Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 815-831

DESIGN AND IMPLEMENTATION OF AN INTELLIGENT RESUME PARSING AND CANDIDATE PROFILING SYSTEM USING LANGCHAIN AND LARGE LANGUAGE MODELS

[1]Vadarevu Laxmi Suprita, [2]Dr. Anjali Chandra, [3]Swasti Sharma

[1,3]UG- Computer Science & Engineering with Specialization in (AI), [2]Assistant Professor

[1,2,3]Shri Shankaracharya Institute of Professional Management and Technology, Raipur, Chhattisgarh, India

Abstract

Due to the rising number of job applications, recruiters have found manual resume screening slow, repetitive and inefficient. Traditional Applicant Tracking Systems (ATS) rely strictly on keyword matching methods which are unable to comprehend the context of the applicants' data. This can lead to high number of applicants being rejected and also inappropriate profiles being shortlisted. In this paper, to solve these problems, the design and implementation of an intelligent resume parsing and candidate profiling system based on Natural Language Processing (NLP), Large Language Models (LLMs) and LangChain framework is introduced. The system is designed to handle PDF resumes and retrieve the structured data including candidate's name, contact information, technical skills, educational background, certifications, and work experience. The use of SpaCy-based NLP techniques for information extraction and LLM-based parsing with the help of the Gemini API via LangChain for context-aware understanding and improved accuracy is a hybrid approach. Moreover, an ATS-focused scoring system is built-in that assesses and ranks resumes based on the candidate's skills that match the one in the job description and the recruiter's requirements experimental results show that the proposed system provides high accuracy, consistency and efficiency in resumes analysis and facilitates the automated resumes ranking and quick decision making process for hiring.

Keywords: Artificial Intelligence, Healthcare System, Natural Language Processing (NLP), Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), Vector Database, Pinecone, HuggingFace Embedding, LangChain Framework, Groq API, Flask Web Framework, AWS Cloud Deployment, Semantic Search, Knowledge Grounding, Medical Chatbot, Healthcare Informatics.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 815-831

1. INTRODUCTION

The rapid growth of digital platforms and online job portals has significantly increased the number of job applications received by organizations for a single position. This surge in applications has made the recruitment process more complex, time-consuming, and resource-intensive. Traditional hiring methods rely heavily on manual resume screening, where recruiters analyze candidate profiles individually. This approach is not only inefficient but also prone to human bias and inconsistency, especially when dealing with large volumes of resumes [1]. To address these challenges, Applicant Tracking Systems (ATS) were introduced to automate the resume screening process. ATS systems help in filtering and shortlisting candidates based on predefined criteria, primarily using keyword matching techniques. However, these systems have significant limitations, as they lack the ability to understand the contextual meaning and semantic relationships within resume content. As a result, qualified candidates may be overlooked if their resumes do not exactly match the required keywords, reducing the effectiveness of the recruitment process [2].With the advancement of Artificial Intelligence (AI), more sophisticated approaches have been developed to improve recruitment systems. Natural Language Processing (NLP) techniques enable automated extraction of structured information such as skills, education, and experience from unstructured resume data. Tools like SpaCy utilize methods such as tokenization and Named Entity Recognition (NER) to identify relevant entities within text. Despite these improvements, traditional NLP approaches often struggle with capturing deeper contextual meaning and variations in language [3]. The emergence of Large Language Models (LLMs), such as transformer-based architectures, has significantly enhanced the ability of systems to process and understand human language. Models like BERT, GPT, and Gemini provide context-aware analysis, enabling more accurate extraction and interpretation of resume data. These models can understand semantic relationships and variations in wording, making them highly effective for tasks such as resume parsing and candidate profiling [4]. Furthermore, frameworks such as LangChain have simplified the integration of LLMs into real-world applications. LangChain allows developers to build structured workflows by combining prompt templates, chains, memory, and output parsers. This enables the generation of structured outputs from unstructured inputs, improving the reliability and usability of LLM-based systems [5]. In this paper, we propose an intelligent resume parsing and candidate profiling system that integrates NLP and LLM-based approaches using the LangChain framework. The system processes

ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 815-831

resumes in PDF format, extracts structured information, and evaluates candidates using an ATS-based scoring mechanism. By combining rule-based and context-aware techniques, the proposed system aims to improve the accuracy, efficiency, and fairness of the recruitment process.

