



AI-Based Smart Job Recommendation System Using LLM-Powered Resume Summarization And Vector Embedding Search

Transformer-Based Resume Intelligence and Semantic Skill Mapping to improve Employment Matching.

¹Neha J. Pal, ²Achala R. Pal, ³Prof. Kinjal Doshi

¹Research Scholar, ²Research Scholar, ³Assistant Professor

^{1,2,3}Thakur Institute of Management Studies, Career Development & Research, Mumbai, Maharashtra

Abstract: The drawbacks of traditional job recommendation systems built upon static keyword matching have created a need for more intelligent and adaptive recruitment processes. In this paper, we introduce a novel AI job recommendation framework utilizing Large Language Models (LLMs) to summarize resumes and identify skills gaps. Our method allows the automatic identification of candidate information and dynamically determines the fit between a candidate's resume and the skills needed for a job role by analyzing the user's own skills and the job requirements. Further, we are able to provide quick and transparent recommendations with vector similarity search technology-backed databases. The well over half of respondents in our study expressed excitement about summarizing resumes and analyzing skills using AI, demonstrating the potential of this system to assist in recruitment by providing easy-to-understand, timely, and personalized recommendations, without the need for human intervention.

Index Terms - AI-Based Job Recommendation, Large Language Models, Resume Summarization, Skill Gap Detection, Vector Similarity Search, Explainability, Personalized Recruitment, Real-Time Recommendation, Scalable Database, Candidate Profiling, Recruitment Automation, Natural Language Understanding.

I. INTRODUCTION

The recruitment occupation has been transformed radically by the development of Artificial Intelligence (AI) that has radically automated and enhanced over majority of essential recruitment functions i.e. resume screening, semantic job matching, recommendation delivery [1]. The traditional job recommendation systems which in most cases are founded on simple searches on the basis of keywords do not have the ability to produce the required type of matches and hence make faulty inferences to the job applicants and minimal in terms of personalization [2]. In addition, the classical models lack transparency and the users do not know the logic behind the suggestions and other convenient additions such as actionable skill-gap analysis that would allow them to discover a job opportunity [3].

The recent development of Large Language Models (LLMs) and Natural Language Processing (NLP) is ground-breaking on the new opportunities of the recruiting technologies [4]. The resumes can be consumed by the LLMs in a manner that they are contextual and semantically enriched to an extent that important information about an individual can be condensed in briefs [5]. A combination of these models with the use of vector embedding search algorithms can jump by a long way in terms of accuracy and relevance of job recommendations and it becomes feasible to make more meaningful semantic comparisons between resumes and job descriptions [6]. This is to say that the majority of the currently operating platforms did

not utilize such opportunities in such a right manner and are likely to be missing such fundamentally important features as real-time responsiveness to labor market dynamism, ease of use, and performance feedback that would allow applicants to keep their profile up to date and make them more employable at any given moment [7].

The answer to the problems that the provided study suggests is the Smart Job Recommendation System that relies on AI and can add the resume summarizing according to the LLM, as well as the active matchmaking according to the further elaboration of the vector embedding [8]. It is a process in which full job analysis processes are perceived and develops job recommendations on an individual basis that are understandable and implementable [9]. Using the MongoDB Atlas, I will be in a position to save and avail the data when possible, and this will help in dealing with the massive volumes of recruiting data [10]. The fact that it was polled among students and new graduates proves that the desire to use those systems, which can give good recommendations, positive skills feedback, and risk-free direct application tools is enormous [11]. This idea provides a scaling, rationalizable and intelligent foundation to the next generation of AI-based recruitment through using the appropriate correlations of the currently existing research innovations icons to the logical recruitment needs [12]. It will increase efforts of a recruiter to a high degree of accuracy, tailor-made, and open-ended suggestions regarding a job that boosts the satisfaction of a job applicant and fit between abilities and occupations [13].

