FMDB *Transactions on Sustainable Computing Systems*

Semantic and Structural Optimization of Resumes via LLMs for Improved ATS Matching

Ojas M. Agarwal[1], Devi Karuppiah[2,*], Lekshmi Kalinathan[3], Benila Sathianesan[4], Monish Sudhagar[5], Vikram Ramkumar[6], Advaith T. Ajith[7]

[1,2,3,4,5,6,7]School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India.
ojas.magarwal2021@vitstudent.ac.in[1], devi.k@vit.ac.in[2], lekshmi.k@vit.ac.in[3], benila.s@vit.ac.in[4], monish.sudhagar2021@vitstudent.ac.in[5], vikramramkumar.r2021@vitstudent.ac.in[6], advait.tajith2021@vitstudent.ac.in[7]

Abstract: This paper reviews a novel LLM-powered pipeline for generating resumes that are dynamically tailored to job descriptions and optimised for Applicant Tracking Systems (ATS). In today's hiring landscape, most resumes are filtered before reaching a recruiter, primarily because of the rigid keyword-matching and formatting rules enforced by ATS algorithms. This presents a significant challenge for job seekers, especially those who cannot tailor their resumes for each application. Traditional resume tools often offer static templates or simple keyword suggestions, but generally fail to contextualise content and provide useful feedback. By combining structured parsing, semantic alignment, and iterative scoring, our system overcomes these limitations and generates highly customised, ATS-friendly resumes with minimal user effort. We demonstrate, through token-based and latent-space evaluations, that our optimised resumes nearly double their effectiveness compared to unmodified versions, improving alignment scores on average by 85%. This presents our approach as a valuable tool that enables job seekers to compete in increasingly competitive markets, while also representing a technological advancement.

Keywords: Large Language Model (LLM); Applicant Tracking System (ATS); AI Resume Generation; Semantic Alignment; Job Description Matching; ATS Algorithms; Technological Advancement.

Received on: 23/11/2024, **Revised on:** 17/02/2025, **Accepted on:** 28/03/2025, **Published on:** 07/09/2025

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSCS

DOI: https://doi.org/10.69888/FTSCS.2025.000477

Cite as: O. M. Agarwal, D. Karuppiah, L. Kalinathan, B. Sathianesan, M. Sudhagar, V. Ramkumar, and A. T. Ajith, "Semantic and Structural Optimization of Resumes via LLMs for Improved ATS Matching," *FMDB Transactions on Sustainable Computing Systems*, vol. 3, no. 3, pp. 192–202, 2025.

1. Introduction

The history of recruitment practice has taken on an unprecedented pace with the addition of artificial intelligence, in the guise of Applicant Tracking Systems (ATS). These are computer programs designed to screen resumes against pre-specified requirements, primarily through keyword searches, which have revolutionised the process of resume construction and screening. However, most job candidates unknowingly send non-ATS-compliant resumes and are screened out of contention despite being qualified [1]. To fill this gap, powerful tools such as large language models (LLMs) (such as ChatGPT) have been developed to generate, optimise, and customise resumes to meet ATS components. LLMs offer contextual understanding,

*Corresponding author.

reorganise, and restructure content, and incorporate keywords related to a specific field to optimise for an automated system [2]. Resume writing is also made easier through LLMs, thanks to their ability to enhance authenticity and readability through natural language processing [4]. Recent studies show that, while using ATS feedback and LLM results output, your chances of passing automated filters [3]. Furthermore, LLM parsing, compared to established machine learning algorithms, utilises LLMs as a form of analysis and achieves significant improvements in accuracy and relevance over traditional machine learning methods, indicating that the tools for acquiring talent are also changing [5]. As companies increasingly depend on automated systems to handle large volumes of applications, the need for LLM-powered solutions that generate ATS-optimised resumes becomes not only relevant but essential.

1.1. Overview of Resume Score Evaluation

The process used at the initial step of hiring, called resume score evaluation, is a key phase in shortlisting candidates. This early evaluation will be conducted by Applicant Tracking Systems (ATS), which takes the highest-scoring resumes for the job description and recommends them to the hiring manager. A score is typically determined in part by parsing the information in the resume, including the number of keywords, their semantic meaning, and the organisation of the information. However, traditional ATS models typically take in a limited amount of content and few creative ideas. This model has structural limitations in understanding the meaning of content within its context, which can lead to bias when evaluating resumes [1]. The addition of large language models (LLMs) has made this whole process subtler.

