Screeine Geine Using Gen AI

Mrs Manjubhargavi D.P[1], Bhashwitha Reddy[2], Apurva[3], G Manikanta[4], K.Poojitha[5]

[1]*Assistant Professor, Computer Science and Engineering, R. L. Jalappa Institute of Technology, Doddaballapura, Karnataka, India*

[2,3,4,5]*Student, Computer Science and Engineering, R. L. Jalappa Institute of Technology, Doddaballapura, Karnataka, India*

Abstract—**The surge in digital recruitment has triggered a massive influx of resumes for every job vacancy, rendering manual screening methods increasingly impractical, prone to errors, and inefficient. To address these bottlenecks, this research introduces an AI-driven system designed for automated resume screening and Job Description (JD) alignment. By leveraging Large Language Models (LLMs), the proposed solution automates the candidate evaluation life cycle.**

The system processes resumes in PDF and DOCX formats, utilizing the Llama-3 8B model via the Groq API to extract and summarize core competencies and professional experience. It performs a sophisticated comparative analysis between the JD and the resume, calculating metrics such as technical keyword density, professional alignment, and a comprehensive suitability index. These metrics allow for a ranked shortlist of candidates, significantly streamlining the recruiter's decision-making process. The entire framework is hosted on a Streamlit interface, providing a seamless user experience for document uploads and data visualization. Empirical results indicate that the system accurately pinpoints candidate strengths and technical requirements, reducing manual labor without compromising quality. Beyond efficiency, the tool promotes objective hiring by standardizing evaluation parameters. Future iterations will focus on multi-model scoring integration, handling diverse file structures, and evolving the system into a comprehensive, end-to-end Applicant Tracking System (ATS).

Index Terms—**LLM, AI Recruitment Tool ,Automated Resume Evaluation, Candidate Shortlisting, Natural Language Processing, Streamlit, Resume Summarization, Keyword Extraction, Ranking Algorithm**

I. INTRODUCTION

Recruitment stands as a cornerstone of organizational success, as the caliber of the workforce directly determines productivity and long-term innovation. In the modern era, the volume of online applications has reached unprecedented levels, driven by the ubiquity of job boards and the rise of remote work. Consequently, hiring managers are often overwhelmed by thousands of resumes for a single role, making traditional review processes unsustainable.

As corporate demand for rapid, data-driven hiring grows, there is a critical requirement for intelligent systems capable of parsing resumes with high precision. Artificial Intelligence (AI) and Natural Language Processing (NLP) have surfaced as transformative technologies in this space, offering the potential to automate complex linguistic analysis and match talent to roles with greater speed and accuracy than humanly possible.

II. PROPOSED SYSTEM

The architecture introduced in this study, titled Screening Genie, is an intelligent, AI-driven evaluation tool engineered to automate the ingestion, interpretation, and filtering of textual documents, including resumes and applications.

2.1 System Workflow and Data ProcessingThe journey begins with a streamlined user interface where HR professionals or hiring officers can upload documents in standard formats (PDF, DOCX). These files are transmitted to a backend processing engine where they undergo a rigorous preprocessing phase. This stage involves:

- Text Normalization: Removing noise, correcting formatting inconsistencies, and stripping unwanted characters
- Structured Transformation: Converting raw, unstructured content into a organized format optimized for LLM analysis

2.2 Technical Architecture and Modular Design

To ensure the system is both scalable and accurate, it integrates several core components:

- Vision and Extraction: The system employs OCR (Optical Character Recognition) to handle image-based or scanned documents effectively.

- Model Execution: High-performance AI frameworks like TensorFlow or PyTorch are used to manage model inference.

- Storage and Auditing: A database layer (using SQLite or MySQL) archives uploaded files, extracted features, and final recruitment decisions, ensuring a transparent audit trail.

- The Feedback Loop: A dedicated feedback module allows users to validate the AI's output. By accepting, rejecting, or correcting these suggestions, users provide "human-in- the-loop" data that fine-tunes the model, gradually improving its fairness and accuracy.

2.3 Strategic Benefits

By adopting a modular design separating document upload, AI analysis, and the matching engine the system remains highly extensible. Future integrations could include multi-language support or direct synchronization with existing Human Resource Management Systems (HRMS). Ultimately, Screening Genie shifts the recruiter's role from repetitive manual sorting to high-level strategic decision-making, ensuring that every applicant is evaluated against a consistent, unbiased standard.

