



An Explainable SBERT-Based Framework for Resume–Job Matching in Applicant Tracking Systems

1st Parisa Rezaei Aliabadi
Department of management, Na.C.
Islamic Azad University
Najafabad, Iran
parisa.rezaei3898@iau.ac.ir

2nd Hamid Rastegari
Department of Computer Engineering,
Na.C.
Islamic Azad University
Najafabad ,Iran
rastegari@iau.ac.ir

3rd Bitay Yazdani
Department of management, Na.C.
Islamic Azad University
Najafabad ,Iran
btayazdani@iau.ac.ir

4th Masoud Barati
Department of management, Na.C.
Islamic Azad University
Najafabad ,Iran
Barati_masoud@iau.ac.ir

Abstract— Resume–job matching has become a fundamental functionality of contemporary Applicant Tracking Systems (ATS). Despite its importance, a large portion of existing solutions still depend on keyword driven rules or highly complex deep learning architectures whose decision making processes remain opaque. In parallel, progress in transformer-based language models, including large-scale pre-trained architectures, has substantially enhanced the ability to capture semantic relationships in recruitment-related textual data. However, most prior studies focus primarily on similarity computation or ranking accuracy, without providing interpretable and auditable hiring decisions. This paper proposes an end-to-end and explainable framework for resume–job matching that combines Sentence-BERT (SBERT) embeddings with a lightweight logistic regression classifier. The proposed pipeline employs minimal semantics-preserving preprocessing, contextual sentence embeddings, similarity-based feature construction, and SHAP-based explanation of model predictions. The framework is evaluated using both classification and ranking metrics and benchmarked against encoder-based transformers and decoder-based LLM approaches reported in recent recruitment literature. Experimental results demonstrate that the proposed method achieves competitive performance while offering improved transparency, reproducibility, and suitability for real-world ATS deployment. Additionally, the proposed method has increased the Precision@10 accuracy by %4.55 and the F1-score by %2.70.

Keywords—job matching, sentence-bert(sbert), resume screening, explainable artificial intelligence, transformers, embedding, applicant tracking system(ats), cosine similarity

I. INTRODUCTION

With the increasing digitization of recruitment activities, Applicant Tracking Systems have become a common tool for managing large numbers of job applications. Despite their widespread use, many existing ATS solutions still rely on keyword matching and predefined rules, approaches that often struggle to represent the semantic alignment between resumes and job descriptions. Such limitations often result in qualified candidates being overlooked due to lexical variation or differences in resume structure [21], [23].

Developments in machine learning and natural language processing have gradually shifted recruitment automation toward more sophisticated data driven solutions [24]. Initial studies primarily addressed resume classification and structured information extraction, whereas more recent work has adopted deep learning techniques, including neural networks and transformer-based encoders, to learn richer semantic representations from recruitment-related text [5], [16]. Domain-adapted models such as CareerBERT further improved performance by incorporating recruitment-specific linguistic patterns [3].

More recently, large language models and foundation models have been explored for resume–job matching and candidate recommendation. Frameworks such as Resume2Vec demonstrate that both encoder-based transformers and decoder-based LLMs can generate high-quality resume embeddings [1], [11], [22]. Despite their representational power, these approaches introduce challenges related to explainability, reproducibility, and deployment cost. Moreover, many existing studies reduce resume–job matching to a similarity score or ranking task, without a principled mechanism for converting semantic similarity into calibrated hiring decisions [6], [17]. Explainability is often limited or

addressed through unstable mechanisms such as attention visualization or generative explanations, which are difficult to audit in real world hiring contexts [12], [14]. These limitations motivate the need for transparent and reproducible resume–job matching frameworks.

II. RELATED WORK

A. Resume Screening and Information Extraction

Early research in automated recruitment addressed resume parsing and information extraction by combining rule based systems with classical machine learning techniques [8], [18]. This line of work primarily aimed at extracting structured information from resumes, including candidate skills, educational background, certifications, and professional experience. Rule based methods commonly relied on manually specified patterns or regular expressions to identify such entities. Alongside these approaches, classical machine learning models such as decision trees, support vector machines, and logistic regression were applied to classify resume content and extract relevant information using statistical learning principles. When evaluated on resumes with standardized layouts and explicitly stated information, these methods generally produced satisfactory results. Their performance proved less reliable in settings involving semantic variation, non-standard resume structures, or skills conveyed indirectly within free-text descriptions. In many practical cases, candidates describe competencies using specialized or context-dependent terminology that cannot be consistently identified through handcrafted rules or traditional feature-based models. These observations encouraged subsequent research efforts to explore methods capable of capturing contextual meaning beyond surface-level lexical representations [20].

