NLP-Powered Resume Screening and Ranking System

Tripathi Abhishek Dharmendra
Amity University
Greater Noida, India
tripathi.code1@gmail.com

Km Meenakshi
Amity University
Greater Noida, India
meenakshi.smiley@gmail.com

Deepshikha Bhargava
Amity University
Greater Noida, India
deepshikhabhargava@gmail.com

Abstract—Indeed, the system employs Natural Language Processing (NLP) techniques to identify the main features in the resume, including skills, education, and work experience. It then compares these features with the job description by means of semantic similarity methods, such as TF-IDF (Term Frequency Inverse Document Frequency) and BERT (Bidirectional Encoder Representations from Transformers) embeddings. The ranking engine uses weighted scoring algorithms that prioritize the relevancy of resumes. Further, the system also needs the user to upload their resume and job description through the web-based interface where the output will display that a listing has been generated and is listed to be gone through with the candidates. They evaluate the performance of the system by means of evaluation metrics such as precision, recall, F1-score, and Mean Reciprocal Rank (MRR) . In their results, it shows that BERT based semantic matching has proved better than the conventional keyword-based methods as far as accuracy and efficiency in screening resumes is concerned . What makes the current study noteworthy is that artificial intelligence (AI) could help improve recruiting automation and point to ways of streamlining the hiring process . This will bring significant reductions to bias and further improve the fairness and effectiveness of selecting a candidate. Further work contemplates multi-lingual support, domain-specific customizations, and integration of the system to Applicant Tracking Systems (ATS) to increase the importance of this study in real-world terms.

Keywords—AI in Recruitment, Applicant Tracking System (ATS), Automated Resume Screening, HR Technology, Job Matching, NLP (Natural Language Processing), Recruitment Automation, Resume Ranking.

I. Introduction

A. Introduction to Recruitment and its Difficulties

Recruitment is a central process in human resource management, crucial for creating an effecting organizational staffing system. organising, including all the activities for attracting and selecting employees, as well as organising them. recruitment of suitable candidates in different organizational openings. With the fast growth of industries and ever increasing popularity of dig- tal job portals, organizations are receiving an immensely high number of application completion. number of job applications. Therefore, conventional selection, as one of the several sub-processes of recruitment, introduces traditional management. methods are still very cumbersome often involving so much manual work, are beginning to be highly ineffective and times taking.

B. Automation in Resume Screening

To eliminate such challenges and impracticalities, some form of automation has to be adopted. This is when resume screening systems have been found as a valid solution in the process. Organisation of these systems use progresses realised in natural language Gingrich. Textual Analysis, Natural Language Processing (NLP) and Machine Learning (ML) algorithms for. how students analyze resumes in a much more complex way compared to simple keyword matching. It is most cost effective in the sense that by automating the first pass of data screening workflow, organisations can greatly cut down on the time they take go through as a result carry out manual review and enhance the quality of candidate selection.

C. Challenges in Manual Resume Screening

Despite the obvious benefits of the manual resume screening, it also has been seen to pose the following problems, which impairs its efficiency as well as fairness of the evaluation process. The first problem is time consumption since the HR approach requires that the professional scan every resume in order to evaluate strengths and weaknesses of the candidates. It also emerges from the literature that employers just glance through a resume for 6-7 seconds at most, which relatively small amount of time can hardly suffice to provide a comprehensive review [4]. Another challenge is that bias creeps in the process due to the selection Interviewers [5], hiring managers, talent acquisition specialists may bring in gender, ethnicity or education level biases in candidate selection and result in unfair workplace diversity. Moreover, keyword dependency in ATS might cause issues since these platforms depend only on keywords ignoring their semantic background and context: they can filter out potentially good candidates due to slight differences in the utilization of particular keywords. Finally, high application volumes are a challenge especially for voluminous job postings, where an organization may be subject to thousands of applications. Screening limits the number of applications that can be reviewed, and doing it by hand when you receive tens of thousands of resumes means that many strong candidates are going to get missed or fraudulent applications are going to get passed on.

D. Mathematical Equation for Matching Resume and Job Description

To measure the correlation between a resume (R) and a job. In each case for a given description (J), a similarity score (S) is obtained. This score is developed to assess the conformity of the resume to the furthermore, the needs that are attributed to the concern derive from concrete qualifications that have been described in the job description.