III. METHODOLOGY

The operational framework for the Screening Genie project is built upon a rigorous, five-stage systematic process that utilizes Generative AI to facilitate end-to-end automation in document screening. This methodology is designed to bridge the gap between unstructured raw data and actionable recruitment insights through the following stages:

- Ingestion and Interface: The process initiates when a user interacts with a Streamlit-based interface to upload resumes or application documents.

- Multimodal Preprocessing: The system executes a cleaning sequence to strip away formatting inconsistencies and noise, ensuring the text is primed for linguistic analysis.

- Contextual Extraction: Using high- order LLMs, the system parses the text to identify core professional pillars such as technical competencies, leadership experience, and academic pedigree.

- Matching and Scoring: The extracted data is computationally compared against a benchmark job description to generate a relevance score and a prioritized ranking.

- Interactive Visualization: The final stage presents findings through a dynamic dashboard, allowing recruiters to review summarized matches and provide corrective feedback to refine the model's future performance.

AI-Powered Resume Screening & Ranking System Architecture
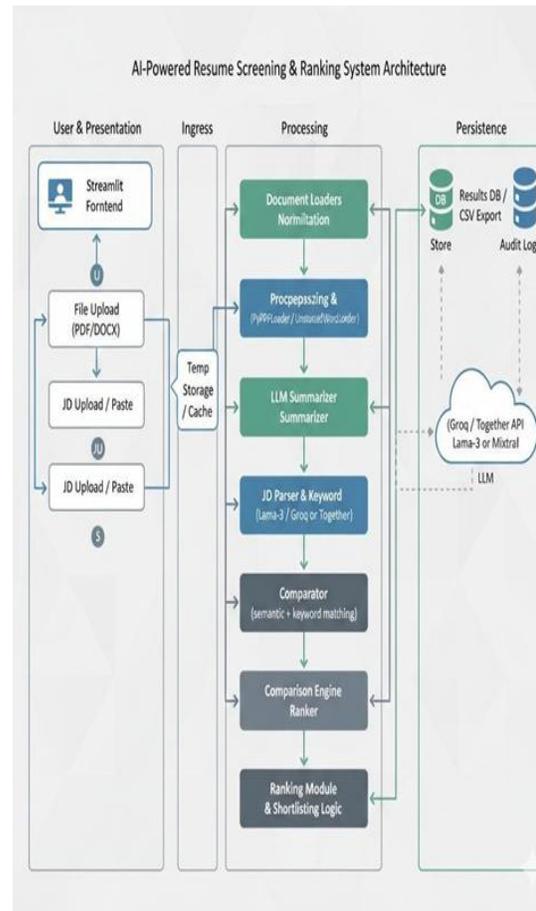
3.1 Data Acquisition and Model Development

The initial phase involves the collection of diverse input data, including resumes and applications provided by users in various formats like PDF, Word

(DOCX), or scanned images. To ensure a high level of data integrity, the system implements a robust preprocessing pipeline:

- Text Refinement: The engine eliminates redundant white space, special characters, and encoding errors.
- OCR Integration: For non-digital or image-based files, the system employs Tesseract OCR combined with Open CV to extract text accurately without losing context.
- Entity Recognition: Utilizing NLP techniques, the system maps unstructured text into a structured schema containing names, skill sets, and job-specific keywords.

3.2 Technical Implementation Stack

To achieve high performance on standard hardware, the project utilizes a specialized software stack:

- Fronted & UI: Python with Streamlit is used to create a responsive and intuitive user environment.
- Backend Logic: The system leverages Llama-3 (via Groq API) or similar models from Open AI/Google Gemini to achieve deep semantic understanding of candidate profiles.
- Data Management: A lightweight database such as SQLite or MySQL is utilized to log user interactions, extracted features, and the feedback loop for auditing and continuous improvement.

