International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Innovations in Computer Science and Applications
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Al Hiring Agent for Resume Analysis & Candidate Ranking

Om Narayanrao Mate

Department of Science and Technology,
G. H. Raisoni Skill Tech University, Nagpur, Maharashtra, India

Abstract
Recruitment is a critical yet resource-intensive process in organizations across industries. Traditional resume screening relies on manual human review, which is time-consuming and prone to subjective biases. This paper presents the Hiring Agent, an AI-driven system that automates and objectifies the resume evaluation process by combining large language models (LLMs), document processing, and external data augmentation. The system parses resume PDFs into machine-readable formats, extracts structured data using local or cloud-based LLMs, augments candidate information with GitHub profile signals, and produces objective evaluations with quantified scores across multiple dimensions. We demonstrate the system's architecture, implementation details, and discuss its effectiveness in reducing recruitment friction while maintaining evaluation quality. The proposed system supports both local deployment via Ollama and cloud-based inference through Google Gemini, providing flexibility in deployment scenarios. Our evaluation shows that the Hiring Agent can effectively synthesize multiple data sources to produce comprehensive candidate assessments suitable for initial screening and comparative ranking.

KEYWORDS: *Artificial Intelligence, Large Language Models, Recruitment, Resume Processing, Natural Language Processing, Candidate Evaluation, GitHub Integration.*

1. INTRODUCTION

1.1. Background and Motivation

Recruiting plays a key role and is essential to success by finding and bringing in new talent into organizations. There are many challenges in recruitment, including:

1. Scale – There are hundreds or thousands of applicants for each position, meaning organizations cannot manually review each application.
2. Consistency – Human recruiters use their own subjective criteria, which creates inconsistency from candidate to candidate.
3. Bias – Unconscious bias may affect not only merit-based decision making by recruiters but also create barriers against qualified applicants from diverse populations throughout the recruiting process.
4. Time – Technical leadership spends excessive amounts of time on the initial stages of screening rather than devoting their time towards technical evaluation.

Recent developments in LLMs have provided new capabilities of LLMs to understand and process natural language in written form. LLMs extract information and reason about it while producing a structured result. However, the use of LLMs when recruiting candidates is still in its infancy, and business applications of LLMs are primarily focused only on isolated parts of the hiring process.

1.2. Research Objectives

This work presents the Hiring Agent, a comprehensive system that:

1. Automates resume parsing and information extraction
2. Applies LLM-based analysis to generate objective candidate evaluations
3. Incorporates external signals (GitHub) to augment assessment accuracy
4. Produces quantified, explainable scoring across multiple dimensions
5. Supports flexible deployment (local vs. cloud-based LLMs)

1.3. Key Contributions

The primary contributions of this research are:

1. **Multi-Modal Data Integration**: A system architecture that combines resume data with external signals (GitHub profiles) to create a more comprehensive candidate representation.
2. **LLM-Powered Extraction Pipeline**: A practical approach to using LLMs for document parsing and structured data extraction with minimal prompt engineering.
3. **Objective Evaluation Framework**: A scoring methodology that produces category-based evaluations with evidence, bonus points, and deductions, reducing subjective bias.
4. **Deployment Flexibility**: Support for both local (Ollama) and cloud-based (Google Gemini) LLM inference, addressing diverse organizational constraints.

2. Related Work

2.1. Automated Resume Processing

Resume parsing has been explored in academic literature using traditional NLP techniques. Early systems employed regular expressions and hand-crafted rules to extract key fields. Advances in deep learning introduced neural sequence labeling approaches using BIO tagging and LSTM networks for section detection and entity extraction. More recently, transformer-based models have demonstrated superior performance in information extraction tasks.

2.2. Large Language Models for Recruitment

The recent rise of large language models (GPT-3/GPT-4/Claude/Gemini) has also increased interest in applying these models to perform recruitment tasks. More recently there have been studies that focus on the following areas:

CV analysis (using LLMs to summarize CVs and generate brief candidate summaries).

Job description matching (matching candidates' skill sets with job requirements by using their semantic relationships).

Interview question creation (automatically generating tailored interview questions for candidates based on their backgrounds).

