



# International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





# AI-Driven Resume Parsing and Automated Interview Question Generation using NLP

Ms.A. Hemalatha<sup>1</sup>, K. Tejaswini<sup>2</sup>, B.Praveen Kumar<sup>3</sup>, J.V.Sandeep<sup>4</sup>, M.Roshini<sup>5</sup>, M.Sai Krishna<sup>6</sup>

Assistant Professor, Department of CSE (Data Science), NSRIT, Vishakhapatnam, India<sup>1</sup>

Student, Department of CSE (Data Science), NSRIT, Vishakhapatnam, India<sup>2,3,4,5,6</sup>

**ABSTRACT:** The rapid growth of digital recruitment platforms has significantly increased the number of job applications, making manual resume screening time-consuming and inefficient. In addition, many candidates lack proper guidance for interview preparation. To address these challenges, this paper proposes an AI-Driven Resume Parsing and Automated Interview Question Generation System using Natural Language Processing (NLP). The system extracts key information such as skills, education, certifications, and experience from resumes and analyzes candidate profiles using machine learning and language models. Based on the extracted data, it generates personalized technical and HR interview questions to improve candidate readiness. The system is implemented using Python and deployed as a web-based application. Experimental results demonstrate that the system effectively parses resume data and generates relevant interview questions. The proposed approach enhances recruitment efficiency and supports candidates in interview preparation, highlighting the potential of AI in modern recruitment systems.

**KEYWORDS:** Artificial Intelligence, Resume Parsing, Natural Language Processing, Machine Learning, Interview Question Generation, Recruitment Automated

## I. INTRODUCTION

Recruitment and talent acquisition are critical processes for organizations seeking skilled professionals in an increasingly competitive job market. With the rapid expansion of online job portals and digital recruitment platforms, companies often receive a large number of job applications for a single position. This increase in application volume has made manual resume screening inefficient and time-consuming for recruiters. Traditional recruitment processes rely heavily on human evaluation, which may lead to inconsistencies, bias, and delays in candidate selection. At the same time, job seekers face challenges in preparing effectively for interviews. Many candidates lack access to structured training resources that can help them understand the expectations of employers and the types of questions commonly asked during interviews. Without proper preparation, even highly qualified candidates may struggle to perform well during technical and behavioral interview rounds.

Recent advancements in **Artificial Intelligence (AI)** and **Natural Language Processing (NLP)** have enabled the development of intelligent systems capable of automating complex text analysis tasks. NLP techniques allow machines to process and understand large volumes of textual information, making them particularly useful for analyzing resumes and extracting meaningful insights from candidate profiles. AI-based recruitment systems can automatically identify candidate skills, match them with job requirements, and assist recruiters in selecting suitable candidates more efficiently. Several research studies have explored the use of machine learning and NLP techniques in resume parsing and automated candidate evaluation. These systems focus on converting unstructured resume documents into structured data that can be analyzed using classification and recommendation algorithms. Similarly, AI-based interview preparation tools have been developed to assist candidates by generating interview questions and providing feedback based on their skills.

Despite these advancements, many existing systems focus only on resume screening or interview preparation independently. The integration of resume parsing and automated interview question generation within a single intelligent system remains an area with significant potential for improvement. The proposed **AI-Driven Resume Parsing and Automated Interview Question Generation System using Natural Language Processing (NLP)** aims to address this gap by combining resume parsing, skill extraction, candidate evaluation, and automated interview question generation into a unified platform.



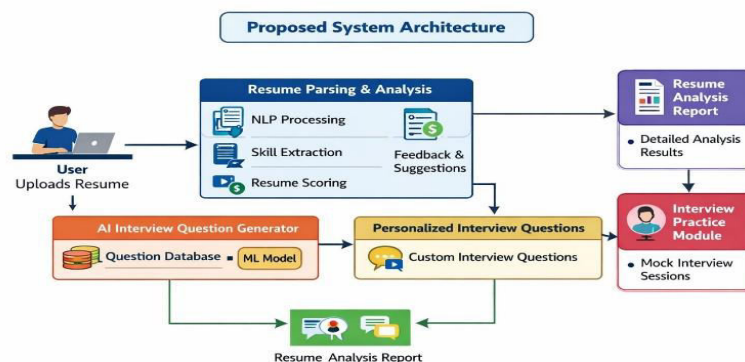
## International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The primary objectives of the proposed system include:

- Automatically extracting candidate information from resumes using NLP techniques
- Analyzing candidate skills and qualifications using machine learning algorithms
- Generating personalized interview questions based on extracted skills
- Providing an intelligent platform for interview preparation and career development

### II. SYSTEM METHODOLOGY



The proposed **AI-Driven Resume Parsing and Automated Interview Question Generation System using Natural Language Processing (NLP)** follows a structured workflow that processes candidate resumes and generates automated interview questions based on extracted skills. The system integrates Natural Language Processing (NLP) and Machine Learning techniques to extract meaningful information from resumes and provide intelligent interview preparation support. The methodology involves several stages including resume acquisition, text preprocessing, skill extraction, and automated interview question generation.