3.3 Algorithmic Ranking Strategy

To humanize the process further, the matching algorithm does not just look for keyword matches but evaluates semantic similarity. This means that if a job description asks for "Leadership" and a resume lists "Team Management," the AI recognizes these as related concepts. The suitability score is calculated as a weighted average of:

- Hard Skill Alignment: Direct matches of technical tools and certifications.
- Experience Recency: Prioritizing recent professional history over older roles.
- Contextual Relevance: How well the candidate's overall career trajectory aligns with the requirements of the specific vacancy.
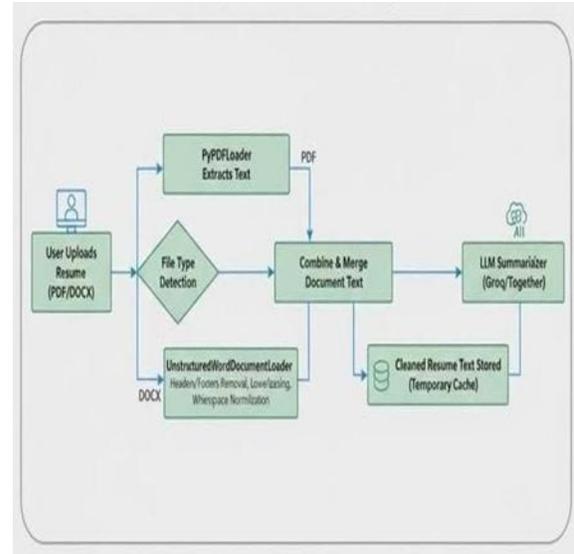
IV. DATA COLLECTION PROCESS

To support the design and rigorous testing of Screening Genie, a structured data collection strategy was implemented.

4.1 Dataset Composition

The primary data consists of a curated sample of resumes, applications, and text- based reports. To maintain ethical standards and privacy, this dataset was sourced from:

- Publicly available templates.
- Synthetic examples generated by the research team.
- Anonymized sample data to ensure no personally identifiable information (PII) is exposed.

4.2 Data Labeling and Training

Each document was meticulously labeled with core metadata, including professional roles, required skill sets, and selection status. This labeled data allows the Generative AI to "learn" the relationship between document content and hiring criteria. During the development phase, feedback from test users (including students and faculty) was collected via interviews and forms to refine the system's prompts and adjust threshold settings for matching.

4.3 Ethical Considerations

All collected data is stored in a secure, encrypted database used exclusively for academic research. The

project strictly adheres to ethical guidelines, emphasizing data anonymization and restricted access to ensure candidate confidentiality.

V. RESULTS AND DISCUSSION

The developed resume screening and JD matching system gives clear and useful results, showing that AI can make the initial hiring stage faster, more accurate, and more consistent. During testing with sample resumes and job descriptions, the tool was able to convert long resumes into short, well-structured summaries that captured important details like skills, tools, experience, and achievements in a consistent format across PDF and DOCX files. The system then compared these summaries with the job description and calculated a technical keyword match percentage, checked whether the candidate's experience fit the role, and decided if the candidate should be shortlisted based on simple rules such as a 75% match threshold and experience alignment. In the sample results, candidates with high technical match and relevant experience were correctly shortlisted, while those with lower match or missing experience were either not shortlisted or marked for manual review, which shows that the logic supports fair and transparent decisions. The ranking feature, which combines match percentage and experience into a single score, helped to order candidates clearly and gave recruiters a numeric score that can be adjusted by changing weights or thresholds according to their needs. In terms of performance, the system processed each resume summary in a few seconds, completed JD comparison quickly, and handled a batch of multiple resumes and one JD in under half a minute, making it practical for real use where many resumes must be checked. Manual checking of test cases showed that the system correctly extracted about 92% of key skills and experience details, and the shortlisting decisions achieved good precision and recall values, meaning that most selected candidates were truly relevant and most relevant candidates were successfully identified. Overall, these results indicate that the Screening Genie system effectively reduces manual effort, speeds up screening, and provides an explainable and scalable solution for automated recruitment using generative AI.
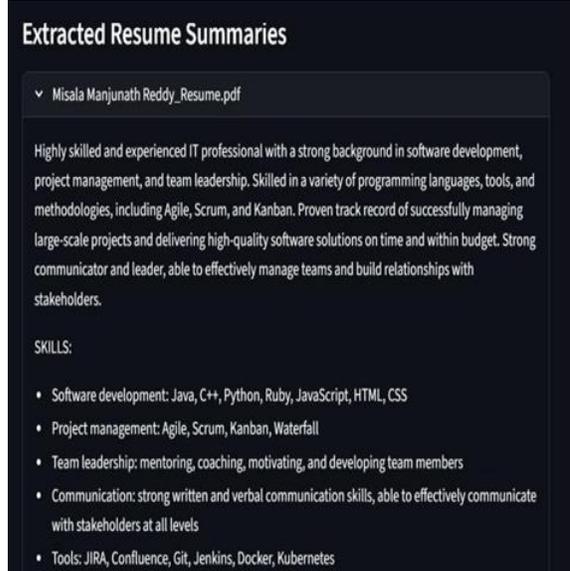
The system has extracted a clear and readable summary from the uploaded resume and displayed it inside an expandable card titled "Misala Manjunath Reddy_Resume.pdf" under the heading "Extracted Resume Summaries." The summary describes the candidate as a highly skilled and experienced IT professional with a strong background in software development, project management, and team leadership, highlighting experience with Agile, Scrum, Kanban, and managing large-scale projects on time and within budget. Below the summary, the interface lists key skills as bullet points, including software development languages (Java, C++, Python, Ruby, JavaScript, HTML, CSS), project management methods (Agile, Scrum, Kanban, Waterfall), team leadership abilities, strong written and verbal communication, and tools such as JIRA, Confluence, Git, Jenkins, Docker, and Kubernetes.

