Skill Gap Detection and Job Fitment via Selection Insight and Rejection Reasoning

Suresh Kumar S
Professor of Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai, India
sureshkumar.s@rajalakshmi.edu.in

Dhivya Bhavani V
Assistant Professor of Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai, India
dhivyabhavani.v@rajalakshmi.edu.in

Priya Darshini S
Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai, India
221801040@rajalakshmi.edu.in

Sandhya J
Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai, India
221801044@rajalakshmi.edu.in

Tanushri G V S
Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai, India
221801055@rajalakshmi.edu.in

***Abstract* - While navigating the competitive job market, matching candidate skills with requirements of the job is a major challenge for both the candidates and the recruiters. A system conceptualized for automating resume screening, identification of the skill gap, candidate-job fitment evaluation, and data-driven explanation of rejection is deployed for employability development and placement readiness. The proposed system integrates NLP and ML processes into four primary modules: Skill Gap Detection, Candidate Fitment Analysis, Resume Rejection Reasoning, and an AI Interviewer. The Skill Gap Detection module utilizes TF-IDF vectorization and visualization processes to match extracted resume skills against job advertisements. The Candidate Fitment component employs Cosine Similarity and Random Forest classification for recommending suitable job positions. The Resume Rejection Reasoning subcomponent analyzes comparisons of accepted and rejected resumes to find potential reasons for rejection. Finally, the AI Interviewer subcomponent generates skill-specific and role-based interview questions for customized preparation. The model can match academic learning to the industry demands because it offers institutions, recruiters, and student organizations an actionable advice.**

Keywords - Natural Language Processing , Machine Learning , Skill Gap Detection, Candidate Fitment, Rejection Reasoning, AI Interviewer, Career Readiness.

I. INTRODUCTION

In the digital age, employability is no longer solely based on academic credentials but also on what can be shown in the way of job-related competencies that keep pace with fast-changing industry requirements. With so many online learning resources and programs available, students and job-seekers are still uncertain about the particular competences that keep them out of those jobs they want. Meanwhile, the hiring managers are hindered from properly processing and evaluating high volumes of resumes with irregular or delayed hiring decisions. This alignment problem between candidate capabilities and organizational demands is a source of inefficiencies for the recruitment ecosystem. Latest technologies have been the driving forces behind automating and optimizing candidate screening. Using these, the information from the resumes can be transformed into useful insights that provide information about matching skills, identifying knowledge gaps, and to suit for required job positions. Machine Learning models further improve this with predictive features, thus making data-driven recruitment decisions for recruiters and candidates easier. This paper introduces an AI-driven framework including NLP and ML for resume evaluation, skill gaps detection, and candidate-job match analysis automatically. Unlike current systems, the new system provides explainable explanations for rejection and has an AI Interviewer module to aid in personal interview practice. The system ultimately focuses to improve employability, decrease recruiter workload, and bridge the gap between education readiness and work readiness.

II. LITERATURE SURVEY

A. REVIEW OF AI-BASED SKILL GAP DETECTION AND JOB FITMENT SYSTEMS

Siswipraptini et al. [1] examine the evolving IT professions and skills needs dynamics using text-mining and clustering analysis. Driven by the projected 15 % increase in IT job openings worldwide and the persistent mismatch between educational output and industry requirements, the article proposes a scientific semantic model for IT job profiling. With datasets gathered from two prominent Asian job boards—Tech in Asia and Job Street Indonesia—the authors examine 1,065 job postings via text preprocessing, term-frequency–inverse-document-frequency weighting, and Average-Linkage Hierarchical Clustering (ALHC) to extract repeated phrases, programming languages, tools, and frameworks constituting IT competencies. The ALHC algorithm allows a hierarchical tree of connected skills and disciplines, successfully unearthing inter-cluster relationships and multi-granular job groupings. The findings identify ten significant IT job clusters, namely web development, software testing, database administration, and information security, each defined by prevailing technical skills like HTML/CSS, JavaScript, SQL, and Python. Industrial relevance is validated by Focus Group Discussions with ten industry experts. In addition to descriptive profiling, the research provides an empirical semantic pipeline for curriculum alignment and workforce analytics enabling educational institutions to realign course structures according to empirical skill demands. In general, this study closes the gap between academic planning and labor-market analytics by integrating text-mining, clustering, and expert validation into a well-integrated model of data-driven IT job profiles.

