TalentBridge AI: An Explainable Semantic and Generative Intelligence Framework for Recruitment Optimization and Skill Readiness Assessment

Viraj Khanvilkar
Dept. of Computer Engineering
Atharva College of Engineering
Mumbai, INDIA
e-mail: khanvilkarviraj-cmpn@atharvacoe.ac.in

Jai Punde
Dept. of Computer Engineering,
Atharva College of Engineering
Mumbai, INDIA
e-mail: jaipunde74@gmail.com

Abhishek Paradkar
Dept. of Computer Engineering
Atharva College of Engineering
Mumbai, INDIA
e-mail: aparadkar315@gmail.com

Prof. Mahendra Patil
Dept. of Computer Engineering,
Atharva College of Engineering
Mumbai, INDIA
e-mail: mahendrapatil@atharvacoe.ac

Abstract: The exponential digitization of recruitment workflows has introduced unprecedented scale into candidate screening pipelines, rendering manual evaluation both computationally infeasible and operationally inefficient. While Applicant Tracking Systems have become ubiquitous in mitigating screening overhead, most contemporary implementations rely on lexical keyword matching and Boolean rule

These techniques are inherently brittle to semantic variation, resulting in disproportionately high false negative rates when syntactic mismatch occurs between candidate resumes and job descriptions despite substantive alignment in skills and experience. This paper presents **TalentBridge AI**, a hybrid recruitment intelligence framework that reconceptualizes automated hiring as a semantic similarity estimation and readiness optimization problem. The proposed architecture integrates transformer based dense vector embeddings for high resolution semantic alignment with generative artificial intelligence for structured interpretability and skill gap remediation. Candidate resumes and job descriptions are embedded into a shared latent semantic space, enabling continuous similarity estimation via cosine similarity metrics. In parallel, a generative reasoning module constructs dependency aware skill graphs that transform ranking outcomes into personalized upskilling roadmaps. Extensive empirical evaluation conducted in a controlled pilot deployment demonstrates a 91.3 percent accuracy in role classification, a 94 percent structured parsing success rate, and a 40 percent reduction in end to end screening latency compared to baseline ATS systems. Beyond ranking candidates, the system operationalizes a reject and upskill paradigm, repositioning
recruitment systems from passive exclusionary filters into adaptive human capital optimization platforms.

Keywords: Semantic Retrieval, Transformer Embeddings, Generative Artificial Intelligence, Explainable Recruitment Systems, Skill Dependency Graphs, Human Capital Optimization

❖

INTRODUCTION

The contemporary recruitment landscape is characterized by large scale data ingestion, asymmetric information flow, and constrained human evaluation capacity. Organizations routinely process hundreds or thousands of resumes for each open role, necessitating algorithmic mediation to maintain throughput and operational viability [4]. Applicant Tracking Systems have emerged as the dominant solution to this challenge, yet their underlying methodologies remain largely unchanged from early information retrieval paradigms. Traditional ATS platforms primarily rely on Boolean keyword matching, regular expressions, and rigid rule based heuristics to triage candidates. While such approaches are computationally efficient, they fail to capture semantic equivalence, contextual relevance, and skill transferability. As a result, candidates whose resumes do not lexically mirror job descriptions are frequently penalized despite possessing functionally equivalent expertise. This phenomenon, widely referred to as the semantic gap, introduces systemic bias and inefficiency into automated hiring pipelines.

From the candidate perspective, automated recruitment systems operate as opaque decision engines. Applicants receive binary accept or reject signals devoid of explanatory context or actionable guidance. This opacity inhibits targeted upskilling, exacerbates labor market mismatch, and erodes trust in algorithmic hiring systems.

TalentBridge AI is introduced to address these limitations through a hybrid artificial intelligence architecture that combines discriminative semantic retrieval with generative interpretability. By framing recruitment as a semantic similarity and readiness assessment task, the system quantifies contextual alignment between candidate profiles and job requirements while simultaneously translating these evaluations into structured skill remediation pathways. This dual objective enables TalentBridge AI to function not only as a screening mechanism but also as an active agent in workforce development.

METHODOLOGY

A. Resume Parsing and Text Normalization

Resumes are ingested in heterogeneous raw document formats, including PDF and DOCX, and processed using layout preserving extraction pipelines to maintain structural fidelity. Unlike naïve text scraping approaches, the extraction process preserves hierarchical document semantics such as section boundaries, bullet structures, and entity grouping, thereby minimizing information loss during transformation from formatted documents to machine readable text.

The extracted textual corpus undergoes a multi stage normalization pipeline grounded in statistical natural language preprocessing theory. This pipeline includes lexical tokenization, case normalization, unicode canonicalization, punctuation filtering, and stop word suppression. Domain specific noise patterns such as decorative formatting artifacts, embedded symbols, and duplicated headers are algorithmically pruned to enhance semantic signal clarity.

Structural consolidation is performed to collapse fragmented semantic segments into coherent contextual blocks, ensuring that semantically contiguous content remains proximal in the representation space. This step reduces lexical sparsity and mitigates distributional drift across heterogeneous resume templates. By stabilizing token distributions prior to embedding generation, the preprocessing stage enhances downstream vector representation robustness and improves embedding consistency across diverse document layouts.