LLMs can analyse resumes not just based on keyword presence, but also by measuring semantic similarities and contextual relationships with respect to the job description. For example, with a system like ResuMatcher, an LLM compares resumes to job descriptions based on semantic similarity. This ultimately provides better accuracy and fairness in the evaluation [6]. These models can also dynamically adjust to various job descriptions, evaluating revised scoring measures based on the job's principles. This is a more personalised and informative evaluation process than strictly relying on keywords and their respective scores [7]. In addition, services such as JustScreen provide detailed score breakdowns, allowing candidates to identify which sections of their resumes need improvement and in which specific areas [8]. LLM-powered scoring models not only help the right candidates increase their chances of selection but also promote transparency and explainability, thereby fostering fairer employment practices [9].

1.2. Motivation

The evaluation of resumes for accuracy and fairness will always be a major challenge in automated hiring systems. Focusing specifically on keywords at the traditional Applicant Tracking System (ATS) level of automated hiring has created a situation in which even the best candidates can be knocked out of contention simply because the relevant terms are lacking in their resumes, which ultimately has created a gap between human reference checks and machine-filtering of resumes [5]. While potentially providing better contextual awareness, LLMS introduces a host of new issues into the automated hiring system. One of the biggest challenges presented by an LLM resume is that it may not align with job expectations worldwide.

An LLM may have all the right descriptors and appear impressive. In reality, it lacks relevance and the ability to demonstrate alignment with the match when evaluating resumes through the ATS [4]. Another major concern is bias in automated systems. Some research indicates that AI models, including LLMs, can exacerbate inherent biases in their training data and disadvantage certain applicant groups [10]. Furthermore, non-standardised resume formats and less standardised styles of use that vary by industry can lead to misinterpretation, even by sophisticated models, in parsing and evaluation [11]. Although transparency and explainability in resume screening systems have been attempted, for example, JustScreen and other models depend on understanding the scoring criteria and provide feedback systems for applicants; however, it is also complicated to build such systems at scale [8].

In summary, developing an effective resume screening process will require ongoing efforts to improve LLM algorithms and to establish ethical guidelines for their use [1]. This paper seeks to create an intelligent system that uses LLMs to generate user resumes for parsing by ATSs. It will create a bridge between how candidates express their capabilities and how the machine interprets them, producing resumes that align with both the user and the screening technology. The system will prepare users for tailoring a personal, ATS-compliant resume using formatting and keywords appropriate to the industry. Additionally, the work utilises an explainable AI for providing real-time feedback on the job-resume fit [12]. For example, the study investigates the usefulness of retrieval-augmented generation for candidate profiles, enabling deeper resume-job fit functionality in the paper [7]. The framework demonstrates that scoring based on graph neural network outputs can further improve resume-job fit, particularly in relatively unstructured formats. In short, the system offers the potential to shift from templated resume tools to smart, contextually supported resume engines. By making the tailoring process more informed and easier to navigate, the system aims to explore new ways for candidates to establish themselves as agents of competitive advantage in job markets.

2. Literature Review

2.1. Historical Overview

Over the past few decades, the process of resume screening has undergone significant changes. Initially, recruitment was entirely manual, and human recruiters reviewed each resume to assess qualifications, experience, and a good fit. Although this process is very methodical, it is also time-consuming, labour-intensive, and open to human bias. Client organisations began seeking a more efficient and consistent way to process resumes, which led to the emergence of early rule-based Applicant Tracking Systems (ATS). The first ATS only performed resume sorting based on keyword functions and structured resume formats, generally eliminating the need for human interaction [13]. ATS became more common during the late 1990s and early 2000s as organisations dealt with an increasing number of applicants to screen. Unfortunately, the rigidity of the ATS criteria eliminated potentially qualified candidates whose resumes did not match those rigid templates. As noted by Rawat et al. [14], the earliest ATS system, developed by Marin Software in 1996, employed a keyword-filtering approach and faced several challenges, including historical bias and limited contextualization.

