversational AI and chatbots have come a long way, thanks to improvements in natural language processing (NLP), machine learning (ML), and large language models (LLMs). Early examples, like those discussed by Meshram et al. (2021) [1], showcase basic chatbot designs, mainly relying on rule-based systems that are enhanced by NLP techniques. These chatbots play a crucial role in automating straightforward, repetitive tasks in areas such as

customer service and support, using AI and NLP to understand and respond to user inquiries. This research highlighted the significance of chatbots in task automation but pointed out their limitations in managing complex queries due to a lack of context awareness.

Building on these initial approaches, Lalwani et al. (2018) [2] investigated a chatbot system that utilizes AI and NLP for more advanced interactions, merging machine learning with NLP to allow for more flexible response generation. This study helped clarify how NLP-enhanced chatbots can transition from simple question-answering to more engaging conversational exchanges. Although the proposed chatbot improved interaction capabilities, it still faced challenges, such as relying on predefined rules and data for crafting responses, which often hindered its ability to grasp nuanced, context-rich conversations.

More recent studies have focused on incorporating more robust ML techniques in specialized applications, as demonstrated by Bal, Jash, and Mandal (2024) [3]. Their research introduced MIBOT, a healthcare chatbot designed for disease prediction, which employs machine learning algorithms to analyze patient inputs and forecast potential health issues. The strength of this chatbot lies in its capacity to interpret unstructured data and make predictions, showcasing the advantages of ML-based chatbots in specific fields. However, MIBOT's effectiveness is closely linked to its training data, which may restrict its adaptability in various contexts or specialized medical areas.

The rise of Retrieval-Augmented Generation (RAG) architectures has significantly advanced conversational AI, allowing for the creation of domain-specific chatbots that offer improved contextual relevance. Kulkarni et al. (2024) [4] investigated how reinforcement learning can be applied to enhance RAG models for these chatbots, showing notable gains in both retrieval accuracy and response quality. Their research lays out a framework for training chatbots that not only pull in relevant information but also craft responses based on previous conversation history. The RAG method tackles the issue of long-term contextual memory in conversational AI, representing a major leap toward making chatbots more responsive and aware of context.

Additionally, Jeong (2023) [5] examined a generative AI service that incorporates large language models (LLMs) within a RAG-based framework, utilizing the LangChain system to improve conversational flow and relevance. This research emphasizes the benefits of merging generative LLMs with retrieval-based systems to develop conversational platforms that can comprehend and adjust to user needs during extended interactions. The findings highlight the crucial role of LangChain in managing dialogue and preserving context, addressing the shortcomings of traditional chatbot systems that often falter in complex, multi-turn conversations.

These studies highlight the transition from basic rule-based chatbots to sophisticated RAG-based systems that leverage LLMs. They emphasize the increasing importance of contextual understanding, memory retention, and adaptability to specific domains in conversational AI. However, the current literature also points out persistent challenges, including the need for scalability and accuracy across various applications. This research seeks to advance these developments by creating a Conversational PDF Analysis System that employs an LLM-driven RAG framework, allowing users to engage meaningfully with document content. This method is in line with recent advancements in conversational AI, aiming to enhance accessibility and information retrieval in document-intensive fields like legal, academic, and corporate settings.

3. METHODOLOGY

The AI-powered job portal chatbot leverages advanced machine learning techniques, natural language processing (NLP), and web scraping to deliver personalized career assistance. The system's development involves several integrated components, each contributing to the seamless functionality and user experience of the chatbot.

A. Data Collection through Web Scraping

To ensure up-to-date job listings and market insights, web scraping techniques are utilized to gather data from various online job portals and professional networks. Libraries such as BeautifulSoup and Selenium are employed to extract job descriptions, required skills, and company details. The scraped data is preprocessed to remove duplicates, inconsistencies, and irrelevant information, forming a clean dataset for analysis.

B. Text Analysis and Preprocessing

The chatbot employs NLP methods for text tokenization, stemming, and semantic analysis. Pre-trained embeddings, such as Hugging Face's transformer models, are used to

convert textual data into vector representations. This enables efficient semantic search and contextual understanding of both user queries and job data.