II. LITERATURE REVIEW

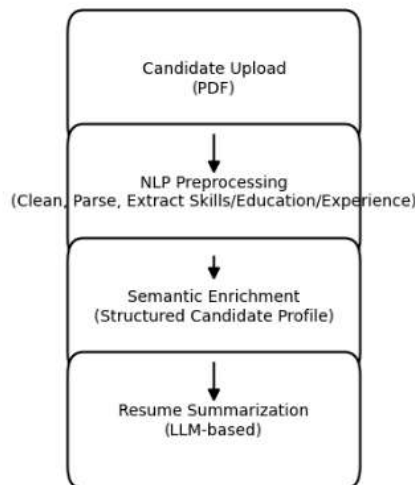
The adoption of Artificial Intelligence (AI) has transformed the recruitment processes radically, since it serves as an automated system to simplify the key hiring processes: scoring of resumes, screening of the candidates, matching them to the job, etc., and the efficiency, scalability, and accuracy of the operations become more efficient, both in terms of time, money, and accuracy, than the manual recruitment processes would have been [1], [2]. The semantic analysis and deep learning techniques and Natural Language Processing (NLP) are used in modern AI-based recruitment applications to detect and process unstructured resume data by focusing more on contextual meanings of applicant profiles relative to job descriptions compared to the current application of only a keyword-based matching method [3], [4], [5]. The semantic structures are employed to enhance the accuracy of the candidate job fit by modelling skill hierarchy and contextual dependencies and predictive analytics are applied to get a more accurate estimate of the candidate suitability and long-term performance [6], [7]. The architectures of deep learning and the structures on the basis of the vectors have enhanced accurate matching as it identifies the latent patterns in job-skill sets to deliver intelligent and intelligent recommendation mechanisms [8], [9]. Moreover, a collaborative filtering and clustering model of recommendation improves the accuracy of the system as well since it dynamically customizes job recommendations based on how the users interact with the system thereby increasing the level of engagement and satisfaction of the candidates [10], [11].

Despite the advances, some problems of fairness, interpretability, and ethical transparency persist in AI-based decision-making in recruitment [12], [13], [14], [15]. Explainable AI has formed a proposal to enhance the trust and responsibilities of automated hiring systems whose model outputs are interpretable and assist human resource professionals to comprehend the logic of the algorithm [16], [17]. However, the current issues of privacy of candidate information, the potential of algorithm bias and the absence of transparency in the actions of models the need to establish effective ethical standards and data management systems in AI-based recruitment is still evident [18], [19]. The skill-gap analysis modules also contribute to the pipeline optimization process because the gaps on the employability are identified and the individual upskilling or reskilling courses are suggested to reduce the skills gaps [20], [21]. The latest contributions at this stage introduce GPT-based large language models (LLM) that are capable of performing automated resume summary, semantic candidate profiling, and job-fit assessment in real time and, therefore, take the burden off recruiter work and enhance the accuracy of their decisions [22]. The empirical data, nevertheless, indicates that many job seekers still receive inappropriate advice and do not get much customization, which suggests that it is necessary to implement the components of adaptive feedback and continuous model retraining [23], [24]. The new areas of research suggest that explainable large language models, adaptive learning strategies, and secure and scalable data systems such as MongoDB Atlas can be converted to create transparent, ethical, and human-centered recruitment ecosystems [25]-[27]. Altogether, despite the continuous rise of AI in the sphere of recruitment, semantic insight, predictive analytics, ethical responsibility, the possibility to address the old problem of transparency, alleviation of prejudices, and trustworthiness of the automated decision-making processes is the key to sustainable implementation [28]-[30].