Ashrafi et al. [2] address massive-scale workforce disruption due the emergence of automation by a novel AI-based re-education and career-recommendation system called Career-gAIde. The article represents the urgent need for rapid skill realignment with the improvement of smart and virtual offices taking the place of traditional jobs. The system is based on Neural Networks (CNN-Random architecture) and is used to automatically match resumes and job descriptions, identify skill gaps, and suggest personalized learning routes to higher-paying, better-matched jobs. Unlike earlier job-matching systems, Career-gAIde not only suggests right-fitting job opportunities but also approximates the user's current salary, forecasts possible future career growth, and suggests aimed re-education through personalized learning resources. The system handles web-crawled job information and resume text through DBpedia-based concept extraction, TF-IDF weighting, and normalization of salary in terms of multiple currencies and experience. Tested on precision and recall values, the system attains 67 % precision and more than 80 % recall in job-offer and skill-recommendation accuracy. The research offers empirical support that smart recommender systems are capable of effectively aiding workforce reskilling, providing scalable, automated support to workers in negotiating insecure post-pandemic job markets. By combining machine learning, NLP, and human-resource analytics, the paper offers a building block model for AI-based career guidance and lifelong learning implementable across various occupational domains.

He et al. [3] introduce an approach, which promotes the person–job matching accuracy by capturing rich interactions among multi-field features effectively. Conventional person–job fit models tend to flatten heterogeneous data sources and compromise the fine-grained dependencies between candidate features and job demands. To bridge this constraint, authors introduce a novel model called MUFFIN (MUlti-Field Features representation and INteraction) that accounts for intra-field and inter-field interactions via self-attention mechanisms. The model initially projects heterogeneous data—categorical, numerical, and textual data from resumes and job postings—into a common latent space leveraging contextual embedding methods like ALBERT for textual inputs. It then presents two important interaction modules, which discovers dependencies between different fields. Dynamic weighting of each feature's importance is achieved, allowing the system to prioritize the most significant candidate–job relationships. Lastly, a multilayer perceptron (MLP) sums these interactions to generate a compatibility score predicting the effectiveness of person–job matching. Experiments on large-scale real-world recruitment datasets show that MUFFIN outperforms conventional baselines and other deep models substantially across evaluation metrics like AUC and F1-score. Overall, this piece of work establishes a robust, end-to-end smart recruitment architecture through the combination of self-attention, learning of feature interaction, and deep contextual representation to effectively represent complex human–job compatibility relationships.

Vukadin et al. [4] presented a multilingual NLP workflow, which is responsible for retrieving data automatically from resumes in multiple languages and in various formats. Acknowledging the inherent variability of processing free-text CVs in different languages and structures, the authors proposed a twin-model design rooted in the transformer architecture and Bidirectional Encoder Representations from Transformers. Their approach enables accurate recognition of high-level document sections—e.g., personal information, education, work history, and qualifications—and finer-grained elements such as names, degrees, job titles, and companies. BERT's multilingual encoder is the back-bone model that provides cross-linguistic generalizability over five languages like English, Polish etc. A second model is dedicated to determining the self-reported competency levels of individual skills. The situation introduces a large-scale annotated corpus of 1,686 CVs, one of the largest multilingual corpora available for this task. Performance measurement with precision, recall, and F1 metrics reflects strong performances with macro-averaged F1-scores of more than 0.82 on all languages. Notably, the built-in attention mechanism enhances interpretability through visualization of what the model is paying attention to while classifying relevant text components. Besides, the paper also explores other optimization dimensions such as reducing the depth of BERT layers to facilitate speed and accuracy trade-off. With an integration of transformer-based multilingual processing, explainability AI best practices, and robust evaluation, this paper constructs a successful, interpretable, and language-invariant framework for smart resume parsing—leverages automation in recruitment analytics and multilingual information retrieval.

Marinai, et al. [5] address a detailed overview of the use of Artificial Neural Networks to Document Image Analysis and Recognition. The paper points out the development of neural networks from being used for character recognition only to being applied to handle wider document processing tasks. The authors opine that while conventional OCR systems operate effectively in the processing of text, ANNs are more universal in processing noisy, complicated, and unstructured documents. Multilayer perceptrons (MLPs) are found to be the most common architecture used in methods like noise elimination, binarization, and classification in the paper. Besides, the book deals with more complex neural models such as Self-Organizing Maps (SOMs) and Recursive Neural Networks (RNNs) applied to hierarchical pattern recognition and conventional data representation. One of the book's key contributions is that the authors put a special emphasis on modular and hybrid neural systems with features to incorporate generative and discriminative learning in an attempt to boost recognition accuracy and robustness to outlying data. Close reading reveals that the paper names feature representation issues, data variability, and computational performance as the problematic ones, and graph representations and convolutional networks as the sought-after directions. Experimental work is seen to demonstrate dramatic advances in segmentation and OCR performance with the implementation of ANN-based systems. Overall, this paper is an introduction linking machine learning with analyzing documents in terms of accompanying in the development of intelligent, adaptive systems efficient of handling complex documents across different languages and formats.