Further efforts were made to improve parsing accuracy and to leverage the simplest applications of natural language processing (NLP) to extract meaningful, structured data (e.g., education, skills, work history). Systems such as PROSPECT, created more than 10 years ago, were among the first to automate candidate screening by scaling textual data review [15]. Despite advances in ATS ideation and screening methods, which are tested at the litmus of speed and reach, traditional ATS systems continue to struggle with unstructured resumes and the unique, nuanced experiences candidates bring. Chavan et al. [16] noted that the systems developed, intended to utilise historical hiring data and machine learning enhancements, still struggled to ensure fairness and comprehensive evaluation [16]. Overall, while the path towards automation has progressed from human execution (sending and reviewing physical resumes) to rule-based and somewhat intelligent ATS systems, it also highlights the limits that remain, namely the understanding of human context embedded within resumes.

2.2. Introduction of ML in Resume Parsing

The emergence of machine learning (ML) in resume parsing and analysis was a landmark moment in the history of technological advancements in recruitment. Previous systems used rule-based approaches (ATS), but machine learning began to identify underlying patterns in resumes rather than simply matching keywords. These solutions were trained to classify resumes based on job titles, education, skills, and experience, thereby improving the accuracy of candidate evaluation. One of the first uses of this approach was the combination of natural language processing (NLP) with ML to parse resumes and rank them based on acceptable job characteristics [17]. Research and analysis on classification and classification methods were performed repeatedly, and many systems have been improved as a result. Roy et al. [18] developed a recommendation system to automate resume shortlisting using ML classification algorithms, which also assisted recruiters in performing their roles more efficiently without compromising candidate selection [18].

Subsequent versions developed advanced models, such as Support Vector Machines (SVM) and BERT, to extract deeper semantic representations from resumes for job characteristic matching. Tian et al. [19] demonstrated how the use of a combination of Latent Semantic Analysis (LSA) and BERT in classification models resulted in a closer match between resume content and job descriptions in business process management [19]. Field-wide systematic reviews confirmed that ML not only increased automation but also provided flexibility and adaptability, allowing for the resumption of screening systems. Sinha et al. [20] reviewed the literature on various screening methods and demonstrated how ML enabled recruiters to address issues of format variation and inconsistent information presented on resumes. These approaches laid the foundation for new generations of intelligent hiring systems, which ultimately led to LLMs and systems capable of providing even better contextual awareness and personalisation.

2.3. Rise of Large Language Models in Recruitment

The emergence of larger language models (LLMs) has profoundly altered the recruitment landscape, enabling deeper contextual understanding and greater automation than previous ML tools. LLMs can understand not only the complexity of job requirements and the scale and unstructured nature of resume data, as previous ML models have, but also the unstructured nature of resume data itself, as recent research has found that organisations can reduce resume-screening time by as much as 75% while increasing candidate-job fit accuracy [21]. LLMs, or large language models such as GPT-3, are being examined for their potential to classify resumes, rank candidates, and even generate new text to describe job descriptions. For example, Rithani and Venkatakrishnan [22] demonstrated the ability to determine and classify resumes into a specific family of IT occupations with high accuracy, enabling high task-specific performance coding. When LLMs are developed in multimodal forms (e.g., video, LinkedIn connectivity, and scratch resumes), access to a qualitative assessment of the candidate is gained beyond simple textual review [23]. In addition to performance improvements, LLMs aid in detecting bias and testing fairness

in hiring. Wilson and Caliskan's [24] research illustrates how an LLM's role in document retrieval can unintentionally provide information that relies on race- and gender-based assumptions. This research highlights the growing importance of implementing ethical frameworks. Finally, the potential of LLMs to generate synthetic data for training and fine-tuning the recruitment model has been validated. Skondras et al. [25] showed that LLMs could generate synthetic resumes, leading to improved classification of job descriptions and better group balance in the datasets. As these models improve, LLMs within recruitment systems will not only enhance recruitment efficiency and speed but will also lead to a reevaluation of candidate judgment criteria and processes.