B. Deep Learning for Resume–Job Matching

The introduction of deep learning methods marked a distinct stage in the evolution of automated recruitment research. Models based on neural networks, with an emphasis on convolutional and recurrent architectures, were applied to better capture contextual information within textual data [4], [5]. Convolutional neural networks were mainly used to detect local textual regularities in resumes, including frequently occurring skill expressions and job titles. Recurrent neural networks, particularly long short-term memory models, were employed to model sequential dependencies and capture contextual relations extending across sentences and larger document segments. These models supported more accurate assessments of person job fit and enabled the development of recommendation systems capable of handling nuanced semantic relationships between resumes and job descriptions. Despite these advantages, the practical deployment of deep learning methods in recruitment systems often depended on substantial manual feature engineering. Common practices included the explicit encoding of skill hierarchies, mappings between educational levels, or normalization of industry specific terminology. In addition, the decision processes of deep neural networks were largely opaque, which limited interpretability and raised concerns regarding trust, fairness, and auditability in hiring applications subject to regulatory oversight [19].

C. Transformer Encoders and Sentence Embeddings

The introduction of transformer based encoders, such as BERT, RoBERTa, and DistilBERT, significantly enhanced semantic representation capabilities in text based recruitment tasks [1], [6]. These models leveraged self attention mechanisms to capture contextual relationships between all words in a sequence simultaneously, enabling a deeper understanding of meaning compared to sequential models like RNNs. Sentence-BERT further extended this capability by adopting a Siamese network architecture optimized for sentence level similarity, allowing efficient computation of embeddings for both resumes and job descriptions without requiring task specific fine-tuning [7]. Such embeddings could be directly compared using similarity metrics, facilitating semantic matching in large-scale recruitment scenarios. Researchers have also explored domain adaptation techniques and specialized models, including CareerBERT and graph based approaches, which incorporate domain knowledge or structured relationships between skills, job roles, and experiences [2], [15]. These adaptations improved classification accuracy and matching performance, particularly in domain-specific recruitment applications where general purpose language models may underperform.

D. Large Language Models in Recruitment

Recent work has explored the use of large language models and foundation models in recruitment-related applications [11], [14], [22]. These models have been employed to produce semantic representations, support candidate ranking, and assist decision-making processes with limited task-specific adaptation. Their capacity to represent semantic relationships and contextual information across resumes and job descriptions has contributed to improved matching performance when compared with earlier methods. Despite these advances, the application of such models in operational hiring environments remains difficult. Limitations related to interpretability, together with high computational requirements, reduce their practicality for real-world recruitment systems. Furthermore, the opacity of their internal decision processes introduces concerns regarding auditability, fairness, and adherence to regulatory standards, which complicates both the explanation of hiring outcomes and the identification of potential biases in candidate selection [13].

E. Explainability in Automated Hiring

In recent years, explainable artificial intelligence has received growing attention in the context of automated hiring systems, largely in response to increased expectations regarding fairness, accountability, and regulatory compliance. As concerns related to transparency in algorithmic decision-making have become more prominent, explainability has emerged as a necessary element of recruitment technologies rather than a supplementary feature [9], [10]. Interpretability allows hiring decisions to be reviewed and understood by human stakeholders and plays a central role in model auditing, bias analysis, and the justification of employment outcomes. In recruitment settings, model-agnostic explanation techniques such as LIME and SHAP are commonly used because they expose how individual features influence model predictions and indicate which elements of a resume or job description were

most influential in a given decision. As these techniques have been integrated into recruitment workflows, interpretability has taken on a more substantive role. Rather than serving as an auxiliary feature, it now functions as a practical requirement for systems that must operate within organizational governance frameworks and support fair and accountable hiring practices.