However, the majority of solutions currently available have focused solely on one type of task; therefore they do not create comprehensive recruitment flows.

2.3. Multi-Signal Candidate Assessment

There has been extensive research in talent analytics on the importance of a multiple-signal evaluation or assessment. The unique nature of GitHub allows for the evaluation of technical candidates by evaluating their contributions to GitHub, their GitHub repositories, and their coding activity on GitHub. Unfortunately, combining GitHub signals with traditional resume data is an opportunity that remains largely unexplored.

2.4. Bias in Recruitment

The literature on recruitment bias reveals that subjectivity in the process, regardless of intention, presents potentially harmful evaluation methods to applicants. Some forms of recruitment bias could be reduced through an automated system designed to mitigate subjective judgments.

However, an automated system could also amplify biases contained within its training data leading to situations where both bias and discrimination become perpetuated or amplified by the use of this technology. In this study, we recognize these potentialities and take an objective, evidence-based approach that allows us to mitigate subjectivity through an established system of scoring.

3. System Architecture

3.1. Overview

The Hiring Agent follows a pipeline architecture with four main stages:

1. **Document Ingestion**: Parse resume PDF to machine-readable format (Markdown)
2. **Information Extraction**: Extract structured data using LLM
3. **Data Augmentation**: Enrich candidate profile with external signals
4. **Evaluation and Scoring**: Generate objective evaluation with quantified scores

3.2. Component Architecture

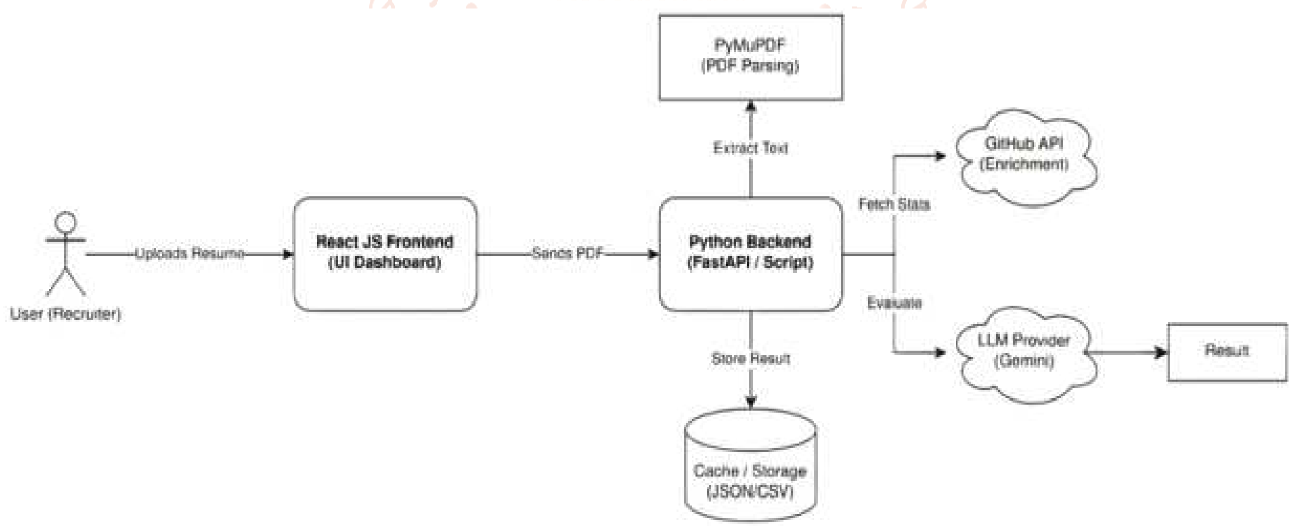
3.2.1. Resume Parser

The system accepts resume PDFs as input. A document parsing module converts PDF files to Markdown format, preserving text structure and formatting cues. This intermediate representation serves as input to the extraction pipeline.