2.4. Comparative Studies: ML vs LLM for Resume Evaluation

The transition from prior machine learning (ML) techniques referenced in the literature on resume evaluation and screening to LLMs is attracting greater academic investment. ML models, primarily based on structured data and feature engineering, are available to handle automated resume screening and fit, based on the job description, by classifying resumes according to prior placements. With LLMs and their contextual, semantic, and natural language constructs, as well as their enhanced capacity to process contextual data, they are increasingly outperforming classical models in virtually all academic fields [26]. Provided one of the first comprehensive studies of machine learning or ML models against large language models or LLMs when evaluating curriculum vitae. Their exploratory research revealed that LLMs were more aligned with recruiter decisions than ML models, as LLMs provided a richer contextual understanding. In contrast, ML models struggled with resumes presented in different formats. Another important study Vaishampayan et al. [27] was conducted by Vaishampayan et al. [27], compared human versus LLM-based resume matching. The study showed that LLM performed equally well as ML in recall and precision, and also had greater congruence with human judgment in resumes where candidate qualifications are not explicitly stated.

Kurek et al. [28] also offered a compelling demonstration of how zero-shot AI models, including LLMs, perform at least as well and in many cases significantly better than classical ML systems for job candidate matching tasks without requiring labelled datasets; however, their system operated best in a dynamic recruitment environment, which helped accelerate the incorporation of changes in job role descriptions and competencies. Interestingly, Bocharova and Malakhov [29] proposed an end-to-end neural network model called "ResJob- Fit," in which they compared traditional deep learning models to transformer-based LLMs for job-resume matching across multiple datasets; overall, the results consistently showed significant improvements in resume-job matching accuracy when using the LLM approach. These comparisons clearly demonstrate that LLMs are not merely incrementally better; they fundamentally alter the way resumes are interpreted, evaluated, and matched with job opportunities.

2.5. Challenges Identified in Past Research

Despite considerable advancements in automated resume evaluation through ATS and LLM technologies, challenges persist in their implementation. One difficulty is the mismatch between the context of the resume content and job descriptions. It should be noted that LLMs generate scripted content that is grammatically correct; however, the LLM-generated resume content, while polished, may be disconnected from actual job expectations. This disconnection can lead to misleading evaluations of resumes [4]. Furthermore, the issue of bias is a persistent challenge for automated systems today [10]. It has been established that the use of AI in recruitment tools can perpetuate gender, racial, or socioeconomic biases, reflecting biases in training data and leading to inequitable hiring practices. The authors identified a significant potential for systemic bias in LLM-powered screening tools and called for greater transparency and accountability in our AI systems. The second most significant challenge is explainability, which remains particularly challenging, especially with complex LLM architectures. Recruiters and candidates often struggle to understand why a candidate received a particular score or why their resume was ranked higher or lower. Pathak and Pandey [30] identified a rubric and multi-agent evaluation models to help procurement increase trust, accountability, and transparency assessments using AI.

Practical implementation also has challenges. ATS systems have more difficulty parsing different resume formats, as well as inconsistent and overly creative resume layouts, even when used with LLMs. As described by Vijayalakshmi et al. [3], the screening still suffers from structural differences and poor formatting exhibited by the applicants. Finally, concerns regarding data privacy and scalability remain unanswered. The computing power required for widely disseminating LLMs is extremely high, and there are significant issues with storing and processing bidirectional sensitive candidate data. These challenges suggest that LLMs provide high-value promise and capabilities for the recruitment function, but require intelligent and ethical design for optimal operationalisation.

3. Methodology

3.1. System Overview

Modern recruitment increasingly demands precision, personalisation, and adherence to automated screening processes. Applicant Tracking Systems (ATS), now fully integrated into corporate candidate hiring pipelines, create an algorithmic filter effect, annotating and cataloguing candidates' experience information, while distancing them from human review. This mechanisation has created a high demand for resumes that are structured and dynamically oriented with the requirements referenced in a job description. Maximising candidate experience through resume development no longer hinges solely on conventional resume-writing methods. Linear templates, anecdotal rules about formatting, and generalised keyword stuffing offer little insight into the contextual relevance required for a dynamic response. To address such problems, the system we propose in this work employs a resume creation process that utilises deep semantic reasoning and structured data presentation. The system relies on a Large Language Model (LLM), which possesses flexible linguistic and contextual reasoning capabilities, enabling it to accurately conceptualise and read both candidate resumes and job descriptions.