B. RESEARCH GAPS AND NEED FOR THE STUDY

Existing body of research on resume classification, skill extraction, and job match prediction has largely advanced using the advances in latest technologies. Various studies had also discussed automated resume parsing and classification by leveraging BERT-based multilingual models, CNN–BiGRU ensemble systems [6], and transformer encoders to extract structured information from unstructured resumes. Although these methods are better at parsing accuracy and matching resumes to jobs, they chiefly focus on syntactic extraction of text and not holistic assessment of employability skills. Thus, semantic interpretation of candidate profiles, behavioral competencies, and contextual reasons for hiring decisions is still lacking.

Likewise, models like DistilBERT and XLM [9] have enhanced resume ranking and shortlisting by better utilizing similarity scores between the job description and candidate resume. But these systems focus mostly on surface-level matching, without indicating why some candidates are rejected and what competencies are absent. Research such as FairHire [11] and MUFFIN [3] have started incorporating LLMs for unbiased resume assessment and multi-field feature representation, but they fall short of offering interpretability in rejection explanations and directed feedback to users. This deficiency creates a gap between computerized screening systems and human judgment in making hiring decisions.

In the domain of skill gap analysis and career guidance, frameworks such as SGAM [12] and Career-gAIde [2] have leveraged adaptive learning and recommendatory solutions. These models, however, are primarily static—they connect present skills but do not often deliver real-time, one-on-one learning or interview preparation feedback. What's more, current AI-powered mock interview platforms leveraging NLP, CV, and reinforcement learning tackle interactive testing but do not directly relate candidate answers to their resume-derived skill profile. Therefore, existing studies lack to bring together resume analytics, skill gap analysis, and interview simulation under a single employability platform.

Document analysis and OCR-based models have maximized data extraction from semi-structured and structured resumes with CNNs, GTNs, and rule-based approaches but are myopically devoted to preprocessing alone and not career-fit prediction. Further, research on learning path generation—while effective in educational environments—is yet to be integrated with employability analytics for the purpose of advising students towards employment readiness.

Therefore, there is a huge research gap in creating a holistic, explainable, and AI-based platform that integrates resume analysis, skill gap identification, rejection explanation, and AI-powered interview preparation. The system proposed addresses this by utilizing NLP, ML, and AI conversational components to deliver end-to-end assistance—from identifying gaps in skills to creating customized interview questions and improvement feedback. This blending closes the gap between computer-aided assessment and man-like inference, and the project as such is an essential part of intelligent employability solutions.

III. RELATED WORK

The recent advancements had transformed recruitment analytics, resume categorization, and prediction of job fitment beyond one's imagination. Resume screening through automation has emerged as a valuable tool for contemporary companies that strive to be efficient and impartial in hiring. Conventional manual screening is usually erratic, prejudiced, and incapable of processing massive volume of applicant data. Consequently, several AI-based frameworks have been put forward to increase the transparency and interpretability of assessment systems.

Early work in data extraction from resumes introduced multilingual BERT-based architectures for obtaining structured details from unstructured resumes and laid the groundwork for resume parsing automation. Similarly, Automated Resume Classification Using Ensemble Learning [6] designed a CNN–BiGRU hybrid model that made use of pre-trained embeddings for resume text classification. Though these systems were more accurate in data extraction, they generally consisted of syntactic feature matching rather than semantic understanding of employability context.

Later developments like Job Applications Selection and Identification with NLP and ML [7] examined OCR, NER, and TF-IDF features based on SVM, RF, and KNN classifiers [8] to enhance parsing accuracy and classification. Ranking of resumes with advancements like XLM [9] employed cosine similarity and Euclidean distance measures in candidate ranking with respect to job descriptions. Whereas these methods improved candidate filtering, they did not remediate underlying analytical reasoning for rejections or provide actionable feedback for skill development.

In parallel, Candidate Recommendation Chatbot with WordNet-based Answer Evaluation [10] and FairHire: Bias-Free AI Interview and Resume Analysis Platform [11] brought conversational and fairness-focused AI to hiring. FairHire applied LLMs in particular for adaptive Q&A and real-time scoring to counter recruitment bias. But these systems were only able to evaluate and did not have integrated modules for generating feedback and developing skills. Likewise, MUFFIN an innovation [3] introduced a sophisticated model incorporating textual, categorical, and numerical features for job fit prediction, but it did not offer explainable rejection reasoning or skill suggestions.