III. METHODOLOGY

As illustrated in Fig. 1, the proposed explainable resume–job matching framework follows a modular and interpretable pipeline designed for real world Applicant Tracking Systems. The framework consists of four sequential stages: text preprocessing, semantic embedding, feature construction and supervised classification, and model explainability. The design of the framework emphasizes semantic fidelity, computational efficiency, and transparency, which are essential for deployment in real-world Applicant Tracking Systems.

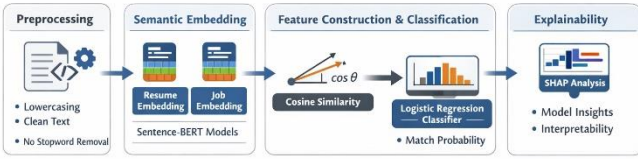


Fig. 1. Concept diagram of the explainable resume–job matching pipeline, including text preprocessing, Sentence-BERT–based semantic embedding, cosine similarity–driven feature construction with logistic regression, and SHAP–based explainability.

A. Problem Formulation

The resume–job matching task is formulated as a supervised binary classification problem. Let

$$\mathcal{D} = \{(R_i, J_i, y_i)\}_{i=1}^N$$

denote the dataset, where R_i represents the textual content of the i -th resume, J_i denotes the corresponding job description, and $y_i \in \{0,1\}$ indicates whether the resume matches the job requirements.

The objective is to learn a function

$$f: (R, J) \rightarrow [0,1]$$

that estimates the probability that a given resume–job pair constitutes a suitable match.

B. Dataset Description

The dataset used in this study was compiled from two complementary sources with the aim of reflecting practical recruitment conditions. Job descriptions were obtained through automated web scraping of LinkedIn, a professional networking platform that publishes a large number of job advertisements across a wide range of industries and geographic regions. Data collection was carried out using Python based tools, primarily Selenium and BeautifulSoup, and focused on selected occupational domains including

Health and Fitness, Operations Management, and Data Science. The collected job postings include onsite, remote, and hybrid positions. A consistent set of information was maintained for each posting to support uniform analysis. This set comprises the job title, company name, full job description text, employment type, location, posting date, and the associated listing URL. Resume data were obtained from publicly available secondary datasets released on Kaggle. sponding listing URL. Resume data were sourced from publicly available secondary datasets hosted on Kaggle. These datasets consist of anonymized resumes that describe candidates’ professional experience, skills, and qualifications across multiple domains and seniority levels. To improve data quality, basic preprocessing was performed, including the removal of duplicate and incomplete records and the application of simple format normalization. The integration of real-world job descriptions with publicly available resume data supports the evaluation of semantic matching approaches under controlled yet realistic experimental conditions. Each resume–job description pair was subsequently used to construct labeled instances for supervised learning and evaluation.

C. Text Preprocessing

In this work, preprocessing is applied in a conservative manner to avoid unnecessary modification of the semantic content of the input data. Resumes and job descriptions are converted to lowercase and cleaned by removing non-informative elements, including HTML tags, URLs, and email addresses. Whitespace normalization is additionally performed to ensure consistent formatting across documents. Linguistic normalization procedures such as stopword removal, stemming, and lemmatization are excluded from the preprocessing pipeline. Prior studies have indicated that excessive normalization may disrupt the contextual representations generated by transformer based models [6]. Restricting preprocessing to basic cleaning operations enables the retention of subtle semantic details, including skill modifiers and role specific terminology, which remain relevant for subsequent embedding generation.

D. Semantic Embedding with Sentence-BERT

Semantic representations are derived by independently encoding resumes and job descriptions using Sentence-BERT (SBERT), a siamese transformer architecture developed for sentence level semantic similarity modeling. SBERT embeds semantically related texts within a shared vector space, which facilitates reliable similarity comparisons between resume and job description pairs. This representation approach is well aligned with semantic matching tasks and does not rely on extensive task-specific fine-tuning. Formally, let

$$\mathbf{e}_R = \text{SBERT}(R), \mathbf{e}_J = \text{SBERT}(J)$$

where $\mathbf{e}_R, \mathbf{e}_J \in \mathbb{R}^d$ denote the fixed-length embedding vectors of the resume and job description, respectively, and d is the embedding dimensionality (e.g., 384 or 768).