Instead of following fixed formatting conventions or keyword-matching techniques, the system views resume creation as a bidirectional mapping problem: mapping a user's qualifications to the job requirements specified in the job posting, both implicitly and explicitly. The system maps qualifications to job postings during a multistage pipeline. The pipeline begins with data extraction and ends with a fully customised resume document optimised for ATS. All three stages of the pipeline are theoretical, modular, and deterministic. Modularity creates a level of interpretability and flexibility because each transformation from raw input to document format preserves a history of that transformation, which can be continuously improved based on the quality of the generated documents. The process also includes feedback loops and performance measures that not only generate a resume but also measure and evolve its quality over time. By establishing this work not just as an automation tool but as a system in the evolution of intelligent systems for career support systems, in which the aim is significantly beyond just meeting the digital filters, but to positively improve a candidate's profile through the adoption of digital data freedom in hiring.

3.2. Working of the System

The system architecture is modular and allows automated resume generation based on an exact job description (and to be an ATS company). The architecture has five sequential phases: Input Module, Resume Building Pipeline, Rendering Engine, Scoring Engine, and Feedback Loop. Each module is designed to navigate a resume that represents the individual's authentic experience, while also meeting the user's expectation of what a job posting embodies. The operational flow and intermodule dependencies are further illustrated in Figure 1, which shows the architecture of the auto-resume building LLM pipeline.

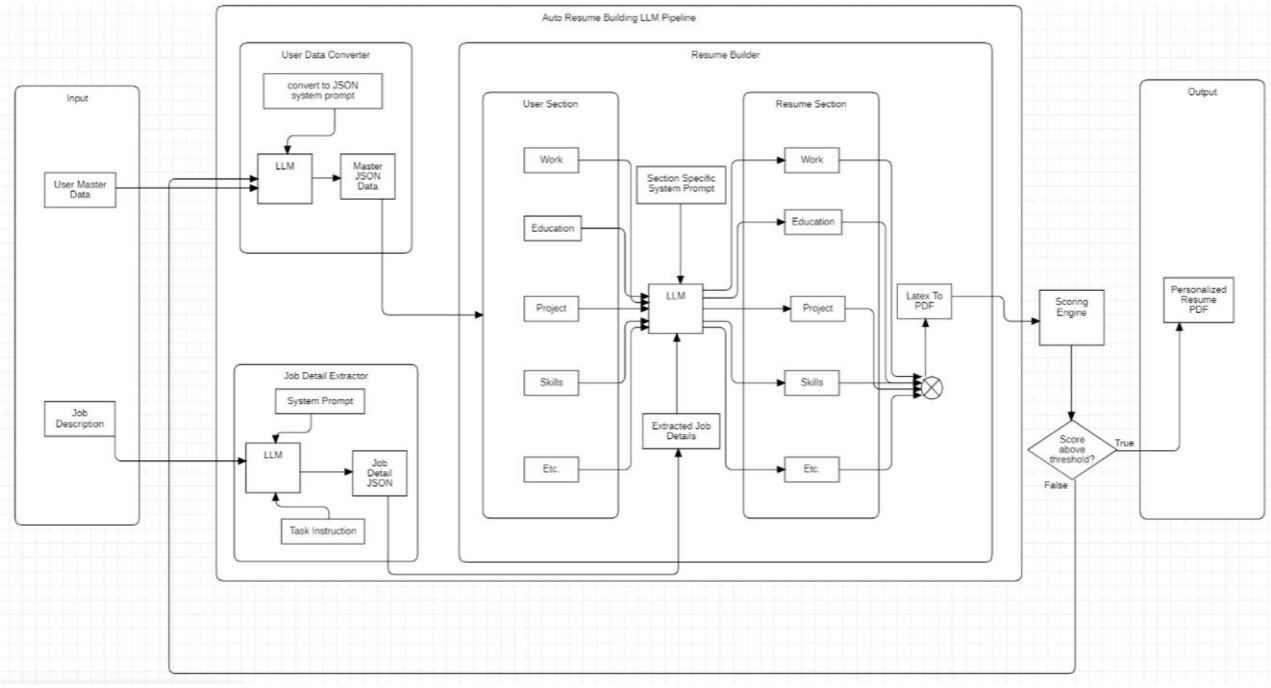

Figure 1: Architecture of the resume-building LLM pipeline

3.2.1. Input Module

The system initiates the pipeline by accepting two distinct inputs: the user-provided master resume and the job description (JD). The master resume is a comprehensive, active profile of the user that encompasses all relevant qualifications, including educational background, skills, work experience, certifications, and papers. The input can be provided as a PDF file or plain

text. At the same time, JD (raw text or URL) is treated as the contextual basis for the resume <adaptation process. The JD may include explicit requirements for qualifications, skill sets, role-specific competencies, and terminology specific to that domain.

