



AI-POWERED RESUME SCREENING AND JOB RECOMMENDATION SYSTEM

Ms. Gawane R.R.¹, Shravani Tele²

¹ Faculty Guide & Supervisor | ² Dattakala Group of Institutions, Faculty of Engineering, Pune, Maharashtra, India, India

Corresponding author: shravanitele2203@gmail.com

ARTICLE ID IJREK-2026-00112	ARTICLE TYPE Research Article	RESEARCH AREA Artificial Intelligence	PUBLICATION Volume 1, Issue Issue 1 (2026)
RECEIVED 19 October 2026	REVISED 23 October 2026	ACCEPTED 20 October 2026	PUBLISHED 24 October 2026

Abstract

Recruitment teams may receive a large volume of resumes for a vacancy, making manual screening time-consuming and inconsistent. Job seekers may also struggle to identify vacancies that align with their skills, education, experience, and interests. This paper proposes an AI-powered system that extracts structured information from resumes, compares candidate profiles with job descriptions, and recommends relevant opportunities. The pipeline combines document parsing, text preprocessing, skill extraction, TF-IDF or sentence-embedding representations, and a transparent matching score. Candidate-facing results explain why a job was recommended, while recruiter-facing results support—rather than replace—human review. The design includes privacy controls, bias evaluation, and mechanisms for correcting extracted information. A reproducible evaluation plan uses extraction quality, Precision@K, Recall@K, ranking metrics, and subgroup error analysis. No experimental results are claimed because the proposed system has not been benchmarked for this manuscript.

Keywords: Artificial Intelligence; Resume Screening; Job Recommendation; NLP; Information Retrieval; Skill Matching; Fairness

1.1 PROBLEM STATEMENT

Design a system that extracts useful information from diverse resumes and ranks job opportunities by relevance while limiting extraction errors, unfair bias, privacy risks, and misleading recommendations. 1.2 Objectives • Extract skills, education, experience, and relevant terms from resumes. • Represent job descriptions and candidate profiles in a comparable form. • Generate explainable job-match scores and ranked recommendations. • Evaluate extraction and ranking quality using documented test data. • Include privacy, fairness, correction, and human-review safeguards. 1.3 Scope This paper describes an academic prototype for resume parsing and job discovery. It does not recommend rejecting applicants automatically, infer protected characteristics, or replace interviews, validated assessments, or recruiter judgment.

2. LITERATURE REVIEW

Information retrieval provides foundational methods for ranking documents by relevance. TF-IDF and cosine similarity are common baselines for matching resume text with job descriptions because they are simple and interpretable. They can underweight context and synonyms. Semantic embeddings may represent related phrases more closely, but must be evaluated for quality and bias. Resume parsing is sensitive to tables, columns, scanned pages, and inconsistent section headings. Job recommendation is affected by vacancy freshness, incomplete descriptions, location preferences, and changing career goals. Research on algorithmic decision-making also warns that historical hiring data can encode past inequalities; excluding protected attributes does not necessarily remove proxy effects.

2.1 RESEARCH GAP

A useful prototype should not focus only on ranking accuracy. It should combine parsing with explainable matching, candidate feedback, privacy-by-design, bias checks, and human oversight.

3. PROPOSED SYSTEM

The candidate workflow allows a user to upload a resume, review extracted profile fields, set preferences, and receive ranked vacancies with explanations. The recruiter workflow allows an authorised user to create a vacancy and review matching profiles as one input to structured human assessment.

Figure 1. Conceptual system workflow: Resume Upload + Job Listings → Document Parsing → Text Cleaning & Skill Extraction → Candidate / Job Profiles → Matching and Ranking Engine → Explainable Recommendations → Human Review

3.1 MAIN MODULES

TABLE I

Module	Purpose
Resume ingestion	Validate file type, size, and consent.
Document parsing	Extract text and detect malformed or scanned documents.
Profile extraction	Identify skills, education, experience, and preferences.
Job catalogue	Store job descriptions, skills, location, experience, and freshness.
Matching engine	Calculate lexical and semantic relevance.
Explanation and feedback	Show matching factors and allow corrections.
Admin and audit	Manage access, retention, versions, and system events.

3.2 SYSTEM ARCHITECTURE

A prototype may use Python and FastAPI for APIs, PostgreSQL for structured data, private object storage for uploaded documents, and React for the user interface. File processing should use size limits and isolated execution. Resumes must not be exposed through public URLs. Candidates should access their own profiles, and recruiters should only access applications they are authorised to review. 4. Methodology

4.1 DATA COLLECTION AND PREPARATION

Evaluation data can be assembled from consented, anonymised sample resumes and job descriptions whose usage terms permit analysis. Synthetic examples can test edge cases but should not be presented as representative real-world evidence. Remove duplicates, document sources and dates, and separate evaluation data by candidate or vacancy group to avoid nearduplicate leakage.

4.2 RESUME PARSING AND PREPROCESSING

- Validate file type, size, and extractable text before processing.
- Extract PDF/DOCX text and use OCR only for scanned documents.
- Normalise whitespace and section headings while retaining source text for correction.
- Identify skills with dictionaries, phrase matching, or a trained entity-extraction model.
- Mark uncertain fields for candidate verification.