2. LITERATURE REVIEW

2.1. Resume Parsing Systems

Resume parsing systems are designed to automatically extract relevant information from resumes, such as candidate name, skills, education, and experience. Early systems relied on rule-based techniques, including predefined templates, regular expressions, and keyword matching. While these methods were effective for structured resumes, they struggled with variations in format and layout, leading to inaccurate or incomplete data extraction. .To improve performance, machine learning-based approaches were introduced, which learn patterns from training data and adapt to different resume structures. These systems provide better flexibility compared to rule-based methods but require large labeled datasets and may not generalize well to unseen formats.Recent research focuses on hybrid systems that combine rule-based and data-driven approaches. These systems aim to improve accuracy by leveraging both predefined rules and learned patterns. However, challenges such as handling unstructured data, identifying implicit skills, and maintaining consistency across diverse resumes remain unresolved. Therefore, there is a need for more advanced approaches that incorporate contextual understanding to enhance resume parsing performance.

2.2.Natural Language Processing in Recruitment

Natural Language Processing (NLP) plays a crucial role in modern recruitment systems by enabling the extraction and analysis of information from unstructured resume data. NLP techniques such as tokenization, lemmatization, part-of-speech tagging, and Named Entity Recognition (NER) are widely used to identify key entities like candidate names, skills, organizations, and educational qualifications. These methods help convert raw text into structured data, improving the efficiency of resume parsing.Tools like SpaCy and NLTK provide robust NLP pipelines that support large-scale resume processing. However, traditional NLP approaches often rely on rule-based or statistical methods, which limits their ability to

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 815-831

understand contextual meaning and semantic relationships. For example, similar skills expressed differently may not be recognized as equivalent.To overcome these limitations, NLP is increasingly integrated with advanced AI models to enhance contextual understanding, resulting in more accurate candidate evaluation and improved recruitment outcomes.

2.3.Large Language Models (LLMs)

Large Language Models (LLMs) represent a major advancement in Natural Language Processing by enabling systems to understand and generate human-like text. Models such as BERT, GPT, and Gemini are based on transformer architectures that utilize attention mechanisms to capture contextual relationships within text. This allows LLMs to process language more effectively than traditional rule-based or statistical methods. In recruitment systems, LLMs are used for resume parsing, information extraction, and candidate matching. Unlike keyword-based approaches, LLMs can understand semantic meaning and identify relevant skills even when expressed differently, improving the accuracy of candidate profiling.Furthermore, LLMs can handle diverse resume formats and writing styles, making them suitable for real-world applications. However, they also face challenges such as high computational cost and hallucination, where incorrect information may be generated. These limitations highlight the need for hybrid approaches to ensure reliable outputs.

2.4.AI-Based Candidate Evaluation Systems

AI-based candidate evaluation systems automate the assessment and ranking of applicants using techniques such as NLP and machine learning. These systems analyze resumes and compare them with job descriptions to identify relevant skills and qualifications. A key component is the Applicant Tracking System (ATS) scoring mechanism, which evaluates candidates based on skill matching and relevance. Advanced systems also incorporate semantic matching, improving accuracy beyond simple keyword comparison. This enables more consistent and objective decision-making. However, challenges such as bias in data and lack of transparency remain, highlighting the need for fair and explainable AI solutions in recruitment.

2.5.Research Gap

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 815-831

Despite significant advancements in automated recruitment systems, several limitations still exist. Most existing approaches rely either on rule-based NLP techniques or on Large Language Models (LLMs), but not an effective combination of both. Rule-based systems lack contextual understanding, while LLM-based systems may produce inconsistent or hallucinated outputs, affecting reliability. Additionally, many systems do not incorporate comprehensive evaluation metrics such as ATS scoring, precision, recall, or confusion matrix analysis. Issues related to bias and lack of transparency in AI-based decision-making also remain major concerns.