3.2.2. Resume Building Pipeline

This pipeline is the central computational motion behind the system and is made up of three modules:

- **User Data Converter:** After extracting content from the user's raw resume and parsing it with a Large Language Model (LLM), the raw resume file will be converted to a JSON schema. This processing will follow prompt engineering, and through a protocol based on prompts, the LLM will parse out the various contents inside of the resume and categorize those contents (such as Work Experience, Education, Skills, and Pappers) with a structured document to describe the resume contents by labeling contexts and creating the meaning from the many semantically complete contexts found in the resume. Based on this reported process, the framework will enhance the system's capacity to search, select, and reconstitute content to customise the resume.
- **Job Details Extractor:** The JD is processed in parallel with this user data and parsed to extract the main expectations and constraints. Another instance of LLM parses this input to determine, identify, and organise the required qualifications, core competencies, preferred technical stack, levels of experience, and strategic keywords. Similar to user data, this information is structured in JSON format, enabling precise comparisons and alignments.
- **Resume Builder:** The final phase of the pipeline takes the parsed user data and the structured JD and passes them to the resume builder. Using section-specific LLM prompts, the builder generates robust resumes with relevant qualifications and job-specific requirements. Any sections, from Skills to Papers to Experience, can be adapted to maximise the attributes most important to the job requests, while maintaining all the intended authenticity and styling. This focused prompting strategy addresses both the customisation for each job and the document's overall integrity.

3.2.3. Rendering Engine

Once the resume content is generated, it flows into the rendering engine, which was developed with LaTeX templating via Jinja, along with the creation of a predefined professional resume document. The rendering engine dynamically populates the resume with engine-controlled formatting options, section order controlled by input priority, and dynamic layout options on the output end. The final resume is professional in look and presentation, compliant with ATS parsing rules, and displays a current trend layout; it is also ready to print as a PDF.

3.2.4. Scoring Engine

To determine the degree of alignment between the generated resume and the JD, the system utilises a scoring framework based on metrics derived from NLP techniques. The metrics we include are keyword match, skill match, sentence structure, section density, and readability. The score serves a dual purpose. First, it provides an objective assessment of resume quality, and second, it serves as a go/no-go for initiating the iterative refinement process.

3.2.5. Feedback Loop

If the resume scores are below the prescribed threshold, the system will enter a feedback loop. The user is then prompted to append or clarify, e.g., missing certifications or skills not included, which may help align with the JD. This additional information is reinserted into the pipeline, and the process repeats until the resume satisfies the required optimisation score. This iterative design will help ensure that the eventual output is not only tailored but is similarly positioned for consideration within automated recruitment systems for similar positions. In summary, the working methodology is a closed-loop, LLM-based system that simulates subject-matter expert-level resume personalisation with minimal user effort at the outset. It has been demonstrated that semantic intelligence, structured data modelling, and user-centred design can come together to address deeply rooted inefficiencies in the resume writing process.

3.3. Evaluation Metrics

To measure the alignment of a generated resume with its associated job description, the system employs a series of document similarity measures that assess lexical and semantic overlap between two text representations: the candidate's optimised resume and the job requirements. The scoring mechanism is designed not only to assess the presence of keywords and the relevance of sections, but also to provide a score based on how accustomed the text is to a specific context. Three primary similarity measures are used in the scoring engine: Cosine similarity, Jaccard similarity, and overlap coefficient. Each similarity measure offers a distinct mathematical framework for assessing document alignment.

3.3.1. Cosine Similarity

Cosine Similarity measures the angular distance between two document vectors representing term-weight distributions in high-dimensional space, typically derived using techniques such as Term Frequency-Inverse Document Frequency (TF-IDF). Cosine similarity measures the similarity between two documents based on their term usage patterns, disregarding document length. Cosine similarity ranges from 0 to 1 (cosine similarity ranges from -1 to 1, but for purposes of measuring document similarity, -1 has no value), where one indicates the documents are aligned perfectly in direction, or identical direction of term distribution, and 0 represents orthogonal vectors and, therefore, zero common terms. Mathemat- ically, given two vectors A = [a1, a2, . . . , an] and B = [b1, b2, . . . , bn], the cosine similarity is computed as:

$$\text{Cosine Similarity} = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2} \cdot \sqrt{\sum_{i=1}^{n} b_i^2}} \quad (1)$$

Generally, an increase in cosine similarity indicates improved semantic alignment between the resume and the job description, while a decrease suggests altered language or divergence in topicality.