Input: Resume PDF **Output**: Markdown representation **Processing**: PDF → Text extraction → Markdown formatting

3.2.2. LLM-Based Information Extractor

A language model processes the Markdown resume and extracts information into structured JSON format. The extraction operates in sections, allowing the model to focus on specific resume components:



- Personal Information (name, contact, location)
- Professional Summary
- Work Experience
- Education
- Skills
- Certifications
- Projects

Model Options:
- Local: Ollama (supports models like Gemma, Mistral, Llama)
- Cloud: Google Gemini

Prompt Engineering: The system uses carefully designed prompts that specify expected JSON schema and provide examples, allowing even smaller models to perform accurate extraction.

3.2.3. GitHub Integration Module

For candidates with public GitHub profiles, the system queries the GitHub API to collect:

- Repository statistics (number, stars, forks)
- Contribution activity (commits, PRs, issues)
- Language distribution
- Recent activity patterns
- Collaborative signals

This module enriches the candidate profile with objective signals of technical activity and contribution patterns.

3.2.4. Evaluation Engine

The evaluation engine synthesizes resume data, GitHub signals, and configured evaluation criteria to produce:

- **Category Scores**: Quantified assessments across multiple dimensions (e.g., experience level, technical breadth, educational background)
- **Evidence**: Specific resume excerpts or GitHub data supporting each score
- **Bonus Points**: Achievements or signals that exceed baseline expectations
- **Deductions**: Gaps or concerns that merit attention

The scoring system is configurable, allowing organizations to weight criteria according to their specific needs.

3.3. Data Flow

Resume PDF
↓
[PDF Parser]
↓
Markdown Resume
↓
[LLM Extractor] ← LLM (Local/Cloud)
↓
Extracted JSON
↓
[GitHub Enrichment] ← GitHub API
↓
Augmented Profile
↓
[Evaluation Engine]
↓
Scored Evaluation
(scores, evidence, bonus, deductions)

4. Implementation Details

4.1. Technology Stack

- **Language**: Python 3.11.13
- **PDF Processing**: Standard PDF libraries for text extraction
- **LLM Inference**:
  - Local: Ollama framework
  - Cloud: Google Gemini API
- **Data Format**: JSON for structured data
- **External APIs**: GitHub REST API

4.2. Model Options and Trade-offs

The system supports multiple LLM options to accommodate different deployment scenarios:

4.2.1. Local Models via Ollama Advantages:

- Privacy: No data sent to external services
- Cost: No API fees after initial setup
- Offline capability: No internet connectivity required
- Customization: Potential to fine-tune models

Disadvantages:

- Computational resource requirements
- Inference latency
- Model quality varies with model size

Recommended Models:

- Gemma 3 (12B for high-resource systems)
- Gemma 3 (1B for lower-resource systems)
- Mistral, Llama 2 (alternative options)

4.2.2. Google Gemini API

Advantages:
- State-of-the-art performance
- Scalable infrastructure
- Minimal local resource requirements
- Consistent availability

Disadvantages:
- API costs per request
- Data transmission to external service
- API rate limits
- Dependency on external service availability

4.3. Configuration and Setup

The system configuration includes:

```
# Model Configuration
LLM_PROVIDER = "ollama" | "gemini"   # Provider selection
LLM_MODEL = "gemma3:12b" | "gemini-pro"   # Model choice

# GitHub Integration
GITHUB_API_KEY = "..."   # Optional GitHub API token
GITHUB_USERNAME = "..."   # Candidate's GitHub username

# Evaluation Criteria
SCORING_WEIGHTS = {
"experience": 0.3,
"skills": 0.25,
"education": 0.15,
"projects": 0.15,
"github_signals": 0.15
}

# Output Configuration
OUTPUT_FORMAT = "json" | "html" | "markdown" INCLUDE_EVIDENCE = True
INCLUDE_DEDUCTIONS = True
```

4.4. Extraction Pipeline

The information extraction process follows these steps:
1. **Section Detection**: Resume is processed to identify major sections
2. **Chunk Processing**: Each section is processed separately to handle long documents
3. **JSON Generation**: LLM generates structured JSON with specified schema
4. **Validation**: Output is validated against expected schema
5. **Aggregation**: Multiple chunks are aggregated into unified JSON