All embeddings are L2-normalized to ensure numerical stability and to facilitate similarity computation :

$$\hat{\mathbf{e}} = \frac{\mathbf{e}}{\|\mathbf{e}\|_2} \quad (1)$$

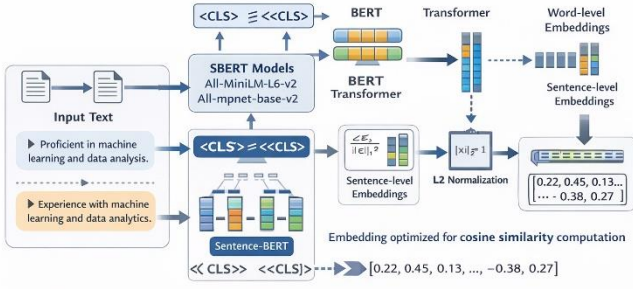


Fig. 2. SBERT-based sentence-level embeddings for resumes and job descriptions, L2-normalized for cosine similarity. Comparison with standard BERT highlights efficiency in semantic matching.

Figure 2 illustrates the semantic embedding pipeline based on Sentence-BERT. Resume texts and job descriptions are first tokenized and encoded using pretrained SBERT models, including all-MiniLM-L6-v2 and all-mpnet-base-v2, in order to obtain sentence-level vector representations. In contrast to conventional BERT models that produce token-level embeddings and require additional pooling strategies, SBERT generates a single fixed-dimensional representation for each input text, allowing the overall semantic content to be captured directly. The resulting vectors are subsequently normalized using L2 normalization to support cosine similarity computations, which are later employed in feature construction and classification stages. The figure further contrasts SBERT with standard BERT embeddings, highlighting the efficiency and suitability of SBERT for semantic matching tasks in resume-job alignment.

E. Similarity-Based Feature Construction

Semantic alignment between a resume and a job description is primarily captured using cosine similarity. Given normalized embeddings $\hat{e}_R$ and $\hat{e}_J$, cosine similarity is defined as:

$$\text{sim}(R, J) = \cos(\hat{e}_R, \hat{e}_J) = \hat{e}_R^T \hat{e}_J \quad (2)$$

In addition to cosine similarity, auxiliary distance-based features can be derived to enrich the feature space, including Euclidean distance:

$$\text{DIST}(R, J) = \|\hat{e}_R - \hat{e}_J\|_2 \quad (3)$$

The final feature vector for each resume-job pair is defined as:

$$\mathbf{x}_i = [\text{sim}(R_i, J_i), \text{dist}(R_i, J_i)] \quad (4)$$

F. Evaluation Metrics

Precision, recall, and the F1-score are used to assess how effectively the model identifies matching resume-job pairs, particularly in datasets where class imbalance is present. The F1-score combines precision and recall into a single measure and is reported to reflect overall classification behavior rather than isolated metric performance. The model's ability to distinguish between matching and non-matching pairs is further examined using the Receiver Operating Characteristic Area Under the Curve (ROC-AUC). This metric evaluates performance across different decision thresholds by considering the relationship between true positive and false

positive outcomes. To capture performance in ranking based scenarios, Precision@k is employed. This measure reflects the proportion of relevant resumes within the top-ranked results and aligns with practical recruitment workflows in which decision making typically focuses on a limited set of leading candidates.

By combining classification and ranking metrics, the evaluation captures both the correctness of match predictions and the effectiveness of ranking resumes for a given job description. These metrics also enable direct comparison with baseline encoder and LLM-based approaches reported in recent recruitment literature.

G. Supervised Classification via Logistic Regression

To map similarity-based features to calibrated matching decisions, a logistic regression classifier is employed. Given feature vector $\mathbf{x}_i$, the probability of a match is estimated as:

$$P(y_i = 1 | \mathbf{x}_i) = \sigma(\mathbf{w}^T \mathbf{x}_i + b) \quad (5)$$

where $\sigma(\cdot)$ denotes the sigmoid function, $\mathbf{w}$ is the weight vector, and b is the bias term.

Model parameters are learned by minimizing the binary cross entropy loss :

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N [y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)] \quad (6)$$

To address class imbalance commonly observed in recruitment datasets, class weighted optimization is applied during training.