The Skill Gap Analysis Model (SGAM) [12] proposed the Design Science Research Methodology (DSRM) for skill gap mapping across job roles and staff, which further led to early quantification of employability skill deficiencies. Progressing from this, Career-gAIde [2] for re – learning had laid down an AI-enabled recommendation framework for learning pathways and higher-tiered employment transitions through CNN–Random NN pipelines. However, both models worked on static data sets and did not have dynamic adaptability to real-time resume databases or interactive user interfaces.

Artificial intelligence-based interview systems like Personal Placement Assistant (PPA) with RAG and LLMs [13] and AI-Driven Mock Interview System with NLP, CV, and RL [14] symbolized multimodal evaluation's next level, with the combination of voice and visual indicators for adaptive feedback. AIced – Prep : RAG + LLM-Powered Interview Assistant [15] applied retrieval-augmented generation using FAISS retrievers and Mistral LLM for question generation, emphasizing the increasing presence of LLMs in interview preparation at an individual level. However, these models were interview-focused and unattached to resume analysis or gap detection, with the pre-interview evaluation issue remaining unsolved.

In the domains of job recommendation, Automated Job Recommendation System for Fresh Graduates [16] utilized demographic and background profiling using ML algorithms to match candidates to appropriate positions, while IT Job Profiles Using Hierarchical Clustering [1] used Agglomerative Linkage and text mining for clustering analogous job profiles. While successful in data clustering, these research studies lacked the personalized measures of evaluation and rejection justification models that were specific to the needs of a company. Resume preprocessing and OCR-work dependent like Gradient-Based Learning for Document Recognition [18] and Artificial Neural Networks for Document Analysis [5] enhanced data extraction and segmentation accuracy with the aid of CNNs and MLPs. Likewise, Rule-Based Semi-Automated OCR Postprocessing for Complex Documents supported data cleaning in the context of multi-column formats. The above contributions were also restricted to preprocessing and not employability evaluation.

Parallel learning research like LD-LP: Learning Diagnosis–Learning Path Method [17] and make learning adaptive according to users with personal recommendation [19] applied reinforcement learning and knowledge tracing to recommend personalized learning pathways. Although effective in learning contexts, they have not been applied in employability networks to bridge skill gaps within real-time recruitment scenarios.

IV. PROPOSED SYSTEM

System Overview

Informed by the understanding of the contemporary literature on automated recruitment, skill analytics, and employability prediction, the model contemplated in this paper will be a database-driven integrated system to screen resumes automatically, identify skill gaps, identify required candidate-job matching, and tailor interview preparedness.

The system architecture designed is theoretically divided into four interconnected layers: Skill Gap Detection, Candidate Fitment Analysis, Resume Rejection Reasoning, and AI Interviewer Module. Each layer provides a unique analytical insight into the overall employability evaluation process. The detection layer of Skill Gap uses text-mining, keyword extraction, and vectorization methods such as TF-IDF, Word2Vec, and Cosine Similarity to identify matches between resume skills from candidates and needed skills in target job descriptions. This is what enables the evaluations to be made by the model on missing and under-represented skills required for different job roles.

The Candidate Fitment Analysis layer accomplishes this using Supervised Learning algorithms like Random Forest Classifier, SVM, and Logistic Regression. These measures essentially consider the extent to which an applicant's profile is predictive of job success with a specific job group. This phase monetizes employability in score-based or probabilistic terms and generates an objective employability score for recruiters and candidates.

Resume Rejection Explanation module offers explainability through the use of explainable AI (XAI) methods It looks back at previously rejected resumes and makes inferences about likely reasons for previous disqualification from previous employment trends, individual needs of the company, and successful profile requirements. The explanation process not only provides candidates with constructive feedback but also enables companies to make their selection filters more sophisticated.

Lastly, AI Interviewer module employs NLP-based question generation and semantic analysis techniques to formulate skill-specific, role-specific, and adaptive questions for interview. By its mimicking of actual interviews, it gets the candidates ready and does self-evaluation and gives the recruiters a standardized yet developing pre-screening instrument for recruits.