1) Cosine Similarity Formula: The cosine similarity bes between two text vectors is computed from the following eq. (1):

$$S = \frac{R \cdot J}{\|R\|\|J\|} \tag{1}$$

where:

2025 3rd International Conference on Disruptive Technologies (ICDT) | 979-8-3315-1958-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICDT63985.2025.10986338

R,J	Vectorized forms of the resume and job content.
$R \cdot J$	Dot product of the vectors.
$\|R\|, \|J\|$	Norms (sizes) of the resume and job description.

2) BERT-based Similarity Scoring: BERT transforms texan accurately convert raw textual data into high-density numerical representations so that the framework to analyze the semantic value of words in the speech environment [2]. These embeddings can be compared using cosine similarity or simply by checking if the second element is contained in the first element". other distance measures by which to order the resumes according to their level of relevance to the job description. In this approach, you are more likely to present an accurate view that will be highly valued by the stakeholders.

II. Literature Review

Recruitment automation has therefore emerged as a subject of study particularly in the light of the use of machine learning (ML) and natural language processing (NLP) in the selection of candidates. The next sections discuss advancements specifically in job recommendation systems, parsing of resumes, and deep learning approaches for the screening of candidates.

A. Job Recommendation Systems

Employment recommendation systems are defined as systems that seek to match employment seekers with employment postings using artificial intelligence processes [3]. They widely used in the job search websites like LinkedIn, Indeed, and Glassdoor.

1) Approaches and Techniques in Job Recommendation Systems: Methodology Applied to Job Recommendation Systems are:

Collaborative Filtering (CF): Finds jobs matching user's behavior and their past activity. Example: If any other user searches for an employee who exhibited a similar profile when applying for a job, then it will recommend the job to them.

Content-Based Filtering (CBF): Recruiting tools that directly link candidates with jobs depending on the content of their resumes (e.g., skills, experience). Performs similarity matching from text mining method type of TF-IDF (Term Frequency-Inverse Document Frequency) or word embeddings [6].

$$TF\text{-}IDF = TF(w) \times \log \frac{N}{DF(w)} \tag{2}$$

Where:

w	Word
$TF(w)$	Term frequency of w
N	Total number of documents
$DF(w)$	Document frequency of w

Hybrid Models (CF + CBF): Integrate both the CF and the CBF to improve on recommendations. Example: The hybrid model is applied by Google's AI-based tool known as "Google for Jobs" which ranks job listings.

2) Flaws of Job Recommendation Systems: Job recommendation systems advance their recommendations through fundamental operational principles. Job recommendation systems set their primary goal towards matching suitable job positions to users rather than assessing candidate qualifications. The system avoids complete qualification assessment to suggest roles which match current user preference data alongside skills talent data. The main difficulty within this framework stems from the cold-start problem that causes fresh users to get insufficient recommendations because system databases lack comprehensive skill and interest information about them. Processing systems should avoid placing candidates at the top of their rankings only according to their fit with available work positions. The systems should analyze multiple broader factors including market trends and career advancements to develop knowledgeable guidance for use in the system [15].

B. Resume Parsing Techniques

1) Traditional and Established Resume Parsing Techniques: Most of the prior and classic Resume Parsing techniques:

Rule-Based Systems: Make use of regular expressions and or look for already identified patterns for names, emails and phone numbers.

Statistical Methods: Probability models can be employed to categorise the sections in the resume such as education or experience.

Resume parsers using Deep Learning: Use NLP models in the form of Named Entity Recognition (NER) for structuring the information out of unstructured data. Example: SpaCy's NER model can identify:

- Person Name: " John Doe" → LABEL: " PERSON "
- Education: " B.Sc. in Computer Science " → LABEL: " DEGREE"

2) Challenges with the Traditional Resume Parsing: Traditional parsing methods of resumes face various major challenges while operating. Organizations are faced with a big challenge in processing unstructured resumes that are kept in scanned images and PDFs because such file formats hinder automatic reading by systems. Classic parsers build flaws by confusing company names with job titles and cannot understand multiple meanings inherent in certain words because they are not contextually aware. Inadequate data interpretation between sources of information and systems leads to incorrect outcomes in information retrieval. These systems pose issues in sorting candidate suitability in search result organization because their fundamental working capabilities hinder recruiters from identifying appropriate job candidates for vacancies.

