Chapter 7

A Hybrid NLP and Transformer-Based System for Resume Analysis and Interview Question Generation

Sheela S Maharajpet
Department of MCA, Acharya Institute of Technology, Bangalore – 560107, India
Dhananjaya S.M
Department of MCA, Acharya Institute of Technology, Bangalore – 560107, India

Abstract: Job seekers cannot predict which interview questions may come from their resumes. Hence, preparing for this part of recruitment may be tough for them. An AI-based system overcomes this issue by examining resumes to create corresponding questions concerning competency and experience trends regarding the job. This paper presents an AI-Powered system that helps candidates to get ready by generating the personalized interview questions from their resume. The system first identifies the key sections in the resume such as skills, education, projects, and certification. Then it uses a fine-tuned language model (FLAN-T5) to create 5 to 10 relevant questions that match the candidate's resume. These questions are divided into types like Technical, Behavioral, and General to give balanced practice. To ensure the usefulness even when some parts are missing or unclear in the resume, fallback methods are used to still generate the helpful questions. Tests on a variety of resumes show that over 85% of the questions are relevant and improve candidate confidence. This AI-based approach provides a practical and personalized tool for job seekers to improve their interview preparation. It reduces guesswork and helps users focus on the most important parts of their experience, ultimately increasing their chances of success.

Keywords: AI-Based Interview, Resume screening, Machine Learning, Natural Language Processing

Citation: **Sheela S Maharajpet, Dhananjaya S M. A Hybrid NLP and Transformer-Based System for Resume Analysis and Interview Question Generation. Machine Learning in Research and Practice: A Multidisciplinary Perspective. Genome Publications. 2025; Pp61-70.**
https://doi.org/10.61096/978-81-990998-5-2_7

1. INTRODUCTION

The hiring process has evolved into a competitive, technology-operated landscape. Traditional manual methods have been complemented by digital platforms, online job boards and social networking, which can reach the global talent pool. However, the challenges remain in the management of large versions of the applicant data and ensuring justified candidates. AI has emerged as an important tool in recruitment, increases efficiency, fairness and scalability [1]. Online job portals have increased the application volume, manual resume screening is likely to be labor-intensity and prejudice. Keyword-based discoveries often fail to identify appropriate candidates due to boundaries in natural language processing. Automated re -start screening system, NLP and equality models, take advantage of the relevant candidate details and align them with job details, improve efficiency and fairness [2]. AI-Placed platforms increase recruitment by assessing the facial expressions, voice patterns and language of the candidates, reduce human bias [3]. However, moral concerns, such as prejudice and transparency, persist in the AI-based hiring system, require fair and accountable practices [4]. With their strong contextual representation, BERT-based models have driven NLP forward, helping to boost opportunities for resume parsing, job assignments, and skill extraction across fields. [5]AI devices also provide real -time response and individual job recommendations, although moral challenges [6] remain. By improving NLP keyword selection and reducing prejudice, reduces the incompetent in disabled screening, the fair candidate is enabled in shortlisting [7]. As awareness of algorithmic bias increases, research has also started to embrace awareness about fairness in the hiring phase. Recent reviews, have assessed fairness-aware approaches, adversarial debiasing, and explainable AI for facilitating fair hiring outcomes [8]

2. LITERATURE SURVEY

There have been recent advances in deep learning, natural language processing (NLP), and algorithms for data extraction, which provide shape and direction to the evolution of resume parsing. The diversity and complexity of present-day resumes pose challenges to traditional systems of rule-based parsing, thereby leading to inefficiency in candidate assessments. Recently developed methods make use of one or more deep learning frameworks, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers, to improve the precision of extracting structured information. However, challenges associated with these approaches still remain: how to resolve linguistic ambiguities; how to ensure fairness in treatment; and how to integrate parsing systems with applicant tracking systems. Further studies aim to pursue multilingual parsing and ethical AI methods within hiring [9].