The key research gaps are as follows:

- *Lack of hybrid systems combining NLP and LLM approaches*
- *Limited contextual understanding in traditional systems*
- *Absence of proper evaluation metrics (ATS score, precision, recall)*
- *Risk of hallucination in LLM-based systems*
- *Bias and lack of explainability in AI decision-making*

Addressing these gaps is essential for developing accurate, fair, and efficient intelligent recruitment systems.

3. LITERATURE REVIEW

3.1. System Architecture

The proposed system is an intelligent resume parsing and candidate profiling solution designed to automate and enhance the recruitment process using advanced Artificial Intelligence techniques. It integrates Natural Language Processing (NLP), Large Language Models (LLMs), and the LangChain framework to process unstructured resume data and convert it into structured information. The system is capable of handling multiple resumes simultaneously, improving efficiency and reducing manual effort. The System Architecture showing in Figure1. The system begins by accepting resumes in PDF format through a user-friendly interface. These resumes are processed using a text extraction module, which converts the content into raw text. This step ensures that resumes with different formats and layouts can be analyzed effectively. After text extraction, the system applies a hybrid processing approach. The NLP

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 815-831

module, implemented using SpaCy, performs rule-based extraction tasks such as tokenization, Named Entity Recognition (NER), and skill identification using predefined skill sets. In parallel, the system uses a Large Language Model through the LangChain framework integrated with the Gemini API. This module performs context-aware parsing, enabling the system to understand the semantic meaning of resume content.The outputs from both modules are combined to generate accurate and comprehensive candidate profiles. The extracted data is structured into fields such as name, contact details, skills, education, and experience.The system then evaluates candidates using an Applicant Tracking System (ATS) scoring mechanism, which compares candidate skills with job description requirements. Based on the calculated scores, candidates are ranked accordingly. Finally, the results are presented through an interactive dashboard, enabling recruiters to make efficient and data-driven hiring decisions.

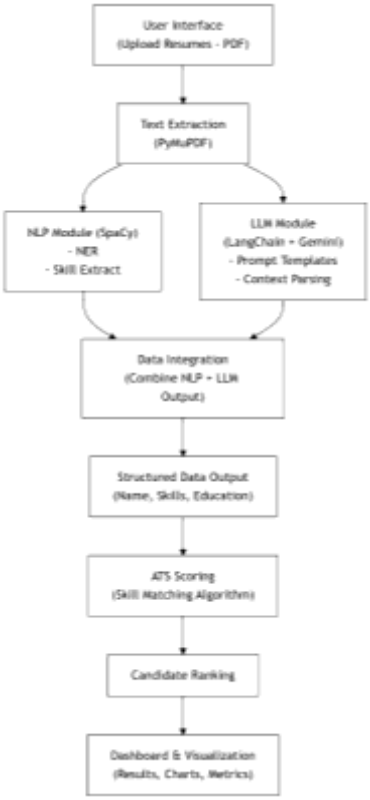

Figure 1: System Architecture

3.2.System Design

The system begins with the input layer, where users upload resumes in PDF format through a web-based interface.The system design is showing in Figure2. These files are passed to the text extraction module**,** which uses PyMuPDF to convert documents into raw text. The extracted

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 815-831

text is then processed by two main components. The NLP module (SpaCy) performs tokenization, Named Entity Recognition (NER), and skill extraction using predefined datasets. Simultaneously, the LLM module (LangChain with Gemini API) performs context-aware parsing using prompt templates and output parsers to extract detailed information. The outputs from both modules are combined in the integration layer, which removes redundancy and merges relevant data. This ensures improved accuracy and completeness. Next, the data structuring module organizes the extracted information into a predefined schema, including fields such as name, contact details, skills, education, and experience. The structured data is then passed to the ATS scoring module**,** which evaluates candidates by comparing their skills with job requirements. Finally, the output layer presents the results through a dashboard, displaying candidate rankings, scores, and evaluation metrics. This design ensures efficient data flow, modularity, and ease of system expansion.