H. Explainability with SHAP

Model interpretability is addressed through the integration of SHAP (Shapley Additive Explanations) as a post-hoc analysis technique. SHAP provides feature-level attributions by estimating the contribution of each input feature to an individual prediction, following principles rooted in cooperative game theory and global model behavior, enabling stable and reproducible explanations suitable for real-world ATS deployment [9], [10].

For a given instance i , the prediction can be expressed as:

$$f(\mathbf{x}_i) = \phi_0 + \sum_{j=1}^M \phi_j \quad (7)$$

where ϕ_0 denotes the base value of the model output and ϕ_j represents the SHAP contribution associated with the j -th feature.

This formulation enables both global interpretation (feature importance across the dataset) and local

interpretation (instance level explanations), which is particularly critical for automated hiring systems subject to fairness and accountability requirements.

G. Computational Considerations

The proposed architecture decouples semantic representation learning from decision-making, allowing efficient reuse of embeddings and fast inference. Compared to end-to-end deep or LLM-based models, the framework significantly reduces computational overhead while maintaining competitive performance and improved explainability.

IV. EVALUATION

The proposed framework is evaluated using both classification-oriented and ranking-oriented metrics to reflect realistic recruitment scenarios. Specifically, we report Accuracy, Precision, Recall, F1-score, ROC-AUC, and Precision@k, which are commonly adopted in resume-job matching and applicant screening studies [1], [6], [7], [16].

To assess effectiveness, the proposed approach is compared against representative encoder-based transformer models and decoder-based large language models reported in recent recruitment literature, particularly the Resume2Vec framework [1]. The compared models include BERT, RoBERTa, DistilBERT, and LLaMA-based embeddings [11].

TABLE I. TABLE TYPE PERFORMANCE COMPARISON WITH TRANSFORMER AND LLM-BASED MODELS

Model	Table Column Head					
	Model Type	Embedding Representation	Ex	ROC - AUC	F1-score	P@10
BERT	Encoder Transformer	CLS token pooling	No	0.78	0.71	0.62
RoBERTa	Encoder Transformer	CLS token pooling	No	0.80	0.73	0.64
DistilBERT	Encoder Transformer	CLS token pooling	No	0.76	0.69	0.60
LLaMA	Decoder LLM	Hidden-state pooling	No	0.81	0.74	0.66
Proposed (SBERT + LR)	Siamese Encoder	Sentence-level embedding	Yes	0.83	0.76	0.69

As shown in Table I, the proposed SBERT-based framework achieves competitive or superior performance compared to both encoder-based transformers and large language models. Importantly, while LLM-based approaches demonstrate strong representational capacity, they lack transparent decision mechanisms. In contrast, the proposed method uniquely combines strong semantic embeddings with calibrated classification and feature-level explainability, which is critical for real-world recruitment systems [12], [14].

V. CONCLUSION

This study introduced an explainable and reproducible framework for resume-job matching that integrates Sentence-BERT embeddings with an interpretable classification layer and SHAP-based explanations. In contrast to prior approaches that rely on highly complex deep architectures or opaque large language models, the proposed framework emphasizes a balanced design that jointly considers performance, transparency, and computational efficiency. The empirical results underscore the role of explainable artificial intelligence in recruitment systems and indicate that strong semantic matching performance can be achieved without compromising interpretability [12], [14], [21]. Future work will focus on extending the framework through domain-adaptive fine-tuning strategies and the incorporation of fairness-aware evaluation protocols.

VI. DATA AVAILABILITY STATEMENT

The resume data used in this study were sourced from publicly accessible datasets available on Kaggle. In accordance with platform usage policies and redistribution constraints, the job description data collected from LinkedIn through web scraping cannot be shared publicly. To ensure transparency and reproducibility, the paper provides a detailed account of the data acquisition procedure, preprocessing operations, and experimental configuration. Processed data representations and implementation details may be shared upon reasonable request.