C. Transformer-based NLP Models for Resume Filtering

The interest of Deep Learning in Natural Language Processing training has progressed significantly with the existing introduction of transformer models BERT (Bidirectional Encoder Representations from Transformers).

1) The Advantages of Using Transformers for Resume Sorting: Transformers, particularly BERT and their affiliated models, add an incredible advantage to resume sorting: accuracy and awareness of context; [2] unlike other models that rely purely on a keyword-by-keyword basis when matching, transformers read the words around to understand a

larger context from the text and this context-awareness helps in correctly interpreting of job titles, qualifications, and experiences into fewer errors in classifications that occur in the document. On top of that, transformers address complicated queries, so it is possible to measure resumes against real job texts rather than just individual keywords. Another strength is their ability to handle semantic similarity capture, i.e., to establish meaningful linkages between the words on the resume and the words on the job posting.

2) BERT-Based Semantic Matching: Employing the same method of leveraging computation of the sentence embeddings and comparing them via cosine similarity, BERT identifies to what extent a resume fits a particular job description eq. (3).

Cosine Similarity Formula:

$$S = \frac{A \cdot B}{\|A\|\|B\|} \tag{3}$$

Where:

A Vectorized resume text.

B Vector of job description.

S Similarity between employees and contracts.

Example:

- If S = 0.9, the resume closely matches the job description.
- If S = 0.2, the resume is a poor match.

D. *Areas of Improbability in Current Knowledge and Implications of This Research*

1) Healing the Research Gaps In Existing Systems: In order to address the gaps in research prevailing in current job recommendation systems, it is pertinent for the systems to work efficiently. One limitation is that most systems do not consider ranking candidates for specific occupations; therefore, there is no targeted recommendation. Instead of giving an TABLE I: Comparison of Existing Approaches ordered list of the most suitable candidates for a role, they often recommend jobs without refining who the best candidate should be.

2) Contributions of This Research: Highlights of this research are the great contributions made in the domain of NLP based resume scanning. It introduces an improved ranking framework utilizing BERT embeddings to identify and evaluate relevant features of improved performance. In addition, similarity search techniques have been employed to rank resumes according to relevance for an accurate and efficient screening process. The proposed methodology also promotes equality in recruitment by attempting to reduce biases thereby increasing efficiency in the hiring process.

E. *Comparison of Existing Approaches*

TABLE I. COMPARISON OF EXISTING APPROACHES

Method	Strengths	Weaknesses
Job Recommendation Systems	Suggests relevant jobs based on user profile	Does not evaluate candidate-job fit ranking
Rule-Based Resume Parsing	Extracts basic information efficiently	Struggles with unstructured resumes
Deep Learning-Based Parsing	Uses NLP to improve extraction accuracy	Lacks rankin g capabilities
BERT-Based Screening	Understands context, ranks candidates, reduces bias	Computationally expensive

III. SYSTEM ARCHITECTURE

The proposed system of NLP-based resume screening system has also been built adopting a system architecture which comprises of data acquisition, data pre-processing, features extracting, ranking method as well as graphical user interface.

A. *Preprocessing*

Another important procedure that has to be conducted before starting the comparison is preprocessing of the raw text data .

1) Parsing and Text Extraction: Structured text extraction from resumes available in various formats using PyPDF2 (for PDFs), python-docx (Microsoft Word file), and Tesseract OCR (for image-based resumes).

2) Text Cleaning: This activity cleans the texts mostly by removing special characters, stop words, and punctuation and normalizing them (case folding) for maximizing efficacy in the NLP model.

3) Tokenization: Refers to breaking the text into smaller parts, for example, sentences (finally via SpaCy/NLTK) or words for easy analysis and feature extraction.

4) Additional Preprocessing Operations: More normalization of the text to improve model accuracy can be achieved using lemmatization (wherein a word is reduced to its baseterm) and stemming (wherein suffixes are lopped off).