C. Vector Storage and Retrieval

A FAISS (Facebook AI Similarity Search)-based vector storage system is implemented to enable fast and accurate retrieval of job-related information. Text chunks from scraped data are divided into chunks, and each chunk is transformed into a vector representation using the pre-trained embeddings from Hugging Face's sentence-transformers/all-MiniLM-L6-v2 model. This model is chosen for its effective balance between computational efficiency and semantic accuracy, allowing for relevant results while optimizing processing time.

D. Machine Learning for Personalization

User inputs, such as resumes, skills, and preferences, are processed using machine learning models. Algorithms analyze user data to identify trends and match profiles with suitable job listings. The system continuously adapts to user feedback, refining recommendations over time for improved personalization.

E. Conversational Interface with RAG Model

The system is built around a Retrieval-Augmented Generation (RAG) model, which improves upon traditional language models by incorporating document retrieval. This RAG configuration utilizes the Groq API to tap into a 70-billion parameter Llama 3.1 model. It is fine-tuned for conversational scenarios with a low temperature setting (0.2) to produce coherent and factually accurate responses. The RAG process consists of two main stages:

- Retrieval: When a user asks a question, the system initially retrieves relevant segments from the FAISS vector store through a nearest-neighbor search.
- Generation: The retrieved text, along with the user's question, is then fed into the Llama model, which crafts a response based on the pertinent document content and the context of the conversation.

F. User Interaction and Feedback Loop

A user-friendly interface built with Streamlit allows job seekers to interact with the chatbot. Features include uploading resumes, browsing job suggestions, and receiving real-time responses. A feedback mechanism is embedded to capture user satisfaction and refine the model's recommendations dynamically.

G. Multi-Turn Context Management

The system employs LangChain's ConversationBufferMemory to manage multi-turn dialogues. This ensures the chatbot maintains context across interactions, enabling it to respond intelligently to follow-up questions and refine its assistance based on the conversation history.

H. Deployment and Scalability

The system is designed for scalability, ensuring it can handle increasing user queries and large datasets. It is deployed using cloud infrastructure, with modular architecture for future enhancements such as multilingual support and multimodal data integration (e.g., images and tables in job descriptions).This methodology demonstrates a structured approach to developing a sophisticated chatbot that blends web scraping, machine learning, and conversational AI to transform the job search experience.

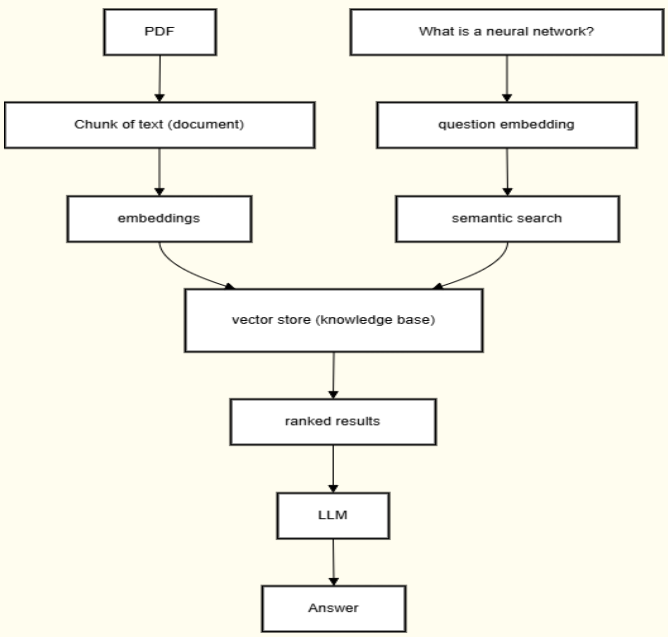

Figure 1(Methodology)

4. IMPLEMENTATION AND RESULTS

This project uses Python 3.9 for development because of its strong support for deep learning frameworks and its popularity in the NLP community. The versatility and readability of Python, along with its large community support, make it a great choice for creating and deploying advanced machine learning models.