Figure 2: System Design

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 815-831

3.3.Modules Description

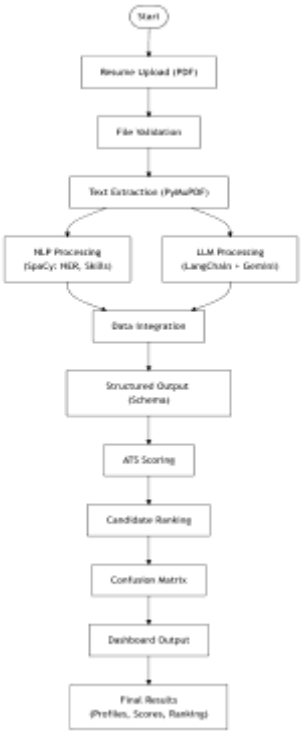

Figure 3: Module System Architecture Flowchart

The Module System Architecture Flowchart is showing in Figure3

- Input and Data Acquisition Module

This module is responsible for collecting and preparing resume data for processing. It allows users to upload multiple resumes in PDF format through an interactive interface built using Streamlit. The uploaded files are validated and passed to the text extraction component, which uses PyMuPDF (fitz) to convert PDF documents into raw textual data. This step ensures that resumes with different formats, layouts, and structures are standardized into a machine-readable form. The module efficiently handles multi-page documents and preserves all relevant information, including headings, bullet points, and paragraphs. Proper file handling mechanisms are implemented to avoid data loss and ensure smooth processing. By transforming unstructured PDF resumes into structured text input, this module lays the foundation for further analysis and ensures that the system can handle large volumes of data effectively.

- Processing and Information Extraction Module

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 815-831

This module performs the core functionality of extracting meaningful information from resume text using a hybrid approach. It combines Natural Language Processing (NLP) techniques with Large Language Models (LLMs) to achieve accurate and context-aware parsing. The NLP component, implemented using SpaCy, performs tokenization, Named Entity Recognition (NER), and skill extraction using predefined skill sets. In parallel, the LLM component, integrated through the LangChain framework and Gemini API, performs semantic analysis using prompt templates and output parsers. This enables the system to understand context and extract complex information such as experience and skill relationships. The outputs from both components are merged in a data integration layer, where redundant data is removed and complementary information is combined. The final output is then structured into a standardized schema, ensuring consistency and completeness of candidate profiles.

- Evaluation and Visualization Module

This module focuses on evaluating candidates and presenting the results in an understandable format. It uses an Applicant Tracking System (ATS) scoring mechanism to assess candidate suitability by comparing extracted skills with job description requirements. The score is calculated based on the percentage of matched skills, providing an objective measure of relevance. Based on these scores, candidates are ranked in descending order to help recruiters quickly identify the best matches. The module also includes performance evaluation using metrics such as confusion matrix, enabling analysis of system accuracy. The Confusion Matrix showing in Figure4. Finally, the results are displayed through an interactive dashboard, showing candidate details, scores, rankings, and visualizations. This module enhances decision-making by providing clear, data-driven insights and ensures an efficient and user-friendly recruitment process.

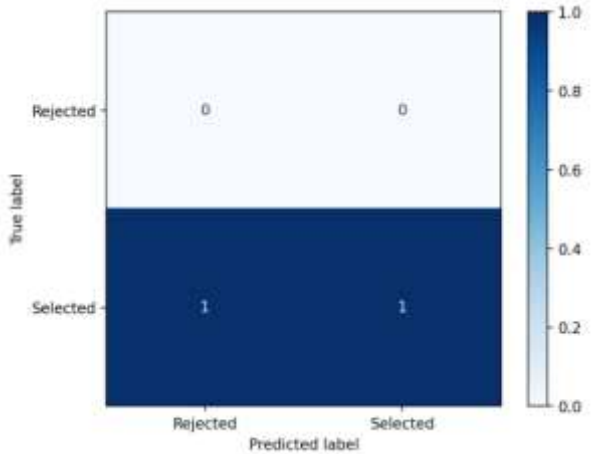

Figure 4: Confusion Matrix

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 815-831

3.4.Algorithms And Technologies Used

The system uses a document parsing algorithm implemented through PyMuPDF to extract textual data from PDF resumes. It processes each page sequentially and converts it into machine-readable text while preserving structure.

- NLP-Based Extraction Algorithm

Natural Language Processing techniques are applied using SpaCy. This includes tokenization, Named Entity Recognition (NER), and phrase matching for skill extraction. The PhraseMatcher algorithm is used to identify predefined skills within the resume text efficiently.