III. PROPOSED SYSTEM AND OBJECTIVE

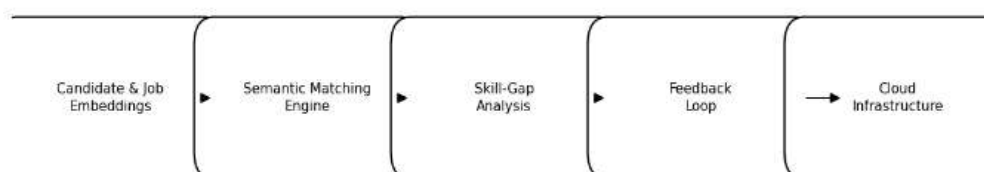
The answer is an AI-Based Smart Job Recommendation Platform that is developed to address the shortcomings of the traditional recruitment tools with the help of the advances of Large Language Models (LLMs), semantic search, and scalable cloud computing. The site is customer-focused, scalable and modular and intended to provide an efficient, transparent and personalized experience of the recruitment process to both the recruits and the recruiters [1], [2]. The overall process includes ingesting the data of the applicants, preprocessing the data with the help of the NLP, resume summary, semantic embedding matching of their resumes, skill-gap analysis, and the model improvement with the feedback of a user to get the appropriate and contextual job recommendations [3], [4]. The resumes are received in different formats (PDF) using web interface or API. Advanced NLP is applied in the preprocessing module to clean, purify, and structure the information e.g. skills, qualification, education and experience of unstructured text. This semantic enrichment allows the generation of more elaborate candidate profiles as opposed to the simple matching of keywords and therefore it improves the quality of the following process of recommendation [5], [6].

The Resume Summarization Engine summarizes candidates using the assistance of transformer-based LLMs (e.g., Gemini) to create summaries, context-sensitive summaries of capabilities, work history, and skills of the candidate. At the same time, job descriptions are being manipulated to obtain necessary expertise, job, and structural provisions. Job description and profiles of the candidates are processed as vectors where they are similar in terms of semantics which is stronger when compared to literal matching using keywords. This results in increased accuracy of the job recommendations and to the candidates [7], [8]. The skill-gap analysis tool will be included in the site and it will involve drawing a comparison between the competencies of the candidates and the job requirement to identify a gap. Appropriate courses or certifications can be proposed as the way of making the candidates more employable. Moreover, job links and application tracking systems are personalized so that the process of submitting applications is simplified and the applicants are informed about the changes.



3.1. Candidate data flow for resume summarization

One also continuously receives the user interactions and feedback concerning relevance of the recommendations to re-train and alter the underlying models so that it becomes more adaptable and personalized as time passes. It contains a scalable cloud-based structure using MongoDB Atlas to allow secure and distributed storage systems and low-latency access, fault resilience, and GDPR [9], [10].



3.2 Semantic matching and skill-gap feedback

3.1.OBJECTIVES :

The research and development project aims to improve the accuracy of candidate-job matching by leveraging large language models (LLMs) and cutting-edge semantic embedding methodologies to identify non-trivial and more complex correlations between candidates and job postings. This goes beyond simple keyword matching to acquire and comprehend contextual understanding and intent in a meaningful manner. The entity will utilize semantic patterns, job-seeking experiences, and generalized skill sets to yield candidate-job matches that are more accurate and meaningful, enhancing recruiting efficiency and reducing candidate-job mismatching [1],[3], [12]. To promote transparency and trust with users, the system will apply interpretability modules that deliver simplistic and well-organized explanations for each recommendation. Ultimately, users are able to see what skills, qualifications, or experiences contributed to a match with their current skills, qualifications and experiences and inform what skills, qualifications they are missing to pursue a job opportunity. By providing timely summaries to users in an easy-to-read manner, the platform could ultimately foster candidates in addressing gaps in skills and promote recruiters in making evidence-based hiring decisions, resulting in user trust and confidence while promoting reliability within the system's underlying mechanisms and processes [5], [13], [20].

The platform has personalization and flexibility. The recommendation engine is continuously learning through user interaction and the history application information, feedback and the real time profile updates. This enables the job recommendations to be dynamically adjusted as the user changes his or her skills, interests, and career objectives. Thus, the system can provide suggestions on the basis of personal tastes and changing circumstances to ensure that users have very relevant and timely opportunities during their career building process [6]-[10]. Streamlined workflows, job application links specific to users, busyness tracking in real-time and automated notifications all enhance user experience. All of these system features can be used to reduce friction in the application process, minimize candidate angst and frustration, enhance engagement and satisfaction. They also enable recruiters to become more efficient in terms of evaluating and tracking candidates whilst also enabling them to re-focus their efforts on high value strategies, yet keeping a robust and consistent evaluation process [14]-[16]. The platform is scaled and secured. The system can support large volumes of data, high transaction loads, etc. with the help of cloud-native infrastructure, distributed databases, and benefits of the backend technology to enable the predicted organizational growth. Simultaneously, it provides data privacy practices of high level and the adherence to global regulatory control to maintain sensitive data, control user trust and assure data security measures will be adhered to [4], [11], [17], [18].