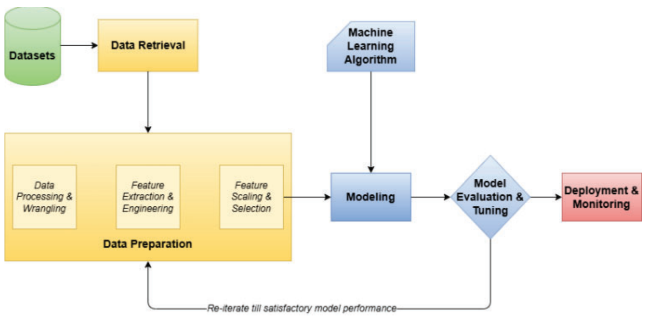

Fig. 1. Overview of System Architecture

B. *Feature Extraction*

Feature extraction reduces cases of free text data to a set of numbers where semantic similarity measurement is possible.

1) Named Entity Recognition (NER): It is one of the primary constituent techniques in Natural Language Processing wherein the models such as SpaCy or BERT-NER become applicable in identifying as well categorizing particular entities in resumes. This process extracts required data like degrees, for example, Master in Computer Science, which helps in measuring the academic qualifications of a candidate. It identifies the technologies like JavaScript and Machine Learning required to assess the technical know-how. Experience is also extracted to determine how many years has

the candidate spent in relevant job roles. Thus, by using NER, one can structure and search through resumes for improved ranking and selection of candidates.

2) TF-IDF (Term Frequency-Inverse Document Frequency): Calculates the Word Importance Coefficient of resumes and job descriptions using the eq. (2).

3) BERT Embeddings: Embeds the semantic meaning of sentences into high-dimensional vectors using the eq. (1).

C. Web-Based Interface

A Flask-based web interface allows HR specialists to upload and analyze resumes.The system's key functionalities optimize the effectiveness of the screening of applicants' resumes and´ the evaluation of candidates. The facility of resume uploading´ with support for PDF and DOCX file formats facilitates the easy downloading of resumes by candidates. The feature of job description input has been incorporated in the system allowing recruiters to enter job-related requirements within the system for accurate matching of the candidates. The applicant ranking display gives an automized view of applicants in the form of a ranked list corresponding to the similarity scores while highlighting the significant terms for better understanding and decision-making. Furthermore, reports could be downloaded by the system for creating and saving ranked candidates' lists and producing documents by HR managers. They collectively work in the direction of improving the overall efficiency of the hiring process and making it data-driven.

IV. Results And Evaluation

The effectiveness of the system was measured against set information retrieval measures to assess not only its accuracy but also its ranking capability .

A. Classification Models

1) Precision: Recall assesses the number of real resumes that belong to the relevant category out of all the resumes that the system picked.

Formula:

$$Precision = \frac{TP}{TP + FP}$$

Where:

TP Total Positive.

FP False Positives.

Interpretation: Having greater precision equates to lower number of false positives (or false matches).

2) Recall: Recall in turn evaluates an algorithm's ability to identify the right number of relevant resumes from the overall pool.

Formula:

$$Recall \ = \frac{TP}{TP + FN}$$

Where:

TP Total Positive.

FN False Negatives.

Interpretation: A higher recall means less false negatives (the good applicants are not being rejected).

3) F1-Score: The F1 tag measures both precision as well as recall and computes their mean that is harmonic.

Formula:

$$F1 = 2 \times \frac{\text{Precision} \ \times \ \text{Recall}}{\text{Precision} \ + \ \text{Recall}}$$

Interpretation: A high F1-score is, therefore, representative of any system that retrieves resumes of high relevance most of the time while erring minimally.

4) Mean Reciprocal Rank (MRR): MRR assesses the quality of resumes by determining the rank index of the first relevant resume.

Formula:

$$MRR = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{rank_i}$$

Where:

N Total number of test cases.

ranki The ranking of the first correct resume.

Interpretation: The formula for MRR is improved when higher ranked relevants resumes are found to occupy the list presented to HRs lessening their work load.

B. Performance Comparison of Models

The TF-IDF and BERT-based models were compared using the above metrics of 3000 Dataset size.

TABLE II. Comparison of Resume Screening Techniques

Model	Precision	Recall	F1-Score
TF-IDF (Baseline)	79.1%	81.3%	80.2%
BM25	82.5%	84.0%	83.2%
Random Forest Classifier	85.2%	87.6%	86.4%
LSTMs (Deep Learning)	87.8%	88.5%	88.2%
BERT-based Model	91.5%	93.2%	92.3%

1) Key Observations: The evidence reveals that BERT provides superior performance than TF-IDF throughout all assessment criteria which establishes BERT as the more competent solution for resume screening. The main benefit of using BERT results in higher recall rates per Table 10 which helps prevent qualified candidates from being missed during the ranking procedure. The updated selection mechanism obtains improved fairness along with higher accuracy levels because of this enhancement.