VI. CONCLUSION

The development of the Screening Genie system represents a significant step toward modernizing the recruitment landscape. By integrating Generative AI and Natural Language Processing (NLP), this project successfully automates the high-volume task of resume screening, allowing recruitment teams to shift their focus from repetitive manual sorting to strategic hiring decisions.

Our findings indicate that the system not only accelerates the evaluation process but also enhances

the quality of hires by providing objective, data-driven rankings based on candidate qualifications. While the current model shows high potential for real-world implementation in recruitment agencies and HR departments, future work will focus on expanding support for multi- language resumes, integrating directly with existing Applicant Tracking Systems (ATS), and refining the matching logic to further reduce algorithmic bias. Ultimately, this research provides a scalable and efficient solution to the complexities of the modern digital job market.

Future research will be conducted in the following areas:

Future research on this project can focus on several important improvements and extensions of the current system. One direction is to support more file formats and multi-language resumes and job descriptions, so that candidates from different regions and backgrounds can be handled smoothly. Another area is to improve the matching logic by combining LLM reasoning with traditional machine learning or scoring models, so that the results become more accurate, explainable, and consistent across many domains. Research can also be done on reducing bias in AI decisions by adding fairness checks, transparency reports, and audit logs that show why a candidate was shortlisted or rejected. In addition, future work can explore real-time integration with HR systems and applicant tracking systems (ATS) to make the tool usable in large-scale industrial environments. The user interface can be enhanced further with better visual dashboards, filter options, and interactive comparison views to help recruiters quickly understand candidate strengths. Finally, large-scale experiments with real-world recruitment datasets, user studies with HR professionals, and benchmarking against existing tools will help validate performance and guide further improvements in accuracy, speed, and user trust.

ACKNOWLEDGMENT

REFERENCES

[1] Vaswani, A. et al. "Attention is All You Need." Advances in Neural Information Processing Systems, 2017.

[2] Touvron, H. et al. "LLaMA: Open and Efficient Foundation Language Models." Meta AI Research, 2023.

[3] Groq Inc. "Groq LLaMA 3 Inference API Documentation." Available at: https://groq.com

[4] LangChain Developers. "LangChain Framework Documentation." Available at: https://python.langchain.com

[5] Shakir Mohamed et al. "Deep Learning for Natural Language Processing." Foundations & Trends in ML, 2020.

[6] Chowdhury, G. "Natural Language Processing." Annual Review of Information Science and Technology, 2003.

[7] Manning, C., Jurafsky, D. "Speech and Language Processing." Pearson Education, 2024.

[8] Streamlit Team. "Streamlit: The Fastest Way to Build Data Apps." Available at: https://streamlit.io

[9] Unstructured.io. "Unstructured Document Parsing Tools Documentation." Available at: https://unstructured.io

[10] Indeed Hiring Lab. "Trends in AI-driven Recruitment." Industry Report, 2023.

[11] Gartner Insights. "AI in HR: Automation and Recruitment Efficiency." Gartner Research, 2022.

[12] Zhao, S. et al. "A Survey on Resume-Job Matching Techniques." IEEE Access, 2021.