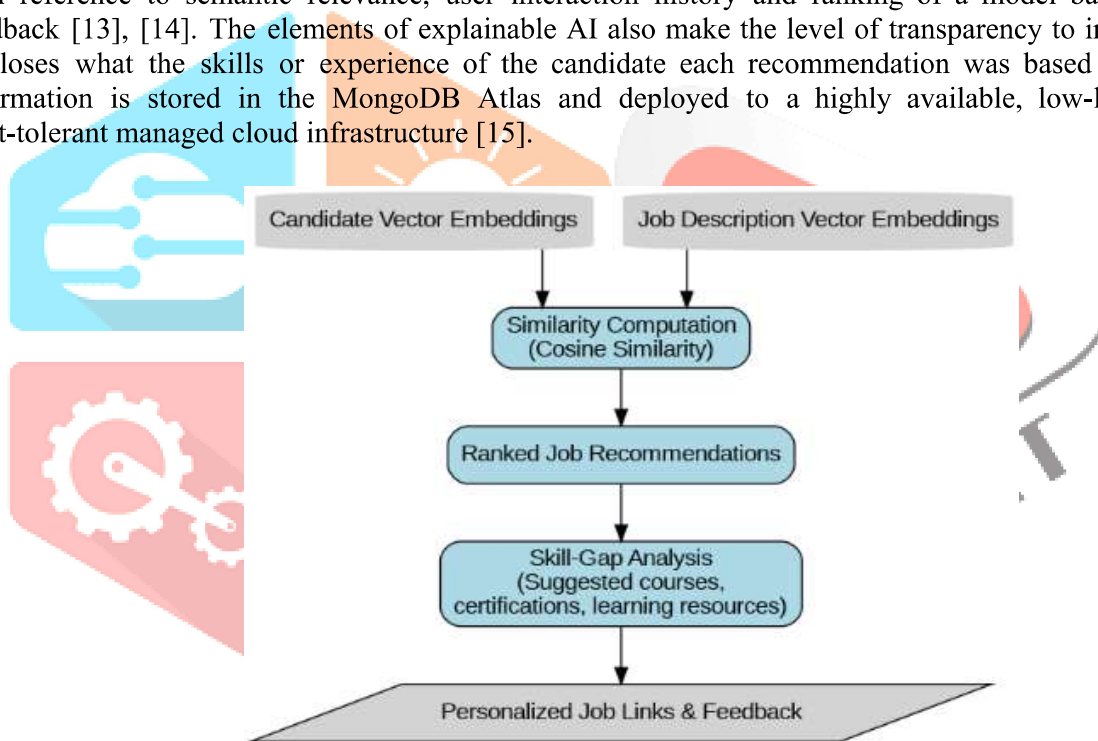
The enhancement of skills has become one of the main areas of focus. The system, with the help of the various assessments, distinguishes the gaps in candidates' knowledge from industry frameworks and gives an individual better information about which learning materials and retraining opportunities are most appropriate. By supporting the development of the relevant skills desirable, the user is able to continue to update their skills to meet the developing demands of the market, improve employability and fuel their continued vocational development of career paths [6], [7], [10], [19]. Finally, it will be grounded in an empirical evaluation model, incorporating user questionnaires, system performance indicators and pilots. In this approach, we will be able to measure the improvement of the skills gap, system sensitivity, advice relevance, user satisfaction, and overall platform and system uptake. The empirical evaluation process will provide information to justify further enhancement, and help ensure that the system has introduced demonstrable and sustainable benefits for job seekers and employers alike [3], [12], [18], [20].