There has been growing interest in automated job interviews as part of the recruitment process due to advancements in AI and NLP. Intelligent systems now assist in creating skill-based questions, adapted to the candidate's background. In their latest research, machine learning models are suggested for automating question generation and suggestion, making use of external knowledge base profiles as well as click-through data. Although these systems enhance interview efficiency, bias and question relevance are some of the major issues with such systems. The next studies aim to provide improvements in the NLP methods that further help in enhancing the accuracy and customization of questions asked in interviews [10].

The increasing popularity of AI and NLP in recruiting has boosted hiring efficiency. Many recent studies have brought to the focus the importance of AI-enabled chatbots in candidate screening, automating interview procedures, and improving candidate experience. A resume is analyzed by machine-learning algorithms, which match candidates with job specifications as pertinent to their skills and similarities. But problems remain, including bias, ethical issues, and context misinterpretation. Current studies are being conducted to make AI models more proficient in achieving equity, transparency, and accuracy in recruitment systems supported by AI [11].

To a large extent, artificial intelligence is greatly responsible for the modes of modern recruitment, by expediting a number of phases of the whole hiring process, which encompasses resume evaluation, candidate grading, and interview coordination. New research is showing the benefits of AI-based recruitment along with the ethical dilemmas that are raised. It's effective and, at the same time, impartial, however, it can also reinforce pre-existing biases if trained on skewed data. The argument in the literature calls for transparency, accountability, and fairness in AI-fueled recruitment for it to be considerate of ethical action. Specialists and researchers have asked for a standard approach to be laid down on prejudice alleviation and space for improvements in AI recruitment [12].

The application of AI and deep learning in resume assessment, undoubtedly, contributed tremendously to the enhancement of the hiring process, automating skill extraction. Recent research proposes a conductor of complex clustering models like deep feature-based K-means clustering refined with AGT to enhance the accuracy of resume classification. These use TF-IDF and GloVe-based vectorization for resume pre-processing and significant feature extraction using CGRUN. Similarly, active research is on dealing with different resumes' formats and clustering method refinement even with the so fine accuracy of the choices they are governing [13].

While AI brings an unprecedented transformation to the recruiting process, it has raised persistent questions of algorithmic bias. Studies point out the underlying reasons for AI-related bias in recruitment as unbalanced training data, faulty algorithms, and biased decision-making trends. Scientists propose various techniques with the aim of addressing bias compensation, such are fairness-aware machine learning, adversarial debiasing, and explainable AI to further establish fairness in job recruitment. Implementing these actions involves usefulness with regard to continuous surveillance, incorporation of multilateral datasets, and transparent management in AI. The ethical frameworks for AI in the future will impact fair and responsible recruitment decisions [14].

Recruitment systems using NLP methods facilitate automated and effective parsing, semantic, and candidate-job matching, saving manual effort in hiring by a significant amount. [15]. Recent approaches such as Resume2Vec have introduced intelligent resume embeddings that enhance candidate-job matching ability through deep semantic relationships significantly improving the accuracy of an applicant tracking system (ATS) [16].

Machine learning models have also been utilized to automate the process of resume screening and ranking, in which classifiers and scoring mechanisms rank applicants based on the relevance of a candidates experience level and level of skills, and reduce recruiting workload [17]. In order to increase contextual alignment, models such as Consultant BERT have also employed a fine-tuning process with a Siamese BERT architecture, where they more accurately match a candidate's profile to a job description, based on more recent job matching benchmarks, this is more robust than those traditional approaches that rely on basic similarity [18].

3. PROPOSED METHODOLOGY

The proposed system is based on a combined NLP and transformer-based system to analyze resumes and formulate interview questions from candidates profile. The approach consists of several stages, which are used to structure resumes into understanding and generate appropriate questions.