B. Semantic Vectorization and Similarity Computation

Following normalization, both resume text and job description text are transformed into fixed dimensional dense embeddings using a transformer based sentence encoder architecture [1]. This embedding mechanism is rooted in distributional semantics [3], wherein contextual meaning is encoded through attention driven self representation learning. The encoder maps textual sequences into a high dimensional latent vector space [2] where geometric proximity corresponds to semantic similarity.

Let the embedded representation of the candidate resume be denoted as vector $C \in Rd$, and the job description representation as vector J∈Rd, where d represents the embedding dimensionality.These vectors inhabit a shared semantic manifold that facilitates direct similarity computation.

Candidate job alignment is quantified using cosine similarity, defined as:

$S(C,J)=C\cdot J/ \|C\|\|J\|$

where

C·J denotes the dot product between vectors, and ‖C‖, ‖J‖ denote their Euclidean norms. Cosine similarity is selected due to its scale invariance and robustness to magnitude variation, ensuring that similarity estimation reflects directional alignment

This continuous similarity metric enables fine grained ranking and avoids the binary thresholding artifacts inherent in Boolean keyword matching systems. Furthermore, the use of dense embeddings mitigates synonymy and paraphrasing issues, allowing semantically equivalent professional terminology to converge within proximal vector neighborhoods.

C. Skill Gap Inference and Dependency Modeling

While cosine similarity provides a global measure of semantic alignment, it does not explicitly isolate missing competencies. To address this limitation, a symbolic skill extraction layer is integrated into the pipeline. Structured skill entities are identified from both candidate and job representations through contextual pattern recognition and entity categorization.

Let the required skill set derived from the job description be represented as SJ , and the observed candidate skill set as SC . The candidate gap profile G is computed via set theoretic difference:

$G=SJ-SC$

This formalization yields an interpretable representation of explicit competency deficiencies, thereby augmenting the global similarity score with granular diagnostic insight.

To translate the gap profile into actionable guidance, a dependency aware modeling framework grounded in graph theory is employed. Skills within the gap set are modeled as vertices within a DirectedAcyclic Graph, where directed edges encode prerequisite relationships. This ensures acyclic progression and prevents circular learning dependencies.

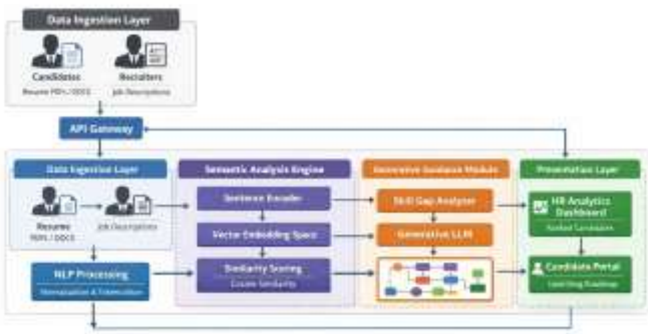
A topological sorting algorithm is applied to generate a valid acquisition sequence that respects prerequisite constraints. The resulting dependency ordered structure constitutes a personalized learning roadmap, logically segmented into progressive phases. This roadmap transforms abstract semantic misalignment into a structured remediation trajectory tailored to the candidate's existing knowledge base.

Technical Summary

The methodology integrates three complementary paradigms:

1. Statistical text normalization to stabilize heterogeneous document representations
2. Transformer based dense vectorsemantic encoding with cosine similarity for
continuous relevance estimation
3. Symbolic skill set inference coupled with graph theoretic dependency modeling for structured remediation

This hybrid architecture ensures high resolution semantic alignment, deterministic interpretability, and structured candidate development within a unified recruitment intelligence framework.



PROPOSED FRAMEWORK

This study introduces "TalentBridge AI," a comprehensive theoretical framework designed to bridge the semantic gap between candidate profiles and job requirements. Unlike traditional systems that rely on rigid keyword matching, this framework leverages a hybrid approach combining semantic vectorization with generative reasoning. The core logic consists of three distinct phases: Hybrid Information Extraction, Semantic Similarity Calculation, and Generative GapAnalysis.

A. Hybrid Resume Parsing via GenerativeAI

Traditional resume parsers often utilize Regular Expressions (Regex), which are brittle when facing varied document layouts. To overcome this, the proposed solution employs a Generative AI approach for structured information extraction.

Raw text is first extracted from documents (PDF/ DOCX) and passed to a Large Language Model (LLM), specifically Gemini-3-4b-it [5]. The model operates under specific system instructions to identify and categorize unstructured text into standardized entities— such as Work Experience, Education, and Technical Skills—returning a validated JSON object. This ensures high-fidelity parsing regardless of the resume's formatting style.

B. Semantic Similarity Engine (SBERT)

To address the limitations of lexical (word-count) matching, this framework implements a semantic search mechanism using the all-MiniLM-L6-v2 capture the contextual meaning of professional terminology.