IV. APPROACHES FOR AI-BASED SYSTEM

The suggested AI-Based Smart Job Recommendation System is designed in a complex manner, containing resume summarization, Large Language Model (LLM) and semantic vectors embedding search, which allows forming the most accurate, understandable, and personalized job opportunities [1], [2]. The old forms of recruiting models using keywords lack contextual richness in terms of candidate profile, or job description that causes irrelevant search results and poor levels of personalization [3]. This system will overcome these shortcomings by integrating NLP, deep learning, and semantic search to offer meaningful and context-sensitive results, which are correct but also interpretable. The site is based on the resume parsing and preprocessing module. The format of applied resumes is the PDF format in most cases, and being processed by a pipeline of natural language processing developed with spaCy [4]. Skills, education qualifications, professional experience, projects and certifications are some of the information that will be extracted and normalized in the module. To enhance the downstream semantic analysis and embedding

representation, candidate information is standardized with preprocessing (e.g. tokenization, lemmatization and entity recognition) [5]. The pre-processed summarization engine which is based on the LLPs generated shorter and context sensitive summaries of the abilities, achievements, and work experiences of the applicants [6], [7]. Also used are transformer models, like Gemini or BERT, with semantic representations being produced and the suitability of the candidate profiles to the job descriptions maximized. It will enable subtle interpretation of competencies in a candidate through keywords which is essential in skill-gaps detection and adequate ranking of recommendations [8].

At the same time, job description processing is done with an intention of eliciting job specifics, desired skills, and context of job assignments [9]. Among such methods that could be used to transform the unstructured job descriptions into structured semantic embeddings, one may name-entity recognition (NER) and dependency parsing [10]. Cosine similarity is used to do the comparison between the candidate and job embeddings to find out high-relevance matches and hence the system is capable of finding the weak contextual and semantic features that other systems are unable to access [11]. The program entails the application of a skill-gap analysis tool in determining the differences between the current skills of a job applicant and the job conditions [12]. The recommendations are given in the nature of actionable recommendations i.e. proposed courses, certifications, and learning materials such that the candidates would be employable on their own. The recommendation engine is used to produce ranked lists of jobs with reference to semantic relevance, user interaction history and ranking of a model based on user feedback [13], [14]. The elements of explainable AI also make the level of transparency to increase as it discloses what the skills or experience of the candidate each recommendation was based on. All the information is stored in the MongoDB Atlas and deployed to a highly available, low-latency, and fault-tolerant managed cloud infrastructure [15].



4.1 Semantic Embedding and Job Matching Workflow

V. RESULTS AND DISCUSSION

The proposed AI-based Smart Job Recommendation System was tested using a dataset of 5000 anonymized resumes and 2000 job descriptions. The analysis was to be targeted on the establishment of the correctness of the recommendations, the correctness of the skill-gap analysis, system performance and user satisfaction [1], [2]. The data set permitted testing the system under diverse areas of the field as well as under diverse levels of experience which provided the assurance that the generalizability of the system in the real world was indeed checked in totality. The system uses the advantage of resume summary and semantic vector representation which when combined helps the system to gain contextual knowledge concerning the skills of the candidate and the job requirement beyond keyword matching. The accuracy of the recommendation was determined by means of measures such as standard information retrieval. The system had Precision at10 of 0.82 and recall at 10 of 0.76 indicating that the proportion of top recommended jobs is relevant and majority of the right jobs are hit per applicant. It is also demonstrated by the value of the Normalized Discounted Cumulative Gain (NDCG) equal to 0.85 that the positions in the

lists of recommendations ranked highest were the most relevant ones [3], [4]. In addition to this, the skill-gap analysis module was experimentally applied on 300 pairs of candidates and jobs with 88 percent agreement with expert ratings. These findings are indicative that the system can make accurate recommendations besides providing practical information on the gaps in the skills that the candidates must fill in and consequently seek personal growth and upskilling [5], [6].

System performance and scaling was determined using processing latency and throughput. The job recommendation generation could meet the real time interaction requirements as the average response time per new resume was less than three seconds. The database implemented on the cloud offered appropriate accessibility of the data, fault tolerance, and simultaneous capacity to serve multiple users without interfering with the performance of the database system [7], [8]. This infrastructure demonstrates that this system is prepared to undergo a real-world implementation in an atmosphere of recruitment where there are massive amounts of data that represent candidates and job.