4.3 CANDIDATE AND JOB REPRESENTATION

A baseline can represent resume and vacancy text with TF-IDF word and phrase features and calculate cosine similarity. A semantic model may produce sentence embeddings. Structured features such as required-skill coverage, experience range, location, and employment type can be scored separately. Protected characteristics should not be used to rank candidates.

4.4 MATCHING SCORE

A transparent score can combine skill overlap, text relevance, experience fit, and explicit preferences: $MatchScore = w1 \times SkillCoverage + w2 \times TextSimilarity + w3 \times ExperienceFit + w4 \times PreferenceFit$. Weights should be selected using validation data and documented. The score is a relevance estimate, not a probability of job performance. Missing information should be shown as unknown rather than automatically treated as negative.

4.5 RECOMMENDATION ALGORITHM

1. Receive and validate resume.
2. Extract text and candidate-approved fields.
3. Retrieve active jobs matching basic preferences.
4. Compute similarity and structured scores.
5. Rank vacancies and generate explanations.
6. Display results for review and correction.
7. Record only necessary feedback and audit events.
8. Evaluate ranking quality and fairness periodically.

5. IMPLEMENTATION PLAN

TABLE II		
Layer	Suggested technology	Responsibility
Frontend	React / TypeScript	Upload, profile correction, filters, recommendations.
Backend API	Python / FastAPI	Authentication, profile endpoints, matching.
Database	PostgreSQL	Profiles, jobs, preferences, feedback, audit metadata.
Document storage	Private object storage	Secure resume storage and deletion controls.
NLP / matching	scikit-learn; optional embedding model	Parsing, features, similarity, ranking.
Testing	pytest and evaluation scripts	API, parsing, ranking, and security tests.

5.1 PRIVACY AND SECURITY

- Collect only necessary data and obtain informed consent.
- Encrypt data in transit and at rest; restrict resume access by role.
- Allow correction of extracted details and deletion where applicable.
- Set retention periods for resumes, profiles, and feedback.
- Do not expose personal details or credentials in logs.
- Keep human review for consequential hiring decisions.

6. EVALUATION PLAN AND DISCUSSION

TABLE III		
Metric		Purpose
Field extraction	precision/recall	Correctness and completeness of extracted entities.
Precision@K		Fraction of top K recommended jobs judged relevant.
Recall@K		Fraction of relevant jobs found in top K.
NDCG@K		Ranking quality with graded relevance.
Coverage and diversity		Whether recommendations cover varied suitable roles.
Fairness error analysis		Whether errors or outcomes differ across lawfully assessed groups.
Latency		Time required to parse and recommend.

7. ADVANTAGES AND LIMITATIONS

7.1 Advantages

- Reduces repetitive job searching and organises large job catalogues.
- Supports skill discovery across different resume wording.
- Explains matching factors and allows profile correction.
- Uses modular components that can be tested independently.
- 7.2 Limitations
- Scanned or poorly formatted resumes can cause extraction errors.
- Similarity does not measure true ability or future job performance.
- Incomplete or outdated job descriptions produce weak recommendations.
- Models may contain bias or fail on unfamiliar terms.
- Privacy, accessibility, and fairness require continuous review.

8. ETHICAL CONSIDERATIONS

Recruitment data is sensitive and can affect employment opportunities. Provide clear notice about data use and retention. Candidates should inspect and correct extracted information. Recruiters should not treat ranking as an objective verdict. Protected attributes and inferred personal characteristics should not determine suitability. Real hiring use should be reviewed against applicable law and organisational policy.

9. FUTURE SCOPE

Future work may include multilingual parsing, improved OCR, semantic retrieval, candidate-controlled preferences, vacancy freshness checks, duplicate-job detection, and accessible interfaces. Further study may evaluate counterfactual fairness, ranking calibration, privacy-preserving analytics, and user experience. Extensions should be validated with representative data and independent review. 10. Conclusion This paper proposed an AI-powered resume screening and job recommendation system combining resume parsing, skill extraction, text and semantic matching, ranked job discovery, and human review. It emphasises explainability, correction, privacy, fairness evaluation, and reproducible metrics. The design can support an academic prototype, but its effectiveness and suitability for recruitment must be demonstrated through careful testing and responsible governance.

REFERENCES

[1] Manning, C. D., Raghavan, P., and Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press. <https://nlp.stanford.edu/IR-book/>

[2] Ricci, F, Rokach, L., and Shapira, B. (Eds.). (2011). Recommender Systems Handbook. Springer. <https://doi.org/10.1007/978-0-387- 85820-3>

[3] Raghavan, M., Barocas, S., Kleinberg, J., and Levy, K. (2020). Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices. Proceedings of FAT* 2020, 469–481. <https://doi.org/10.1145/3351095.3372828>

[4] Mehrabi, N., Morstatter, F, Saxena, N., Lerman, K., and Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. ACM Computing Surveys, 54(6), Article 115. <https://doi.org/10.1145/3457607>

[5] scikit-learn developers. Feature extraction documentation. https://scikit-learn.org/stable/modules/feature_extraction.html

[6] European Union. (2024). Artificial Intelligence Act. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>

Publication archival note: Official scholarly contribution published under CC BY 4.0 International License. Permanent record on <https://www.ijrek.org> | Article ID: IJREK-2026-00112