The procedure includes several main stages:

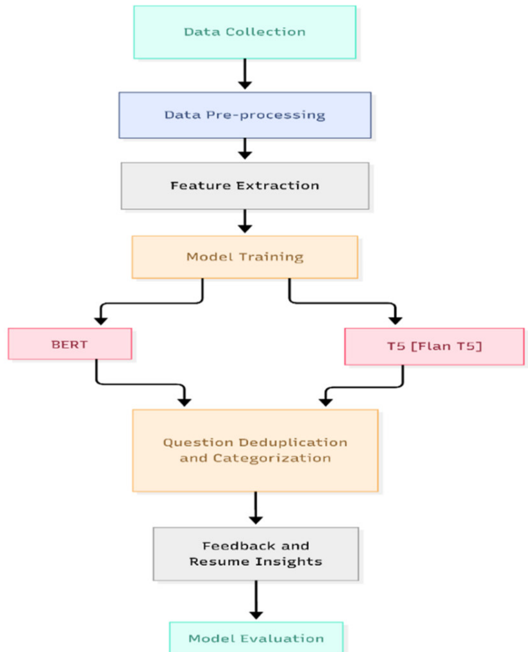

Figure 1: System Architecture

3.1 Data Collection

We collected a large dataset of resumes across multiple industries and roles, which include software engineering, data science, and system administration. The resumes will be used as input to the pipeline and the nature of these resumes reflects the variation of resumes one might see in the actual world. In addition to resumes, we will also be collecting job descriptions and more commonly used interview questions to help guide more question generation.

3.2 Data Pre-processing

Resume data was pre-processed using custom regex expressions, collections of NLP techniques. This includes simple text cleaning, normalization, and tokenization. The content that resumes provided valuable content to be structured in whole sections of content, (i.e. summaries, skills, education, experience, projects, certifications, etc.). Every section is evaluated when compared to its description for completeness using heuristics and token scoring methods.

3.3 Feature Extraction

For every section that we can detect, named entities as well as domain-relevant keywords (like programming languages, tools, frameworks, certificates, and degrees) are extracted using a mix of spaCy, regular expressions and domain-relevant keyword lists. These are the basis for improving suggestions and generating targeted questions.

3.4 Model Training

In our pipeline, we have established three transformer-based models, each dedicated to a specific task. The table below summarizes the components of each individual model, and the role of each model in the pipeline:

Table 1: Model Integration Table

Component	Model Used	Task
Resume Improvement Classifier	BERT	Multi-label classification
Interview Question Generator	FLAN-T5	Context-aware question generation
Deduplication	Sentence-BERT	Similarity detection and filtering

1. BERT: Bidirectional Encoder Representations from Transformers

Google created BERT in 2018 as a machine learning model representing plain text in other formats. Unlike past word embedding models, BERT represents text in both directions to gain an understanding of the left context and right context, and associated meaning. BERT improved significantly on several benchmarks related to question answering, sentiment analysis, and text classification.

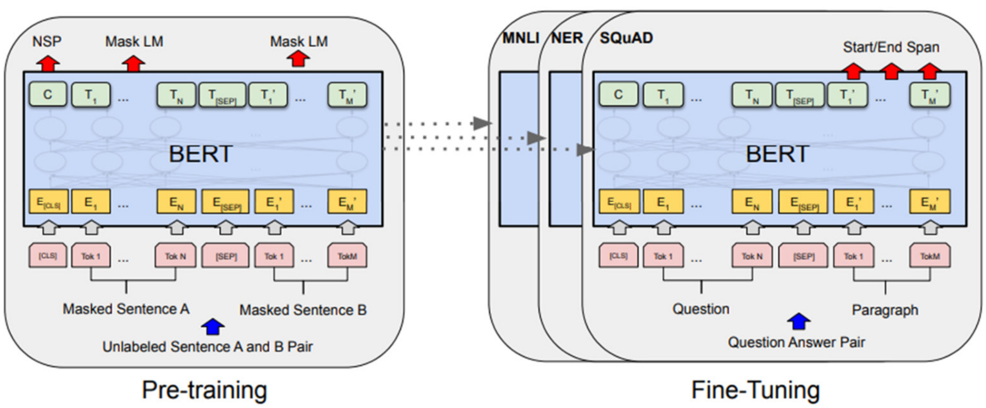

Figure 2: Architecture of BERT model [19]

BERT was fine-tuned for purpose-specific multi-label classification to be able to identify categories for resume improvement including missing_metrics, outdated_stack, and generic_summary. This model will facilitate suggested feedback.