A. Core Libraries and Tools

- LangChain: LangChain serves as the main deep learning framework, tailored specifically for natural language processing (NLP) applications. It provides high-level abstractions and utilities that streamline text processing, tokenization, and sequence modeling. LangChain is employed here to manage and facilitate the complex operations involved in creating a Conversational AI system capable of interacting with and extracting relevant information from large text corpuses.
- Hugging Face Transformers: The Hugging Face Transformers library is used to load and integrate transformer-based models, including pre-trained and custom models like Llama3-1-70B. Transformers are the core architecture for processing text data in this application, allowing for robust semantic analysis and response generation. The library provides pre-trained models, enabling rapid experimentation and fine-tuning without requiring large datasets.
- FAISS (Facebook AI Similarity Search): FAISS is utilized to handle vector storage and retrieval efficiently. By storing embeddings generated from text data, FAISS enables rapid similarity searches, which is essential for retrieving contextually relevant information from PDF data and enabling Retriever-Augmented Generation (RAG) within the chatbot.

B. System Workflow

The chatbot application is designed to answer user queries based on the content extracted from uploaded PDFs. The major components and functionalities include:

1) Document Processing and Parsing: The system can process PDF files, extracting readable content while filtering out irrelevant elements like headers, footers, or any other extraneous text that might compromise the semantic quality of the extracted data.

2) Text Segmentation: After parsing, the text is divided into smaller chunks using LangChain's text-splitting features. These chunks allow for efficient processing by the model and are tailored for creating embeddings that capture significant parts of the document's content.

3) Embedding Creation: Each text chunk is transformed into an embedding, which is a dense vector representation that captures its semantic meaning. The Hugging Face sentence-transformers model is utilized to generate these embeddings, guaranteeing high accuracy and relevance for retrieval tasks.

4) Vector Storage and Retrieval: The system stores generated embeddings in a FAISS vector database, which allows for the rapid retrieval of semantically similar content. When a user asks a question, the system transforms it into an embedding and conducts a similarity search within the FAISS database to find the most relevant pieces of information.

5) Retriever-Augmented Generation (RAG): To improve the chatbot's capability to produce precise and contextually relevant responses, RAG is utilized. The retrieved text segments provide context for the language model, which leverages them to formulate responses that consider both the user's question and pertinent document information.

6) User Interface with Streamlit: A web-based interface has been created using Streamlit, allowing users to upload PDFs, pose questions, and get real-time answers. The straightforward and interactive nature of Streamlit provides a smooth user experience, eliminating the need for advanced web development skills.

This implementation utilizes the powerful features of LangChain, Hugging Face Transformers, and FAISS to create a reliable and precise conversational system. It can understand, process, and respond to inquiries based on documents that are uploaded in real-time. The outcome is a scalable solution that merges Retrieval-Augmented Generation (RAG) with Large Language Models (LLMs) to provide interactions that are rich in context, making it ideal for research and information retrieval tasks.

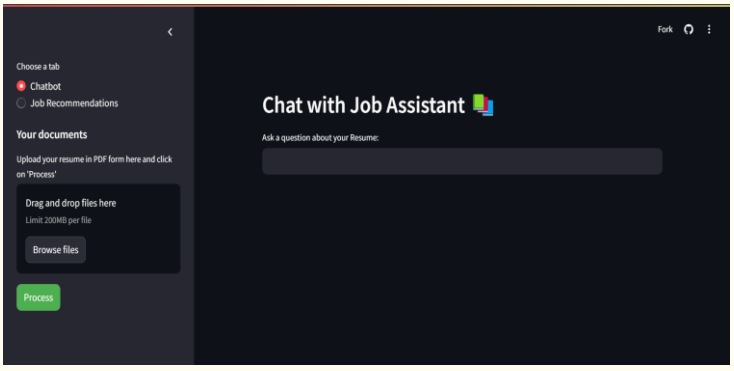

Figure 2 (User Interface)