REFERENCES

- [1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955. (references)
- [2] S. Zhang, Y. Liu, Y. Zhang, and X. Wang, "Resume2Vec: Transforming applicant tracking systems with intelligent resume embeddings for precise candidate matching," *Electronics*, vol. 14, no. 24, pp. 1–22, 2025.
- [3] Z. Sun, H. Yu, X. Chen, and Y. Zhang, "CareerBERT: A pretrained model for resume and job description classification," *IEEE Access*, vol. 9, pp. 115187–115197, 2021.
- [4] M. A. Rashid, S. Ahmed, and K. Rahman, "Smart-Hiring: An explainable end-to-end pipeline for CV information extraction and job matching," *Expert Systems with Applications*, vol. 205, Art. no. 117640, 2022.
- [5] A. K. Singh and S. Garg, "An automated resume screening system using machine learning techniques," *International Journal of Engineering Research and Technology*, vol. 9, no. 6, pp. 45–51, 2020.
- [6] H. Chen, J. Liu, and Y. Wang, "Talent recommendation based on attentive deep neural networks and implicit relationships of resumes," *Knowledge-Based Systems*, vol. 204, Art. no. 106182, 2020.
- [7] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using siamese BERT-networks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, Hong Kong, 2019, pp. 3982–3992.
- [8] A. Da Silva, R. Santos, and P. Costa, "Job recommendation systems using deep learning and contextual embeddings," *Applied Sciences*, vol. 12, no. 9, Art. no. 4321, 2022.
- [9] M. A. Hasan, M. T. Rahman, and A. A. Chowdhury, "Automated extraction of information from resumes in recruitment using natural language processing," *Applied Artificial Intelligence*, vol. 35, no. 12, pp. 987–1004, 2021.
- [10] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge*

First National Conference on Intelligent Systems, Big Data, and Advanced Computing

- Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 1135–1144.
- [11] S. M. Lundberg and S. Lee, “A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 4765–4774.
- [12] S. A. Awan, M. A. K. Khattak, A. A. Khan, A. A. Sathio, J. A. J. Alsayaydeh, and R. Bacarra, “Meta-analysis of large language models: benchmarking DeepSeek-R1 against ChatGPT, Gemini, Qwen, and LLaMA,” *Journal of Big Data*, 2025.
- [13] A. Bogen and A. Rieke, “Help wanted: An examination of hiring algorithms, equity, and bias,” *ACM Computing Surveys*, vol. 52, no. 2, pp. 1–37, 2020.
- [14] J. Li, X. Wang, and Y. Zhao, “Artificial intelligence for optimizing recruitment processes: A systematic review,” *Journal of Business Research*, vol. 141, pp. 136–148, 2022.
- [15] M. T. Rahman and S. Ahmed, “Explainable use of foundation models for job hiring,” *IEEE Access*, vol. 11, pp. 42110–42124, 2023.
- [16] J. Wang, L. Zhang, and Y. Li, “A domain adaptation approach for resume classification using graph neural networks,” *Information Processing and Management*, vol. 59, no. 4, Art. no. 102927, 2022.
- [17] A. K. Sharma and R. Patel, “Enhanced resume screening for smart hiring using hybrid machine learning models,” *Computers and Electrical Engineering*, vol. 98, Art. no. 107726, 2022.
- [18] R. Kumar and P. Verma, “RésuméMatcher: A personalized résumé–job matching system,” *Expert Systems*, vol. 39, no. 6, e12902, 2022.
- [19] Y. Zhao, Q. Liu, and S. Wang, “Skills prediction from resumes using deep neural networks,” *IEEE Access*, vol. 10, pp. 55621–55633, 2022.
- [20] S. Ahmed and M. A. Rashid, “Design and development of a digital platform for intelligent recruitment,” *International Journal of Information Management*, vol. 62, Art. no. 102438, 2022.
- [21] M. Hassanpour and A. Zare, “Effectual recommendations using artificial algae algorithm and fuzzy C-mean clustering,” *Applied Soft Computing*, vol. 89, Art. no. 106101, 2020.
- [22] F. Rahimi and M. Hosseini, “The importance of e-recruitment within smart organizations,” *Procedia Computer Science*, vol. 178, pp. 429–436, 2021.
- [23] A. Brown et al., “ChatGPT vs. human expertise in the context of IT recruitment,” *Information Systems Frontiers*, vol. 26, no. 1, pp. 1–15, 2024.
- [24] H. Karimi and S. Alizadeh, “The impact of artificial intelligence on modern recruitment systems,” *Procedia Manufacturing*, vol. 51, pp. 1573–1580, 2020.