2. T5 (Text-to-Text transfer Transformer):

T5 is a transformer-based model for natural language processing (NLP) developed by Google Research. T5 makes every text-based problem a text-to-text problem, which means it can translate, summarize, answer questions, and classify texts. The architecture of T5 allows the model enough flexibility to achieve reasonable performance across several NLP applications.

We used FLAN-T5 (i.e. fine-tuned T5) that generates interview questions based on the resumes we supplied in prompts. Specifically, the model takes structured prompting of certain sections of the candidates' resumes, and leverages these structured prompts, and their context, to generate specific, coherent interview questions.

3.5 Question Deduplication and Categorization

Generated questions are filtered with Sentence-BERT (SBERT) to remove semantically redundant content. Each question is then categorized as Technical, Behavioral, or General using keyword heuristics and semantic analysis.

3.6 Feedback and Resume Insights

Using the BERT classification and the extracted content, the system produces personalized resume improvement tips (e.g., include metrics, update tech stack), and suggestions to enhance clarity and relevance. Style issues (e.g., extensive paragraphs, generic verbs) are also flagged.

3.7 Model Evaluation

For evaluating the relevance of generated questions, cosine similarity is used to compare SBERT embeddings of the question to relevant resume content (e.g., those chosen as related by the candidate). There's also evaluation of section-wise coverage and diversity of generated questions, as well as a confusion matrix for classification tasks. This structured pipeline supports robustness, scalability, and personalization, providing users an all-in-one AI assistant for resume development and interview preparation.

4. RESULTS AND EVALUATION

To evaluate if this AI system works properly, we implemented a few empirical observational tests on a selected dataset of resumes. The AI systems responses were reviewed across three vectors: relevancy, diversity, and section coverage.

4.1 Experimental Setup

The system was evaluated using 50 resumes across different domains (software engineering, data science, system administration). The resumes were parsed and moved through the entire pipeline, including section detection, entity extraction and then question generation using our fine-tuned FLAN-T5 model. To create a consistent approach, the Sentence-BERT model (all-MiniLM-L6-v2) was used to evaluate the relevancy (via cosine similarity)and for question diversity or similarity examination.

The above figures illustrate the capabilities and flexibility of the resume based question generation system. In figure 3 Resume Section Coverage shows the percentage of resumes that included each section and how frequently the sections were used in the question generation. Education information was available in 100% of the resumes, followed by the Projects (95%), Skills (90%), and Certifications (85%), indicating based on potential content the system sometimes likely had a lot of good content to generate from. Less prominent evidenced sections, such as Experience (70%) and Summary (60%), were used when the section was offered, and the model likely evaluated and used the richest sections of content as many of the sections were blank. In figure 4 Relevance Score Distribution suggests a graph with histogram-style visualizations in the format of past and future representations is appropriate for the distribution of cosine similarity scores produced between the questions and the resume section from which the model utilized to create the questions. There is an appropriate mean of 0.84 with a tight standard deviation, indicating there was legitimate semantic relevance and suitable contextualization available between the questions and their offered content. This validates that the model was producing questions that were grammatically correct and, indeed, relevant in context. The Diversity Score across Resumes displayed in Figure 5 presents a boxplot of SBERT-based pairwise similarity scores from questions generated from the same resume. The median similarity is 0.76 with similarity scores from 0.73 to 0.80, which indicates a good diversity within the questions generated with minimal redundancy, showing that the system can keep a variety of questions while still being relevant to the text of each individual resume. Taken in its entirety, these images show that system can strike a good balance between ensuring the sections were used during in its generation, maintaining semantic relevance and diversity within question generation.