#### 2.1 Resume Data Acquisition

Resume data acquisition is the first stage of the proposed system. Candidates upload their resumes through the web-based interface in formats such as **PDF or DOCX**. These resumes contain unstructured textual information including candidate name, contact details, education, technical skills, project experience, certifications, and work history. Once uploaded, the resumes are stored in the system database and passed to the processing module. The system converts the document content into machine-readable text using document parsing techniques. This extracted textual data serves as the primary input for the Natural Language Processing pipeline that performs further analysis of the candidate profile. Proper resume data acquisition ensures that the system receives consistent input data, enabling accurate extraction of relevant candidate information.

#### 2.2 RESUME TEXT CONVERSION

Before performing resume analysis, several preprocessing operations are applied to improve the quality and consistency of textual data. Resumes often contain formatting inconsistencies, punctuation marks, special characters, and irrelevant words that can reduce the accuracy of machine learning models. The preprocessing stage begins with **text extraction**, where the content of the uploaded resume document is frequently occurring words that do not contribute meaningful information. Further preprocessing includes **text normalization and lemmatization**, which convert words into their base forms to improve semantic understanding. These preprocessing steps transform raw resume data into structured textual information that can be effectively analyzed by machine learning and NLP algorithms.

#### 2.3 Skill Extraction using Natural Language Processing

After preprocessing the resume text, the next stage involves extracting important information related to the candidate profile. This process is performed using **Natural Language Processing techniques**, which identify and extract relevant entities from the resume. The system analyzes the processed text to detect entities such as **technical skills, programming languages, educational qualifications, certifications, and work experience**. Named Entity Recognition (NER) techniques are used to identify these entities and convert the unstructured resume text into structured data.



## International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

### 2.4 Interview Question Generation Module

The system is implemented using Python with a Streamlit-based web interface. Resume text is extracted using PyPDF2 and python-docx libraries. NLP processing is performed using Hugging Face models, and data is managed using Pandas. SQLite is used as the backend database. The system allows users to upload resumes and receive analysis results along with interview questions.

### III. AI MODEL AND NLP PROCESSING

The proposed system utilizes **Large Language Model (LLM)-based Natural Language Processing techniques** to analyze resumes and generate personalized interview preparation content. Unlike traditional machine learning approaches that rely on manual feature engineering, LLM-based models can understand the semantic meaning of textual data and extract meaningful insights from unstructured documents. The system integrates the **Hugging Face API**, which provides access to advanced pre-trained language models capable of processing large amounts of textual information. When a candidate uploads a resume, the extracted text is passed to the LLM-based processing module. The model analyzes the resume content and identifies important information such as technical skills, education, project experience, and certifications.

### IV. SYSTEM IMPLEMENTATION

The proposed AI-Driven Resume Parsing and Automated Interview Question Generation System is implemented using Python. A web-based interface is developed using Streamlit, allowing users to upload resumes and view analysis results. Resume text is extracted from PDF and DOCX files using PyPDF2 and python-docx libraries. The extracted text is processed using Natural Language Processing with Large Language Models through the Hugging Face API to identify skills, education, and experience. Data processing is handled using Pandas, while Matplotlib and Streamlit components are used for visualization. The system ensures secure user authentication using SHA-256 hashing. SQLite is used as the backend database to store user data and analysis results. The complete system operates as an interactive application that provides resume analysis and generates personalized interview questions.

### V. DATASET DESCRIPTION

The dataset used in this research consists of resumes collected from publicly available sources and sample candidate profiles. These resumes include information such as skills, education, certifications, project experience, and work history, and are stored in PDF and DOCX formats. The system extracts text from these documents using parsing libraries and processes it using Natural Language Processing techniques to identify key entities such as programming languages, tools, and technical skills. The extracted data is organized into structured profiles for analysis. The dataset covers multiple domains including software development, data science, and web development, enabling the system to generate relevant interview questions. The processed data is analyzed using an LLM-based NLP pipeline to identify candidate competencies and provide personalized interview preparation support.