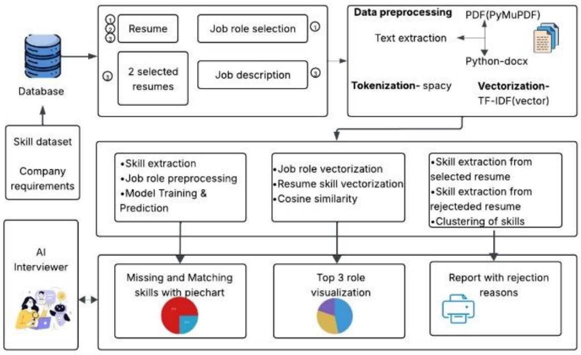

Fig. 1. System Architecture

V. CONCLUSION

Discovery of the gaps between educational skills and industry needs is a top priority in the automated hiring environment. Recent literature shows that most models focus on parsing resumes and candidate-job matching, with fewer integrating the elements that provide explainability, feedback, and recommended improvement paths for the rejected candidates. A conceptual framework - consisting of the tiers of Skill Gap Detection, Candidate Fitment Analysis, Resume Rejection Reasoning, and AI Interviewer - provides an organized approach to resolving these shortcomings.

Through integration of NLP-driven extraction, ML-powered matching with job role profiles, and reasoning modules that map rejection reasons to company requirements, such a model could improve transparency and enable candidates with decision-enabling insights. The AI Interviewer layer to generate role-specific interview questions from a candidate's skill profile further increases preparedness and self-assurance in interview situations.

Having such an integrated system has the capability to yield dramatic improvements in employability forecasting, candidate screening efficacy, and alignment of educational outcomes and selection procedures. Future research opportunities lie in empirically testing the framework within various job markets, creating cross-domain datasets, and incorporating explainable AI methods for understanding model outputs. Scalable deployments, ongoing learning, and deployment in production will enable intelligent recruitment systems to reach their maximum potential as more equitable and effective servants to both employers and candidates.

VI. REFERENCES AND RESOURCES

[1] IEEE Access, *"Information Technology Job Profile Using Average-Linkage Hierarchical Clustering Analysis,"* 2023.

[2] IEEE Access, *"Efficient Resume-Based Re-Education for Career Recommendation in Rapidly Evolving Job Markets,"* 2023.

[3] IEEE Transactions on Computational Social Systems, *"Self-Attentional Multi-Field Features Representation and Interaction Learning for Person–Job Fit,"* Feb. 2023.

[4] IEEE Access, *"Information Extraction From Free-Form CV Documents in Multiple Languages,"* 2021.

[5] IEEE Transactions on Pattern Analysis and Machine

Intelligence, *"Artificial Neural Networks for Document Analysis and Recognition,"* Jan. 2005.

[6] ICACCS Conference, *"Automated Resume Classification System Using Ensemble Learning,"* 2023.

[7] IEEE SCEECS Conference, *"Job Applications Selection and Identification: Study of Resumes with Natural Language Processing and Machine Learning,"* 2023.

[8] ICAEECI Conference, *"A Rule-Based Semi-Automated OCR Postprocessing Method for Aligning Multi-Language Transcripts with Multi-Column Text,"* 2023.

[9] IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE), *"Resume Ranking and Shortlisting with DistilBERT and XLM,"* 2024.

[10] International Research Conference on Smart Computing and Systems Engineering (SCSE), *"Candidate Recruitment Based on Automatic Answer Evaluation Using WordNet,"* 2019.

[11] ICTEST Conference, *"FairHire: AI-Driven Resume Profiling and Technical Interview Automation,"* 2025.

[12] IEEE Engineering Management Review, *"Using the DSRM to Develop a Skills Gaps Analysis Model,"* Dec. 2020.

[13] ICDICI Conference *"A RAG Based Personal Placement Assistant System Using Large Language Models for Customized Interview Preparation,"* 2024.

[14] ICSSAS Conference, *"AI-Powered Mock Interview Platform Using Computer Vision, Natural Language Processing and Generative AI,"* 2025.

[15] ICDICI Conference, *"AIced-Prep – AI Based Mock Interview Evaluator,"* 2025.

[16] CISTI Conference, *"Designing Learning Paths with Open Educational Resources: An Investigation in Model-Driven Engineering,"* 2023.

[17] IEEE Transactions on Learning Technologies, *"LD–LP Generation of Personalized Learning Path Based on Learning Diagnosis,"* Feb. 2021.

[18] Proceedings of the IEEE, *"Gradient-Based Learning Applied to Document Recognition,"* Nov. 1998.

[19] IEEE Access, *"Smart E-Learning Framework for Personalized Adaptive Learning and Sequential Path Recommendations Using Reinforcement Learning,"* 2023.

[20] ICCMC Conference, *"Machine Learning Based Ideal Job Role Fit and Career Recommendation System,"* 2023.

[21] International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN), *"Data-Driven Evaluation of Graduate Job Suitability via Scalable Machine-Learning Architectures,"* 2025.