4.2 Overall Performance Summary

Table 2: System-Level Performance Overview

Dimension	Performance
Resume Section Coverage	94% (≥ 3 sections used)
Relevance Score	0.84 (Mean Cosine Similarity)
Question Count	6.1 (Average per resume)
Deduplication Accuracy	92% (Unique Questions)
Execution Time	10-15s per resume

The performance statistics show a good overview of the system performance. Resume Section Coverage is high, with 94% of the resumes having three or more sections, which translates to more significant variation, producing rich questions. The average Relevance Score is an impressive 0.84, implying the questions have a great deal of semantic distance from the resume content. The system is producing, on average, 6.1 questions per resume, which balances deeper questioning with quantity. There is also high Deduplication Accuracy (92%), which means that most questions are unique and not extracted from the same context, ensuring a variety of questions. The execution time has an average between 10 - 15 seconds per resume, allowing for variation depending on input length and complexity. This execution time is not as fast as real-time systems, however, the additional time is spent with processing that allows for deeper analysis and more relevant output, therefore this is an acceptable trade-off for an application that is focusing on quality.

4.3 Confusion Matrices for Resume Quality Issue Detection

To assess the success of our custom BERT model, we fine-tuned the model on a multi-label dataset of resumes marked with improvement categories: missing_metrics, missing_skills, outdated_stack, or generic_summary. Each label identifies a common structural or content weakness typically found in resumes.

These labels allow the model to provide improvement suggestions so that job seekers are able to hone in on a resume based on more specific, relevant and current information.

The confusion matrices in Figure 6 show the classification performance for each individual label. The model shows good discrimination power with very few false positives and false negatives across our categories.