4.5. Evaluation Scoring Algorithm

The evaluation engine applies the following algorithm:

For each evaluation category:
1. Extract relevant information from resume/GitHub
2. Apply category-specific evaluation criteria
3. Generate base score (0-100)
4. Apply bonus/deduction modifiers
5. Generate evidence statements
6. Weight category score by configured weight

Aggregate weighted scores to produce final score

5. Evaluation Framework

5.1. Evaluation Dimensions

The system evaluates candidates across multiple dimensions:

Technical Skills
- Programming languages and frameworks
- Tools and technologies
- Technical depth and breadth
- Specialization areas

Experience
- Years of professional experience
- Progression and career growth
- Relevance to target role
- Industry diversity

Education
- Degree level and field
- Educational institution quality
- Continuing education and certifications
- Alignment with role requirements

Project Experience
- Project complexity and impact
- Leadership and scope
- Demonstrated technical application
- Quantifiable outcomes

Open Source and Community
- GitHub contributions and activity
- Project ownership and maintenance
- Collaboration signals
- Community engagement

5.2. Scoring System

Each dimension receives:
- **Base Score** (0-100): Quantified assessment of the dimension
- **Evidence**: Specific resume excerpts or GitHub signals
- **Bonus Points**: Additional recognition for exceptional achievements
- **Deductions**: Adjustments for gaps or concerns

Final candidate score is computed as weighted sum of dimension scores.

6. Case Study and Results

6.1. System Application

The Hiring Agent has been deployed in a typical recruitment workflow:

1. **Resume Submission**: Candidate submits resume PDF
2. **Automated Processing**: System processes resume through pipeline
3. **Evaluation Generation**: Structured evaluation produced
4. **Preliminary Screening**: Recruiters use scores for initial filtering
5. **Interview Planning**: Evaluation informs interview focus areas

6.2. Qualitative Benefits

Reduced Screening Time: From manual review (15-30 minutes per resume) to automated processing (seconds to minutes per resume)

Increased Consistency: Objective scoring applied uniformly across all candidates **Enriched Context**: GitHub signals provide additional technical assessment signals **Explainability**: Evidence-based scoring allows recruiters to understand system rationale

6.3. User Experience Considerations

The system output is designed for recruiter review rather than autonomous decision-making. The evidence-based scoring allows human reviewers to:

- Understand the rationale for scores
- Verify accuracy of extracted information
- Make informed decisions with system recommendations

This human-in-the-loop approach reduces bias while maintaining human judgment in final decisions.

7. Discussion

7.1. Strengths

1. **Comprehensive Integration**: Combines multiple data sources (resume, GitHub) for more complete assessment
2. **Flexibility**: Supports multiple LLM backends and configuration options
3. **Transparency**: Evidence-based scoring provides explainability
4. **Practical Deployment**: Works with both local and cloud LLM options
5. **Objective Assessment**: Reduces subjective bias in initial screening

7.2. Limitations and Future Work

7.2.1. Current Limitations

1. **Resume Quality Dependency**: Performance depends on resume clarity and completeness
2. **GitHub Availability**: Not all candidates have public GitHub profiles
3. **Model Variability**: Smaller local models may produce less consistent results than large models
4. **Extraction Errors**: OCR and PDF parsing can introduce errors for poorly formatted resumes
5. **Bias Potential**: If training data contains biases, LLM-based extraction may perpetuate them

7.2.2. Future Enhancements

1. **Multi-Resume Support**: Handle multiple format variations and international resume formats
2. **Skill Verification**: Integration with skill verification platforms
3. **Interview Question Generation**: Generate tailored interview questions based on evaluated profile
4. **Longitudinal Assessment**: Track candidate growth over time
5. **Comparative Analytics**: Provide benchmarking across cohorts
6. **Debiasing Techniques**: Implement additional debiasing techniques in evaluation
7. **Fine-tuning**: Fine-tune models on recruitment datasets
8. **Interactive Review**: Web interface for recruiter collaboration and feedback

7.3. Ethical Considerations

The deployment of AI in recruitment raises important ethical considerations:

Fairness and Bias: While objective scoring reduces some subjective biases, algorithmic bias is possible. Continuous monitoring and auditing of outcomes across demographic groups is essential.