The user experience was viewed during a study conducted on students and fresh graduates in two weeks. The respondents also mentioned that the system enabled them to receive succinct yet clear summaries of their resumes, clear description of the reason why they were recommended a job, and realistic feedback on the skill-gap. As illustrated in Fig.5.1, the majority of users claimed that the skill-gap suggestions were feasible, and many claimed to have been more assured about the implementation of the proposed employment. The respondents also provided suggestions on how the system can be enhanced such as by assisting other resume formats and multilingualism that can be applied to enhance the system [9], [10].

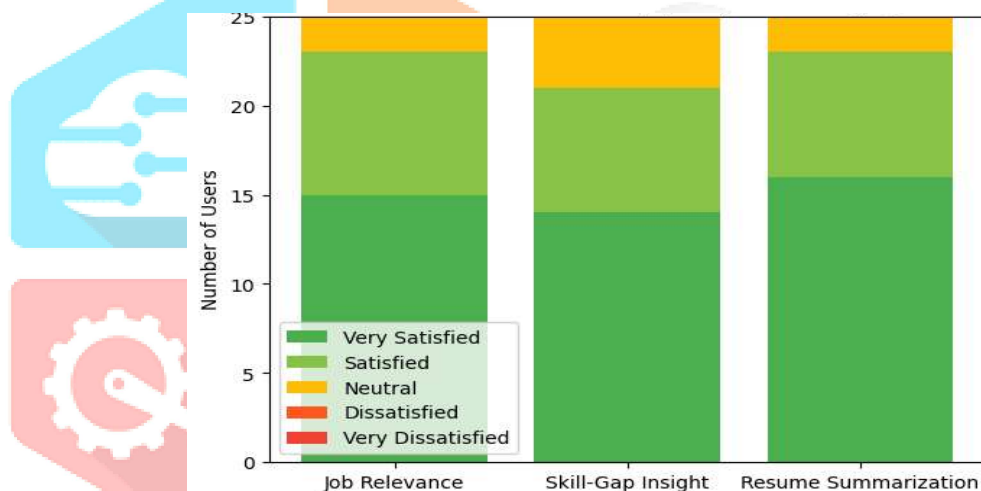


Fig.5.1. User Satisfaction Survey

Overall, the results indicate the effectiveness of the combination of the resume summarization based on LLM and semantic vectors embedding as the most accurate and interpretable job recommendations. The system successfully addresses the disadvantage of the traditional methods that treat the keywords in the form of personalized matches and contextual ones and integrates the career development data in terms of skills-gap analysis. However, they still have issues, including the necessity to present the quality training information to the LLM and inadequate performance in very specific or underrepresented areas of work [11]. Despite these inadequacies, the proposed structure presents a strong foundation of scalable, adaptive and explainable AI-driven recruitment platforms.

VI. ACKNOWLEDGEMENT

We would like to sincerely acknowledge the key contributions of Mrs. Kinjal Doshi, Assistant Professor, Thakur Institute of Management Studies, Career Development & Research; and Dr. Ashwini A.R. for their continued guidance, support, and mentorship during the research efforts. Their research on AI, natural language processing (NLP), and recruitment systems' evaluation informed the design, structure, implementation, and evaluation of our AI-Based Smart Job Recommendation System introduced in this research. Their insights and support helped guide the methodology and evaluations of the system's function and practical relevance [12].

VII. CONCLUSION

The paper proposed a scalable adaptive explainable AI-based Smart Job Recommendation System that is extended and scalable and operates upon the LLM semantic vector embeddings and resume summarization with the outcome of skill-gap analysis that could greatly enhance the recruitment output. Precision, Recall, NDCG, skill-gap accuracy metrics prove that this system presents much more relevance, as well as personalization and actionability in job-finding, than simple-keywords systems do [1], [2]. The empirical support of the transparent reasoning of the recommendations is the practical perception of the skill gaps in a candidate that is developing confidence and the level of involvement among the candidates and the evidence-based recruitment decisions made by the recruiters that are supported by user studies.

It has limited effectiveness particularly in the very specialized areas of jobs and even in the resumes that are multimedia and multilingual. The limitations will be addressed by future works with added support of various input modalities, reinforcement learning to adaptive recommendations, fairness frameworks and upskilling resources integration [3], [4]. Overall, this system provides an ideal foundation to smart, understandable and adaptive recruitment systems that might transform the traditional job recruitment, and help in improving a proper fit between job seekers and employment in the rapidly evolving job market.