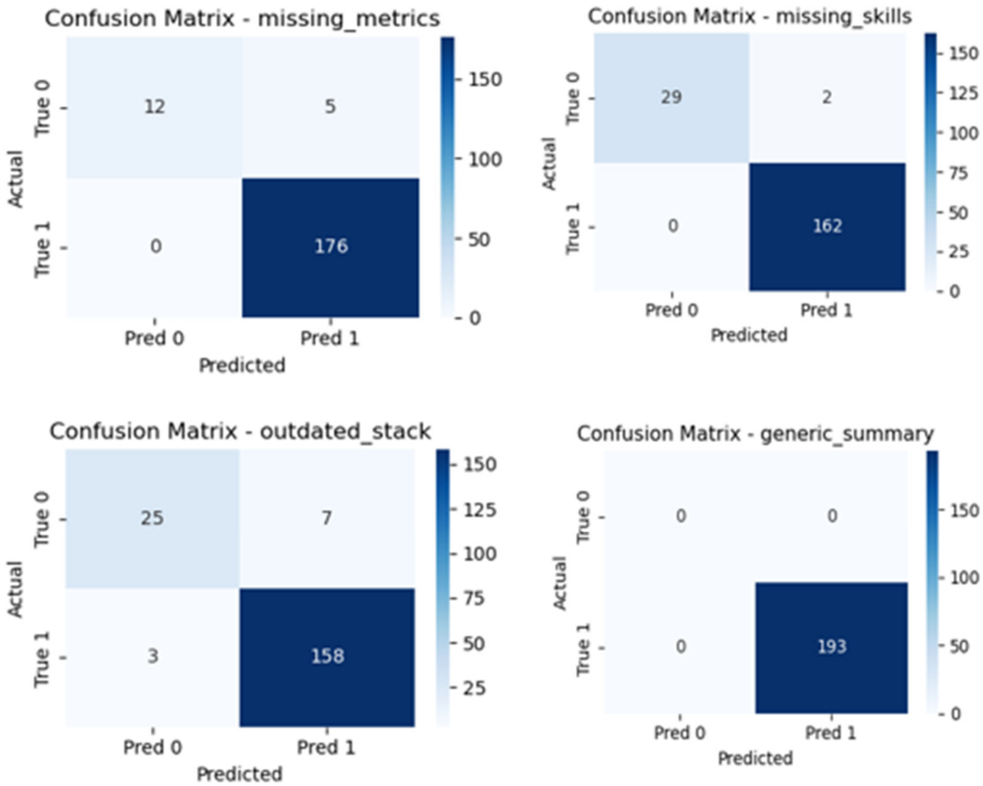

Figure 3 : Confusion Matrices for Each Improvement Label

1. Missing Metrics: Refers to resumes that reported accomplishments, but did not contain quantified achievements, or data-backed contributions (e.g. increased performance by 30%). The classifier was able to correctly identify 176 true positives with only 5 false positives which reiterates the robust sensitivity to identifying data-driven content.
2. Missing Skills: This category is applied to resumes that had either omitted core technical skills like Python, SQL, or Git, which would be considered core key skills of the requisite role. The classifier demonstrated excellent precision with 162 true positives and only 2 false positives, showing that it had some reliability in detecting abilities that were important in omission.
3. Outdated Stack: Detects that resumes are using an obsolete or deprecated technology or role. The classifier had 158 true positives, 7 false positives and 3 false negatives, which allude there was a bit more noise but good performance overall.
4. Generic Summary: Flags resumes that begin with generic statements that reflect weak points (no strengths) and provide no significant value proposition for the reader. This model had perfect classification with 193 true positives and labelled no errors, which most strongly suggests that the model had confidence and pattern in identifying generic summaries.

Given that these classifiers successfully identify quality issues, they offer a process to automate the process for resumes to be automatically thumbed down and provide the user with feedback regarding being flagged for evaluation with relatively no effort on the part of the manual model-trained evaluator.

Classification Report for Resume Improvement Categories

Table 3: Classification Performance Overview

Label	Precision	Recall	F1-Score
Missing Metrics	0.97	1.00	0.99
Missing Skills	0.99	1.00	0.99
Outdated Stack	0.96	0.98	0.97
Generic Summary	1.00	1.00	1.00
Micro Avg	0.98	1.00	0.99
Macro Avg	0.98	1.00	0.99

Table 3 demonstrates the classification performance of the fine-tuned BERT model in the four categories we identified to improve resumes. The model's performance is excellent. F1-scores ranged from 0.97 to 1.00. Labels like Missing Skills or Generic Summary, scored on perfect or nearly perfect scores, indicating the model was highly accurate. While Outdated Stack had slightly lower precision, it still boasted an F1-score of 0.97, a very high level of reliability. Furthermore, both micro and macro averages are 0.99, confirming the model was consistent across all labels and no less valid in identifying the types of issues impacting resume quality categorically.

5. CONCLUSION

In this study, we proposed an AI-based tool that provides support to job seekers by analyzing resumes and generating tailored interview questions. Our system integrates rule-based natural language processing (NLP), entity extraction and a fine-tuned transformer model such as FLAN-T5 for context-aware question generation for our users. Our system incorporates section-aware parsing to be able to facilitate question generation within the context of each section in the resume, SBERT-based relevance scoring and fallback mechanisms in the event that we produced output that is too similar or dissimilar to the user, to ensure we provide a set of questions that are various and relevant to the user's resume. We had reliably good performance with 94% section coverage, average relevance score of 0.84 and 92% question uniqueness. We included resume job-description matching in our system to personalize the performance to the user. Our current limitation is the time taken for question generation as the model inference time is sensitive to many parameters as well as quite slow. Future work will address enhancing performance and include features for ATS-based feedback and mock interviews, transforming it into a fully developed virtual career assistant.

REFERENCES

[1] Vivek, R. (2023). Enhancing diversity and reducing bias in recruitment through AI: A review of strategies and challenges. *Informatics, Economics, Management, 2(4),* 0101–0118. https://doi.org/10.47813/2782-5280-2023-2-4-0101-0118

[2] Daryani, Chirag & Chhabra, Gurneet & Patel, etal., (2020). An Automated Resume Screening System Using Natural Language Processing And Similarity. 99-103.