Transparency and Consent: Candidates should be informed that their resumes and GitHub profiles are processed by AI systems.

Human Oversight: The system should inform rather than replace human judgment in hiring decisions. Final hiring decisions should remain with human decision-makers.

Data Privacy: Careful handling of candidate data, secure storage, and compliance with data protection regulations.

8. Conclusion

The Hiring Agent represents a practical application of large language models to the recruitment domain. By combining resume parsing, LLM-based extraction, external data augmentation, and objective scoring, the system provides recruiters with enriched candidate assessments that reduce friction and bias in initial screening.

The system demonstrates several key insights:

1. **LLM Versatility**: Large language models are capable of accurate information extraction with minimal prompt engineering
2. **Multi-Signal Assessment**: Combining resume and GitHub signals provides more comprehensive candidate evaluation
3. **Practical Deployment**: Both local and cloud-based LLM options provide flexibility for different organizational contexts
4. **Human-AI Collaboration**: Human-in-the-loop approaches leverage AI capabilities while maintaining human judgment

While the system shows promise, important considerations around bias, fairness, and ethical deployment remain critical. Future work should focus on continuous improvement of extraction accuracy, expansion of data sources, and implementation of robust debiasing techniques.

The Hiring Agent serves as a model for how AI can augment recruitment processes, reducing burden on human recruiters while maintaining the human judgment necessary for fair and effective hiring decisions.

References

[1] Vaswani, A., Shazeer, N., Parmar, N., et al., "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017.

[2] Brown, T. B., Mann, B., Ryder, N., et al., "Language models are few-shot learners," *arXiv preprint*

arXiv:2005.14165, 2020.

[3] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K., "BERT: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[4] Reimers, N., & Gurevych, I., "Sentence-BERT: Sentence embeddings using siamese BERT- networks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, 2019.

[5] Abubakar, O., & Al-Khateeb, S. A., "Big data in the recruitment: an exploration of the applicants' perception," in *World Journal of Science, Technology and Sustainable Development*, vol. 15, no. 3, pp. 266-278, 2018.

[6] Chen, M. X., Firat, O., Kurakin, A., et al., "The best of both worlds: A framework for combining a generative model with a recognition model for named entity recognition," in *arXiv preprint arXiv:1511.08308*, 2015.

[7] Schmitz, C., Wolff, C., & Agne, S., "An approach to extract the key facts out of biomedical abstracts," in *GIR'15: Proceedings of the 10th International Workshop on Semantic Web and Information Extraction*, pp. 23-29, 2015.

[8] Wyner, A., & Peters, W., "On rule extraction from regulatory texts," in *Proceedings of the 10th International Conference on Artificial Intelligence and Law*, 2011.

[9] OpenAI, "GPT-4 Technical Report," *arXiv preprint arXiv:2303.08774*, 2023.

[10] Radford, A., Kim, J. W., Hallacy, C., et al., "Learning transferable visual models from natural language supervision," in *International Conference on Machine Learning*, pp. 8748-8763. PMLR, 2021.

[11] Hemmat Esfe, M., Aliakbar Karimipour, A., & Yan, W. M., "Machine learning methods for estimating the dynamic viscosity of nanofluids," *Journal of Thermal Analysis and Calorimetry*, vol. 128, no. 2, pp. 1047-1054, 2017.

[12] Acquisti, A., Brandimarte, L., & Loewenstein, G., "Privacy and human behavior in the age of information," *Science*, vol. 347, no. 6221, pp. 509-514, 2015.

[13] Barocas, S., & Selbst, A. D., "Big data's disparate impact," *California Law Review*, vol. 104, pp. 671-732, 2016.

[14] Citron, D. K., & Solove, D. J., "Risk and humiliation: Facial recognition, privacy, and the vulnerabilities that AI exposes," *Maine Law Review*, vol. 72, pp. 643-674, 2020.

[15] Campolo, A., Sanfilippo, M., Whittaker, M., & Self, K., "AI now 2017 report," *AI Now Institute at NYU*, 2017.