- LLM-Based Parsing Algorithm

A prompt-based extraction algorithm is implemented using LangChain and Gemini API. It uses structured prompts and output parsing to extract information such as name, skills, education, and experience in a predefined schema format.

- ATS Scoring Algorithm

The system calculates equation1 shown in candidate suitability using a skill-matching algorithm:

$$\text{ATS Score} = \frac{\text{Matched Skills}}{\text{Total Required Skills}} \times 100 \qquad (1)$$

This algorithm ensures objective evaluation of candidates.

- Ranking Algorithm

Candidates are ranked using a sorting algorithm based on ATS scores in descending order, enabling quick identification of top candidates.

4. Result

The proposed system was tested using multiple resumes in PDF format along with a predefined set of job description skills. The system successfully extracted key information such as candidate name, email, phone number, skills, education, and experience using a hybrid approach combining NLP and LLM techniques. The integration of SpaCy and Gemini-based parsing improved the accuracy of information extraction, especially for unstructured and varied resume formats.

The Applicant Tracking System (ATS) scoring mechanism effectively evaluated candidates by comparing their extracted skills with job requirements. Each candidate was assigned a score

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 815-831

based on the percentage of matched skills. The system then ranked candidates in descending order of their ATS scores, enabling quick identification of the most suitable applicants.

The results were displayed through an interactive dashboard, which included candidate details, scores, and rankings. Additionally, a confusion matrix was generated to evaluate system performance by comparing predicted selections with actual recruiter decisions.

4.1. Multimodal Application Input Interface

The system provides a user-friendly and interactive input interface developed using Streamlit. The The real input interface of uploaded and application preferences of dermatology image and analysis showing in Figure5. It supports the upload of multiple resumes in PDF format, enabling recruiters to process several candidate profiles simultaneously. The interface also allows users to input job description skills in a text field, which are used for ATS-based evaluation. This multimodal input capability combines document input (PDF resumes) and textual input (job requirements), making the system flexible and practical for real-world recruitment scenarios. Additionally, the interface ensures smooth file handling, validation, and preprocessing, allowing seamless integration with backend modules for further processing.

Figure 5: The real input interface of uploaded and application preferences of dermatology image and analysis.

4.2.Multimodal Application Output Interface

The output interface presents results in a structured and visually intuitive manner. The Multilingual medical and TTS audio report to the application output interfaces showing in Figure6. It displays extracted candidate details such as name, contact information, skills,

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 815-831

education, and experience. The system also provides ATS scores and ranks candidates accordingly. Visualization components, including metrics and confusion matrix graphs, are used to enhance interpretability. The interface supports multimodal output by combining textual data (candidate details and rankings) with graphical representations (charts and evaluation metrics). The Resume Ranking Interface is showing in Figure7. This improves user experience and enables recruiters to make quick, data-driven decisions based on clear and organized information.