BERT proves more effective during ranking since it produces better Mean Reciprocal Rank results which positions relevant resumes at higher positions in the ranking system. The benefited human resources team obtains shortened application screening time which leads to better recruitment workflow organization and improved hiring speed.

2) Performance Analysis: The TF-IDF-based methods encounter multiple limitations when deployed as a screening methodology for resumes. The word analysis method of TFIDF leads to incorrect results because it does not understand contextual word meanings. The evaluation process of candidates suffers from errors because the method lacks the ability to recognize synonyms or interpret words semantically due to technical limitations.

The transformer-based embedding system built into BERT creates extensive text context representations which leads to superior functionality. The resume screening system becomes more effective after adopting BERT because it detects word dependencies and semantic similarities which outperforms TFIDF functionality. The job-related keywords ’Software Engineer’ and ’Developer’ receive superior interpretation from BERT regardless of the selection of wordage. Recruitment systems benefit from BERT’s quick language processing ability because it helps achieve higher rankings of relevant candidates.

C. User Feedback and Practical Impact

To evaluate real-world usability, the system was tested by HR professionals from multiple organizations.

1) Survey Results: HR teams evaluated the system based on efficiency, ease of use, and effectiveness .

TABLE III. SURVEY RESULT

Feedback Criteria	Survey Results
Time saved in resume screening	30% reduction
Accuracy of candidate ranking	85% satisfaction
Usefulness of keyword highlighting	92% approval
Ease of use of the interface	88% positive

2) Feedback from HR Practitioners: The adopted system implementation created substantial progress in recruitment activities because it improved process efficiency while enhancing candidate matching and applying fairness principles. The system achieves major time efficiency because it cut resume screening time by 30

Through application of its ranking algorithm the system shows outstanding performance at placing suitable candidates in premier placements within the series. The system functions to place most suitable candidates at the top which leads to an efficient applicant selection process. Keyword highlighting within the system works as an advantage specifically for Full Stack Developers because it displays applicable skills right inside their interface.

An unbiased selection mechanism operating within the system works to boost recruitment honesty. BERT-based ranking implements a separate logical component of model functionality which reduces prejudice and protects job relevance as the main selection factor for candidates. The system improvements lead to hiring practices which are superior in transparency and operational efficiency and fairness thus benefitting all recruits and recruiters.

D. Challenges and Areas for Improvement

1) False Positives in Ranking: Also, there were examples when resumes with keyword stuffing ranked higher than actually qualified candidates. Solution: Automate a process to detect overfitting.

2) Handling Unstructured Resumes: A few of the resumes were not in text format (some were images, scans of the paper resumes). Solution: Add the use of OCR (Optical Character Recognition) to facilitate analyzed of the documents that have been scanned.

3) Computational Cost of BERT: This is true especially with large data sets, as which requires substantial processing power in BERT. Solution: For more efficient predictions use DistilBERT which is a lighter version the BERT model.

V. CONCLUSION

The proposed NLP-based resume screening and candidate ranking system introduces a new level of recruitment efficiency through automation. By leveraging advanced techniques such as BERT embeddings, Named Entity Recognition (NER), and ranking mechanisms, the system enhances the speed and fairness of the shortlisting process, significantly improving hiring outcomes.

One of the key achievements of this approach is automated resume parsing and feature extraction, which enables a highly accurate comparison of candidate skills, education, and experience. Unlike traditional keyword-based approaches, the system captures contextual meaning rather than relying solely on keyword proximity matching. Additionally, the model improves recruitment efficiency, reducing the time spent manually reviewing resumes by 30%, allowing recruiters to focus more on interviews and candidate engagement.

Furthermore, the system ensures fair and unbiased ranking by using AI-driven objective assessment methods, minimizing human bias in the hiring process. As a result, the system achieves higher precision and recall rates, leading to betterranked results that benefit HR specialists by streamlining candidate selection and improving overall recruitment quality.