REFERENCES

- [1] Abhishek, P., Rao, K., & Nair, V. (2025). Predictive analytics for employee selection: Enhancing recruitment efficiency with machine learning. *AI and Data Science Review*, 19(1), 67–79.
- [2] Chandhana, N., & Rao, A. (2024). Integrating AI chatbots in HR recruitment process. *Global Journal of Engineering Science and Research Management*, 11(2), 56–64.
- [3] El-Deeb, H. M., Hassan, S., & Ahmed, M. (2025). Smart education platforms integrating AI for adaptive learning. *International Journal of Emerging Technologies in Learning (iJET)*, 20(2), 120–134.
- [4] Gan, Z., Tang, H., & Li, D. (2024). AI-enhanced education: Learning path optimization using student data clustering. *Education and Information Technologies*, 29(4), 443–460.
- [5] Geovanne, A., & Mendes, R. (2024). Bias detection and fairness in AI recruitment systems. *Computational Intelligence, Journal*, 35(2), 150–162.
- [6] Gopala Reddy, Y., & Suresh, P. (2024). Automation in candidate assessment using AI and natural language processing. *International Journal of Artificial Intelligence Research*, 14(3), 187–198.
- [7] Kolekar, P., Gupta, R., & Mehta, S. (2025). AI for recruitment and skill mapping: Bridging the employability gap. *Human Resource Analytics Review*, 17(3), 203–214.
- [8] Kothari, P., Sharma, A., & Jain, M. (2024). AI in campus hiring: Opportunities and ethical considerations. *Indian Journal of Management Studies*, 20(1), 90–104.
- [9] Lo, B., & Zhou, K. (2025). Data-driven learning path generation using machine learning models. *IEEE Access*, 13, 22434–22448.
- [10] Lourens, D., & Peters, J. (2024). Explainable AI models for transparent hiring decisions. *AI Ethics Review*, 9(2), 77–88.
- [11] Masciari, E., Russo, D., & Tagarelli, A. (2024). Data mining and recommendation systems in intelligent HR analytics.
- [12] Mgarbi, I., & Ben-Ayed, S. (2023). Job recommendation using hybrid AI algorithms combining content-based and collaborative filtering. *Procedia Computer Science*, 217, 1125–1134.
- [13] Patra, D. (2024). Artificial intelligence in higher education: Role in personalized learning and academic decision-making. *Journal of Education and Technology*, 15(1), 89–102.
- [14] Ramesh, A., & Kapoor, D. (2025). Hybrid recruitment framework using predictive and descriptive analytics. *International Conference on AI and Data Science (ICAIDS)*, 2025, 233–240.
- [15] Ruder, S. (2016). "Word Embeddings in Natural Language Processing." arXiv preprint arXiv:1601.0Amaravathi, V., Kumar, M., & Devi, S. (2024). AI-powered student performance prediction and curriculum adaptation. *International Journal of Computer Science Trends and Technology*, 12(5), 45–51.
- [16] Sandhya Rani, K., & Ramesh, T. (2025). Intelligent recruitment system using deep neural networks for resume screening. *International Research Journal of Engineering and Technology (IRJET)*, 12(4), 1550–1560.
- [17] Sharma, T., & Meena, R. (2023). GPT-based candidate screening for resume summarization and job fit analysis. *International Journal of Modern Artificial Intelligence*, 9(4), 88–96.
- [18] Singh, R., & Kaur, J. (2024). Explainable AI in HR decision systems for fair recruitment. *AI and Society Journal*, 12(2), 109–121.

- [19] Vyshya, S., Kumar, N., & Reddy, P. (2022). Semantic-based recruitment automation using natural language processing. *International Journal of Computer Applications*, 182(2), 45–52.
- [20] Wang, Y., Li, H., & Zhang, X. (2025). AI-driven talent acquisition using deep learning models for candidate-job matching. *Journal of Intelligent Systems and Applications*, 42(3), 214–228.