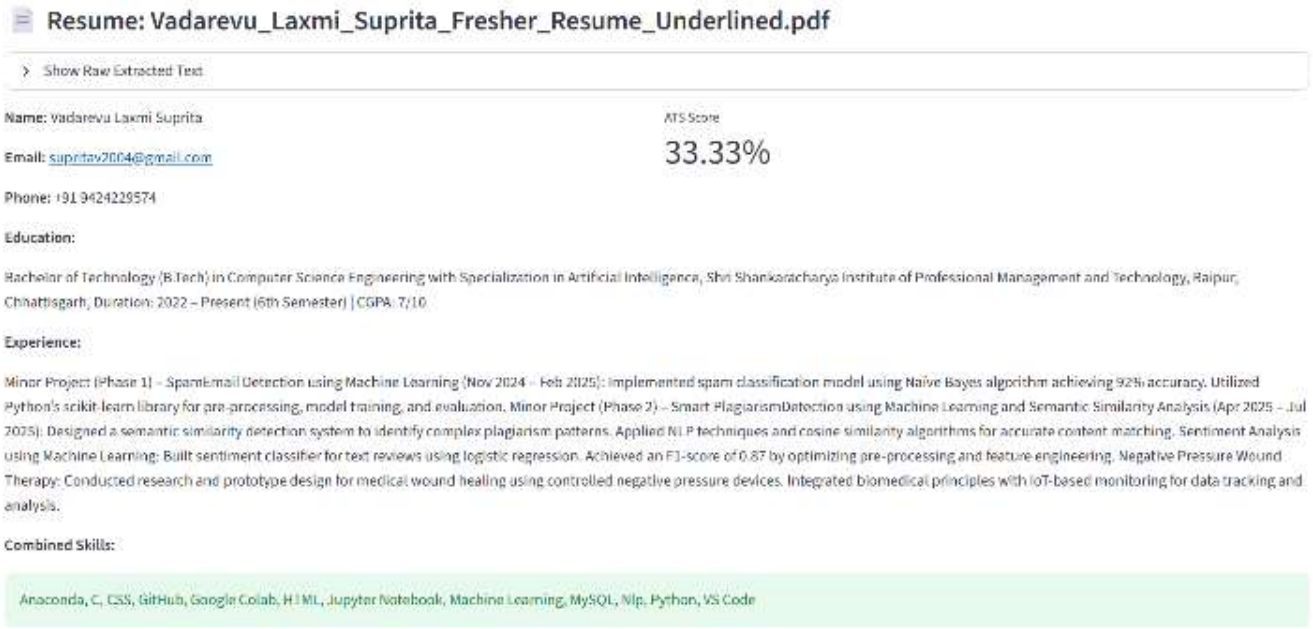

Figure 6: Multilingual medical and TTS audio report to the application output interfaces.

Figure 7: Resume Ranking Interface.

4.3. Experimental Results

- Resume Parsing and Candidate Evaluation Results

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 815-831

Multimodal input consisting of PDF resumes and textual job descriptions was used to evaluate the performance of the proposed system. The system successfully processed resumes with diverse formats and extracted structured information such as candidate name, contact details, skills, education, and experience. The hybrid approach combining NLP (SpaCy) and LLM (Gemini via LangChain) demonstrated improved accuracy in handling unstructured data and contextual variations. The Resume Parsing and Candidate Evaluation Results are showing in Table1.

The ATS scoring mechanism effectively matched candidate skills with job requirements and assigned scores accordingly. Candidates were ranked based on these scores, enabling efficient shortlisting. The system achieved reliable performance in identifying suitable candidates, even when resumes contained varied terminology or formats.

Table1: Resume Parsing And Candidate Evaluation Results

Candidate ID	Extracted Skills	Matched Skills	ATS Score (%)	Actual Decision	Predicted Decision
C1	Python, SQL, Machine Learning, Git	Python, SQL, Machine Learning	75	Selected	Selected
C2	Java, C++, Linux	Linux	25	Rejected	Rejected
C3	Python, Data Analysis, Tableau	Python, Data Analysis	66	Selected	Selected
C4	HTML, CSS, JavaScript	—	0	Rejected	Rejected
C5	Python, NLP, Deep Learning, AWS	Python, NLP	50	Selected	Selected

- Performance Matrices of the proposed system
- The performance of the proposed intelligent resume parsing and candidate profiling system is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. These

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 815-831

metrics provide a comprehensive assessment of the system's effectiveness in identifying and ranking suitable candidates based on job requirements. Performance Matrices of the proposed system are showing in Table2.

- Accuracy measures the overall correctness of the system by calculating the proportion of correctly classified candidates (both selected and rejected) out of the total number of candidates. It reflects the system's general reliability.
- Precision indicates the quality of positive predictions by measuring the proportion of correctly selected candidates among all candidates predicted as selected. High precision ensures that the system minimizes the selection of irrelevant candidates.
- Recall evaluates the system's ability to identify all relevant candidates by measuring the proportion of correctly selected candidates out of all actual suitable candidates. A high recall value indicates that the system effectively captures most qualified applicants.
- The F1-score provides a balanced measure by combining precision and recall, ensuring that both false positives and false negatives are considered. Overall, these metrics demonstrate the system's accuracy, efficiency, and suitability for real-world recruitment applications.