5. FUTURE SCOPE

The current job recommendation system offers basic functionality by matching job roles with users based on a limited set of skills. While effective to an extent, this approach can overlook the nuanced capabilities and aspirations of users, leading to suboptimal recommendations. Enhancing this feature to incorporate a more detailed and skill-specific matching mechanism would significantly improve its precision and relevance. By leveraging a broader dataset and more granular profiling, the system can provide recommendations that are better aligned with a user's unique qualifications and career goals. Additionally, integrating Retrieval-Augmented Generation (RAG) would enable the system to access and analyze real-time data, offering insights into evolving job market trends. This would allow the system to suggest roles that not only match the user's current skills but also align with future industry demands.

The incorporation of the latest LLaMA model offers transformative potential by enabling the analysis of multimedia content, including videos, images, and audio. This advancement would extend the chatbot's capabilities beyond text-based analysis, paving the way for innovative features. For instance, the system could analyze video introductions embedded in resumes, providing personalized feedback to enhance presentation skills. Additionally, it could interpret visual content from job descriptions or listings, such as infographics or multimedia job advertisements, to provide a more holistic understanding of the opportunities available.

By enhancing the precision of job matching and incorporating multimedia analysis, this system can deliver a more comprehensive and adaptable tool for job seekers and recruiters alike. These advancements would not only elevate the quality of recommendations but also expand the scope of applications, enabling the platform to provide deeper insights into professional growth and market alignment.

6. CONCLUSION

This research presents the design and development of an AI-powered job portal chatbot that combines machine learning, web scraping, and conversational AI to provide personalized career assistance. By employing web scraping techniques, the system gathers and preprocesses diverse and up-to-date job data, which is then analyzed using advanced NLP models for contextual understanding and semantic matching. The integration of a retrieval-augmented generation (RAG) framework ensures precise, context-aware responses, while a user-friendly interface enables seamless interaction.

The chatbot not only streamlines the job search process by delivering tailored recommendations but also enhances user engagement through multi-turn conversational capabilities. By utilizing a scalable architecture with FAISS vector storage and fine-tuned language models, the system demonstrates efficiency, accuracy, and adaptability.

This work underscores the transformative potential of AI in modernizing career assistance and addressing the dynamic needs of job seekers. Future enhancements, such as multilingual support, multimodal data processing, and reinforcement learning for improved personalization, will further expand its applicability and effectiveness. This project sets a strong foundation for leveraging AI to redefine how individuals connect with career opportunities.

7. ACKNOWLEDGMENT

We would like to express our sincere gratitude to Ms. Rachana Sharma, our project supervisor, for her unwavering support and guidance. Her insightful feedback and encouragement helped us immensely. A special thanks to Dr. K.K. Singh, Director of Delhi Technical Campus, for providing us with the necessary facilities and resources to carry out this research. His encouragement and support have been invaluable.

8. REFERENCES

[1] Meshram, Siddhant & VR, Megha & More, Tanmay & Kharche, Shubhangi. (2021). Conversational AI: Chatbots. 1-6. 10.1109/CONIT51480.2021.9498508.

[2] Tarun Lalwani , Vasundhara Rathod. (2018). Implementation of a Chatbot System using AI and NLP. International Journal of Innovative Research in Computer Science and Technology (IJIRCST), 6(3),26-30, doi:10.21276/ijircst.2018.6.3.2.

[3] Bal, Sauvik & Jash, Kiran & Mandal, Lopa. (2024). An Implementation of Machine Learning-Based Healthcare Chabot for Disease Prediction (MIBOT). 10.1007/978-981-99-6866-4_32.

[4] M. Kulkarni, P. Tangarajan, K. Kim, and A. Trivedi, "Reinforcement Learning for Optimizing RAG for Domain Chatbots," arXiv preprint arXiv:2401.06800, 2024.

[5] Jeong, Cheonsu. (2023). Generative AI service implementation using LLM application architecture: based on RAG model and LangChain framework. Journal of Intelligence and Information Systems. 19. 129-164. 10.13088/jiis.2023.29.4.129.