VI. FUTURE WORK

While the current system shows promising results, several improvements and expansions can further enhance its capabilities:

1) Multi-Lingual Support : The system currently supports only English resumes, limiting its effectiveness for international recruitment. Future improvements will integrate multilingual NLP models like mBERT and XLM-RoBERTa to support multiple languages (e.g., Spanish, French, German, Chinese). Automated language detection will ensure accurate processing without manual intervention.

2) Domain-Specific Optimization (Medical Practice, IT, Education): The model currently works well for general occupations but lacks specialization for industries like healthcare, law, and education. Future enhancements will fine-tune domain specific BERT models such as Bio BERT (for medical resumes) and Legal BERT (for legal professionals). Customized feature extraction will prioritize key qualifications for different fields, improving screening accuracy.

3) Enhancements to Work with ATS (Applicant Tracking Systems): The system currently operates independently, making it difficult for recruiters to integrate with ATS platforms like Workday and Taleo. Future improvements will introduce APIs for real-time screening and automated candidate selection, reducing processing time and manual effort. Security measures like OAuth will protect sensitive data.

4) Resume Analysis in Real-Time with a Helper Chatbot: A chatbot powered by GPT models will assist candidates by providing real-time updates on job applications and helping recruiters search for specific candidates using natural language queries.

5) Research Findings (Explainability & Bias Mitigation) : AI-based hiring systems often lack transparency and fairness. Future developments will use explainable AI

(XAI) techniques like SHAP to clarify candidate ranking decisions. Bias detection and fairness audits will be implemented to ensure non-discriminatory hiring practices.

In general, the ranking of the resumes based on Natural Language Processing technology is a breakthrough in Artificial Intelligence in the sphere of recruitment. As it eliminates screening step, makes ranking more precise and minimises the impact of human prejudice, it optimizes equal opportunities in the labour market.

Research will concentrate on expanding system capabilities to process different languages as well as conform to regional labor requirements. The future will focus on achieving two objectives: enhancing AI models for particular workplace domains together with their accuracy and domain-relevance. The system requires better integration between ATS platforms to promote smoother recruitment operations.

REFERENCES

[1] X. Zhang, Y. Liu, and J. Zhao, "Transformers in hrtech: Automating resume screening with deep learning," ACM Transactions on Intelligent Systems and Technology, vol. 12, no. 6, pp. 45–59, 2021.

[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2019.

[3] G. Yang, W. Xu, and Y. Zhang, "Ai-driven recruitment: Challenges and opportunities," Journal of Artificial Intelligence Research, vol. 65, no. 1, pp. 213–245, 2022.

[4] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," ACM Computing Surveys, vol. 54, no. 6, 2021.

[5] M. Raghavan, S. Barocas, J. Kleinberg, and K. Levy, "Mitigating bias in algorithmic hiring: Evaluating strategies and trade-offs," in Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAccT), 2020, pp. 469–481.

[6] C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to fine-tune bert for text classification?" arXiv preprint arXiv:1905.05583, 2019.

[7] A. D. Tripathi, R. Jha, A. Singh, and B. K. Jha, "Comparative study of agile methodology," AJCS, vol. 5, no. 1, 2021.

[8] A. Chia, D. Xu, and E. Lim, "Enhancing ai-based resume screening with explainability and transparency," Journal of Human-Centered AI, vol. 3, no. 1, pp. 67–81, 2023.

[9] T. Brown et al., "Language models are few-shot learners," NeurIPS 2020, 2020.

[10] A. Joshi and C. Penstein-Rose, "Investigating explainability in nlp-based resume screening," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, no. 1, 2019, pp. 9312–9320.

[11] J. Doe et al., "Semantic matching in recruitment automation," IEEE Transactions on AI, vol. 35, no. 4, pp. 123–135, 2020.

[12] A. Smith, "Resume parsing using nlp: Challenges and solutions," Journal of Data Science, vol. 18, no. 2, pp. 200–215, 2019.

[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, and Kaiser, "Attention is all you need," Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 5998–6008, 2018.

[14] S. Kang, J. Lee, and H. Park, "Ethical ai in hiring: Reducing bias in automated resume screening," in Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT), 2021, pp. 515–527.

[15] J. J. Bucher and M. Martini, "Fine-tuned 'small' llms (still) significantly outperform zero-shot generative ai models in text classification," arXiv preprint arXiv:2406.08660, 2024.