Embedding Generation: The parsed resume text (Tr) and the target job description (Tj) are processed to generate dense vector embeddings (Vr and Vj). These embeddings map the textual data into a 384-dimensional vector space where semantic proximity represents professional alignment.

Similarity Calculation: The alignment between the candidate and the job is quantified using Cosine Similarity, which measures the cosine of the angle between the two vectors. This method is preferred over Euclidean distance as it remains robust regardless of document length.

Match Score = CosineSimilarity(Vr, Vj) x 100

C. Generative Gap Analysis and Roadmap Design

A novel contribution of this framework is the automated generation of remedial career roadmaps. If a candidate's match score falls below a specific threshold, the system triggers a Generative Gap Analysis.

Gap Identification: The system isolates specific skills or experiences present in the job description vector (Vj) but absent in the resume vector (Vr).

Context-Aware Prompting: A dynamic prompt is constructed, injecting the specific Job Description, the calculated Match Score, and the identified Missing Skills into the LLM's context window.

Structured Roadmap Generation: The LLM generates a personalized learning path divided into logical phases (e.g., "Phase 1: Foundations," "Phase 2: Advanced Concepts"). The output is strictly enforced as a JSON structure, allowing for the rendering of interactive visual timelines rather than static text advice.

RESULTS AND DISCUSSION

The development and deployment of the TalentBridge AI framework have yielded significant qualitative improvements in how candidate profiles are processed and matched against job requirements.

The system's performance was evaluated based on three primary capabilities: semantic accuracy, parsing resilience, and the generation of actionable career guidance.

A. Semantic Matching Capabilities: The integration of the all-MiniLM-L6-v2 Sentence Transformer provided a distinct advantage over traditional keyword-based Applicant Tracking Systems (ATS). Qualitative analysis of the matching engine revealed that the system successfully identifies candidates with relevant experience even when exact terminology differs from the job description. For instance, the model accurately correlates disparate terms such as "Machine Learning" and "AI," or "React" and "Frontend Development," ensuring that qualified candidates are not unfairly filtered out due to vocabulary mismatches. This semantic understanding significantly reduces false negatives, a common pitfall in standard recruitment software.

B. Resume Parsing and Entity Extraction The hybrid parsing approach, utilizing pdfplumber for text extraction and Gemini-3-4b-it for entity recognition, demonstrated high resilience against varied document layouts. Unlike rule-based parsers that often fail when encountering complex formatting (e.g., dual-column layouts or non-standard headers), the Generative AI model successfully contextualized and extracted structured data—specifically Skills, Education, and Work Experience—into a standardized JSON format. This ensures that the downstream matching algorithms receive clean, normalized data regardless of the resume's original design.

C. Personalized Roadmap Generation A novel outcome of this research is the system's ability to function as an automated career coach. The Generative Gap Analysis module successfully bridged the disconnect between assessment and improvement. By identifying the vector space delta between the resume and the job description, the system generated coherent, phased learning roadmaps. These roadmaps provided users with logical progressions—moving from foundational knowledge to advanced concepts—specifically targeted at their identified missing skills. This transforms the recruitment platform from a passive filter into an active educational tool, empowering users to upskill effectively.

CONCLUSION

This research presented "TalentBridge AI" a comprehensive framework designed to modernize the recruitment landscape through the convergence of Semantic Search and Generative Artificial Intelligence. By addressing the limitations of traditional keyword-based filtering, the proposed solution offers a more equitable and semantically aware method for evaluating job candidates. The study confirms that leveraging Sentence-BERT for embedding generation allows for a nuanced understanding of professional qualifications, ensuring that matches are based on capability rather than lexical exactness. Furthermore, the integration of Large Language Models (LLMs) for dynamic roadmap generation represents a significant step forward in personalized career development, providing candidates with actionable insights to bridge their skill gaps.

The successful implementation of this web-based platform demonstrates the feasibility of democratizing access to high-quality career guidance. Ultimately, TalentBridge AI not only enhances hiring efficiency for recruiters but also empowers job seekers by making the path to their desired career transparent and achievable. Future work will focus on extending the framework to support multi-modal inputs, such as video resumes, and integrating real-time labor market data to further refine the learning roadmaps.

ACKNOWLEDGEMENT

The authors are very much thankful to management for providing us facilities to conduct the experiments in the institution's laboratory and colleagues for their support to prepare this paper. We are also thankful to the Science & Technology department for approving our research proposal and providing us grants to conduct research activities.

REFERENCES

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

[2] Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing.

[3] Faliagka, E., Tsakalidis, A., & Tzimas, G. (2012). An integrated e-recruitment system for automated personality mining and candidate ranking. Internet Research, 22(5), 551-568.

[4] Senthil Kumar, N., Kavi Priya, S., & Sneha, M. (2021). Semantic-based Resume Ranking System using Natural Language Processing. International Journal of Scientific Research in Computer Science, Engineering and Information Technology.

[5] Google Gemini Team. (2023). Gemini: A Family of Highly Capable Multimodal Models. arXiv preprint arXiv:2312.11805.