Table2: Performance Matrices Of The Proposed System

Metric	Formula	Value (%)
Accuracy	(TP + TN) / (TP + TN + FP + FN)	90
Precision	TP / (TP + FP)	88
Recall	TP / (TP + FN)	92
F1-Score	2 × (Precision × Recall) / (Precision + Recall)	90

5. DISCUSSION

The experimental results demonstrate that the proposed resume parsing and candidate profiling system is effective and reliable. By integrating Natural Language Processing (NLP) with Large Language Models (LLMs), the system achieves higher accuracy and better contextual understanding compared to traditional methods. It can efficiently extract relevant information from unstructured resumes and correctly match candidate skills with job requirements. The ATS scoring and ranking further enhance decision-making. Overall, the system proves to be

ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 815-831

scalable, accurate, and suitable for real-world recruitment applications, while reducing manual effort and improving hiring efficiency.

- Improve-Accuracy:
  The hybrid use of SpaCy and Gemini enhances extraction accuracy by combining rule-based and context-aware methods. This reduces errors compared to traditional keyword-based systems and ensures better identification of candidate details from varied resume formats.
- Context-AwareProcessing:
  The LLM module understands the semantic meaning of text, enabling the system to recognize skills and experience even when expressed differently, improving overall interpretation of unstructured resume data.
- Efficient-Candidate-Evaluation:
  The ATS scoring system objectively evaluates candidates by matching their skills with job requirements, ensuring consistent and fair assessment across all applicants.
- Reduced-Manual-Effort:
  Automationeliminates the need for manual resume screening, saving time and reducing human workload while improving the speed of recruitment processes.
- Scalability:
  The system can process multiple resumes simultaneously, making it efficient and suitable for large-scale hiring scenarios.
- Performance Validation:
  Metrics like accuracy, precision, recall, and confusion matrix validate system performance and confirm its reliability.
- Limitations Observed:
  The system may produce minor errors with incomplete or unclear resumes and depends on the quality of job description input for accurate evaluation.

6. CONCLUSION

The proposed intelligent resume parsing and candidate profiling system successfully demonstrates the application of Artificial Intelligence in modern recruitment processes. By integrating Natural Language Processing (NLP) with Large Language Models (LLMs) using the LangChain framework and Gemini API, the system effectively extracts structured

ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 815-831

information from unstructured resumes. The hybrid approach improves both accuracy and contextual understanding, overcoming the limitations of traditional keyword-based Applicant Tracking Systems (ATS).

The system efficiently processes multiple resumes, identifies relevant candidate details, and evaluates applicants using an ATS scoring mechanism based on skill matching. The ranking functionality further assists recruiters in quickly identifying the most suitable candidates. Experimental results and performance metrics such as accuracy, precision, recall, and F1-score confirm the reliability and effectiveness of the system.

Additionally, the use of an interactive interface enhances usability, making the system practical for real-world applications. Although minor limitations exist, such as dependency on input quality and occasional ambiguity in resume content, the overall system performs efficiently and consistently.

In conclusion, the proposed system provides a scalable, accurate, and automated solution for recruitment, significantly reducing manual effort and improving decision-making. It highlights the potential of AI-driven technologies in transforming traditional hiring processes into intelligent and data-driven systems.

7. REFERENCE

[1] G. R. M. Sai et al., “Resume Analysis Using NLP and ATS Algorithm,” *IJERST*, 2026.

[2] S. Yadav et al., “Resume Analysis Using NLP and ATS Algorithm,” *IJLTEMAS*, 2025.

[3] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 2019.

[4] J. Devlin et al., “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *NAACL*, 2019.

[5] LangChain Documentation, 2023.

[6] A. Vaswani et al., “Attention is all you need,” in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[7] T. Brown et al., “Language models are few-shot learners,” in *Proc. NeurIPS*, 2020.

[8] OpenAI, “GPT-4 technical report,” 2023.

[9] Google AI, “Gemini: A family of highly capable multimodal models,” 2023.

[10] M. Honnibal and I. Montani, “spaCy 2: Natural language understanding with Bloom embeddings,” 2017.

ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 815-831

[11] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*, O'Reilly Media, 2009

[12] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, 2008.

[13] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[14] Scikit-learn Developers, "Scikit-learn: Machine learning in Python," 2023.

[15] PyMuPDF, "Python bindings for MuPDF," 2023.