[3] Lee, B. C., & Kim, B. Y. (2021). Development of an AI-based interview system for remote hiring. *International Journal of Advanced Research in Engineering and Technology, 12(3),* 654–663. https://doi.org/10.34218/IJARET.12.3.2021.060

[4] Stuss, M. M., & Fularski, A. (2024). Ethical considerations of using artificial intelligence (AI) in recruitment processes. *Education of Economists and Managers, 71(1),* 53–67. https://doi.org/10.33119/EEIM.2024.71.4

[5] Gardazi, N. M., Daud, A., Malik, M. K., Bukhari, A., Alsahfi, T., & Alshemaimri, B. (2025). BERT applications in natural language processing: a review. *Artificial Intelligence Review*, *58*(6), 1-49.

[6] Rathore, S. P. S. (2023). The impact of AI on recruitment and selection processes: Analyzing the role of AI in automating and enhancing recruitment and selection procedures. *International Journal for Global Academic & Scientific Research, 2(2),* 78-93. https://doi.org/10.55938/ijgasr.v2i2.50

[7] Saatci, M., Kaya, R., & Ünlü, R. (2024). Resume Screening With Natural Language Processing (NLP). Alphanumeric Journal, 12(2), 121-140. https://doi.org/10.17093/alphanumeric.1536577

[8] Albaroudi, E., Mansouri, T., & Alameer, A. (2024). A comprehensive review of AI techniques for addressing algorithmic bias in job hiring. *Ai*, *5*(1), 383-404.

[9] Victoire, T. A., Vasuki, M., & Selvi, Y. S. (2024). DeepResume: Deep learning-based resume parsing for candidate screening. *International Journal of Current Science, 14(2),* 377–382.

[10] Qin, C., Zhu, H., Shen, D., Sun, Y., Yao, K., Wang, P., & Xiong, H. (2023). Automatic skill-oriented question generation and recommendation for intelligent job interviews. *ACM Transactions on Information Systems, 42(1),* 27. https://doi.org/10.1145/3604552

[11] Bhagia, Y., Abbas, S. M., Kumar, S., & Maheshwari, S. (2024). Conversational AI: A survey. *International Research Journal of Engineering and Technology, 11(4),* 833–842.

[12] Stuss, M. M., & Fularski, A. (2024). Ethical considerations of using artificial intelligence (AI) in recruitment processes. *Education of Economists and Managers, 71(1),* 53–67. https://doi.org/10.33119/EEIM.2024.71.4

[13] Priyanka, J. H., & Parveen, N. (2024). An efficient resume skill extraction using deep feature-based AGT optimized K-means clustering. *Multimedia Tools and Applications, 83*, 65967–65986. https://doi.org/10.1007/s11042-024-18220-0

[14] Albaroudi, E., Mansouri, T., & Alameer, A. (2024). A Comprehensive Review of AI Techniques for Addressing Algorithmic Bias in Job Hiring. *AI*, *5*(1), 383-404. https://doi.org/10.3390/ai5010019.

[15] Devaraju, S. (2022). Natural Language Processing (NLP) in AI-Driven Recruitment Systems. *IJSRCSEIT, DOI*, *10*.

[16] Bevara, R. V. K., Mannuru, N. R., Karedla, S. P., Lund, B., Xiao, T., Pasem, H., ... & Rupeshkumar, S. (2025). Resume2Vec: Transforming applicant tracking systems with intelligent resume embeddings for precise candidate matching. *Electronics*, *14*(4), 794.

[17] Vishaline, A. R., Kumar, R. K. P., VVNS, S. P., Vignesh, K. V. K., & Sudheesh, P. (2024, July). An ML-based Resume Screening and Ranking System. In *2024 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT)* (pp. 1-6).

[18] Lavi, D., Medentsiy, V., & Graus, D. (2021). consultantbert: Fine-tuned siamese sentence-bert for matching jobs and job seekers. *arXiv preprint arXiv:2109.06501*.

[19] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers) (pp. 4171-4186).