### VI. LITERATURE REVIEW

Artificial Intelligence and Natural Language Processing have improved modern recruitment systems by enabling automated resume analysis. Earlier systems relied on keyword-based methods, which lacked contextual understanding. Recent approaches use NLP techniques such as tokenization, named entity recognition, and text classification to extract meaningful information from resumes and convert it into structured data. Transformer-based Large Language Models further enhance this process by improving contextual understanding. The proposed system builds on these advancements by integrating resume parsing with automated interview question generation in a single platform. In recent years, transformer-based Large Language Models have shown promising results in understanding textual data. These models enable systems to analyze resume content more effectively and generate intelligent insights. The proposed system utilizes AI and NLP-based techniques to automatically analyze resumes and generate personalized interview questions, providing an integrated solution for resume evaluation and interview preparation.



## International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

### VII. EXPERIMENTAL RESULTS

The performance of the proposed **AI-Driven Resume Parsing and Automated Interview Question Generation System using Natural Language Processing (NLP)** was evaluated to determine its effectiveness in extracting resume information and generating relevant interview questions. The system processes resumes uploaded by users and extracts important information such as technical skills, education, certifications, and professional experience using Natural Language Processing techniques. The extracted data is then analyzed using a Large Language Model (LLM) based processing module to understand the contextual meaning of the resume content and generate automated interview questions. To evaluate the effectiveness of the system, standard evaluation metrics such as accuracy, precision, recall, and F1-score were considered. Accuracy measures the overall correctness of the system in identifying relevant skills from the resume and generating suitable interview questions. Precision indicates how many of the predicted relevant skills are actually correct, while recall measures the system's ability to identify all relevant skills present in the resume text. The F1-score provides a balanced measure of both precision and recall. Experimental observations show that the system successfully extracts key information from resumes and generates meaningful interview questions that match the candidate's skill set. The use of LLM-based NLP processing improves contextual understanding of resume content and produces more relevant interview questions compared to traditional keyword-based systems..

### VIII. APPLICATIONS

The proposed system can be applied in various areas of recruitment and career development. It can be used for recruitment automation to analyze resumes and identify suitable candidates, reducing manual effort. It also supports career guidance by helping students and job seekers improve their resumes and skills. Additionally, the system acts as an interview preparation tool by generating personalized technical and HR questions. It can also be integrated into online job portals to match candidate skills with job requirements and recommend suitable opportunities.

### IX. FUTURE WORK

Although the proposed system demonstrates promising results in resume analysis and interview preparation, several improvements can be implemented in future research. One possible enhancement is the integration of more advanced transformer-based language models to improve the semantic understanding of resume content and generate even more accurate interview questions. Future versions of the system can also include automated resume scoring mechanisms that evaluate candidate profiles based on industry requirements and provide detailed feedback for improvement. Another potential enhancement involves integrating job recommendation features that match candidate skills with suitable job opportunities available on recruitment platforms. The system can also be extended to include mock interview simulations using speech and video analysis technologies, which can evaluate candidate responses and provide feedback on communication skills and confidence levels. Additionally, expanding the resume dataset with profiles from different industries and professional backgrounds will improve the system's accuracy and generalization capability. Mobile-based deployment and cloud integration can further enhance the accessibility and scalability of the system.

### X. CONCLUSION

This research presented an AI-Driven Resume Parsing and Automated Interview Question Generation System using Natural Language Processing designed to support both job seekers and recruiters in the modern recruitment process. The proposed system utilizes Natural Language Processing and Large Language Model-based techniques to automatically extract information from resumes and generate relevant interview questions based on candidate skills. The system integrates resume parsing, skill extraction, candidate profile analysis, and automated interview question generation into a single intelligent platform. By analyzing candidate resumes and identifying key competencies, the system provides structured guidance that helps candidates prepare more effectively for job interviews. The implementation of the system using Python, Streamlit, Hugging Face API, and NLP libraries demonstrates the practical applicability of artificial intelligence technologies in recruitment and career development platforms. Experimental observations indicate that the system can successfully parse resume data and generate relevant interview questions tailored to candidate profiles. Overall, the proposed system highlights potential of artificial intelligence in transforming traditional recruitment processes into intelligent and automated career support systems that benefit both organizations and job seekers



## International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

### REFERENCES

- [1] Y. LeCun et al., "Deep Learning," Nature, 2015.
- [2] J. Devlin et al., "BERT: Language Understanding," NAACL, 2019.
- [3] A. Vaswani et al., "Attention Is All You Need," NIPS, 2017.
- [4] S. Bird et al., Natural Language Processing with Python, 2009.
- [5] D. Jurafsky & J. Martin, Speech and Language Processing, 2019.
- [6] T. Mikolov et al., "Word Representations," ICLR, 2013.
- [7] J. Pennington et al., "GloVe," EMNLP, 2014.
- [8] T. Brown et al., "Few-Shot Learners," 2020.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  [ijircce@gmail.com](mailto:ijircce@gmail.com)



[www.ijircce.com](http://www.ijircce.com)

Scan to save the contact details