3.3.2. Jaccard Similarity

The Jaccard Similarity measures the lexical overlap between two documents by counting the number of distinct terms that appear in both documents. Jaccard similarity measures the number of terms they share out of the total distinct terms in both documents, i.e., the ratio of the intersection to the union.

For sets A and B, the metric is given by:

$$\text{Jaccard Similarity} = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

A Jaccard score closer to 1 means that the lexical is more similar, while a Jaccard score closer to 0 indicates more dissimilarity in word choice. This can be especially effective in measuring whether applicants have explicitly mentioned any required terms or skills mentioned in job postings.

3.3.3. Overlap Coefficient

The Overlap Coefficient is a measure of the extent to which two documents overlap. Rather than a union of sets, as with Jaccard, the overlap coefficient normalises the overlap over the size of the smaller set. This tends to be a more sensitive measure when comparing documents of significantly different lengths, such as a job description and a short resume.

The formula is expressed as:

$$\text{Overlap Coefficient} = \frac{|A \cap B|}{\min(|A|, |B|)} \quad (3)$$

If the overlap coefficient is 1, it means that the smaller document is fully contained within the larger document, indicating that the documents are highly relevant to each other. This measure can help identify whether a resume fully includes the most critical elements outlined in a job description. Collectively, these metrics offer a multifaceted perspective on document similarity. Cosine similarity reflects directional alignment in the semantic space; Jaccard similarity measures unique lexical overlap; and the overlap coefficient quantifies the extent of set membership. By considering these different perspectives, the scoring engine provides a solid and explainable assessment of resume-job fit. The differences across metrics reflect discrepancies that can serve as diagnostic considerations, influencing the feedback loop of improving or aligning resumes with job specifications.

4. Results

The results indicate significant improvements in resume-job alignment that can be facilitated by the LLM-based system described in this study. As shown above, for the resume similarity measure, both token-level and semantic-level scores improved before and after examination, demonstrating the effectiveness of this method in creating resumes aligned with job specifications. Table 1 summarises the results of the resume optimisation system, showing the improvement in alignment

between the candidate's resumes and the corresponding job descriptions. The values for job description and candidate resume alignment are based on the cosine similarity metric and were calculated between two sets of text representations: (1) the original, user-submitted resume; and (2) the LLM-produced, processed resume. The evaluation of both resumes occurred in two separate representation spaces: the token space and the latent space, to capture different aspects of textual similarity.

Table 1 illustrates cosine similarity scores calculated in the token space, where documents are represented as term-frequency vectors (typically in the TF-IDF format), which measure surface similarity, i.e., the number of words shared exactly or nearly exactly between the resume and the job description. A higher cosine score in the token space indicates that the resume uses more of the same vocabulary as the job description, which may include keywords, skills, or role-specific phrases. These are important because they are what most ATS (Applicant Tracking System) screening algorithms use to parse a resume. Additionally, most ATS systems prioritise word-level matches when parsing and screening resumes.

Table 1: Token and latent space cosine scores

Token Space Cosine Score		**Latent Space Cosine Score**	
Original Resume Score	**LLM Produced Score**	**Original Resume Score**	**LLM Produced Score**
0.2	0.54	0.17	0.60
0.4	0.71	0.40	0.79
0.43	0.65	0.36	0.61
0.31	0.58	0.32	0.57
0.31	0.53	0.47	0.87
0.42	0.76	0.24	0.48
0.23	0.35	0.50	0.80
0.4	0.75	0.22	0.37

Table 1 also includes cosine similarity scores computed in the latent space, where both resumes and job descriptions were converted to dense vector embeddings using a language model. This representation captures semantic similarity, meaning it considers how similar the two documents are in terms of meaning, regardless of whether the same words are used. Although phrases like "paper oversight" and "team leadership" may not overlap lexically, their latent space can still demonstrate significant alignment if the underlying concepts are complementary. This type of similarity is especially relevant for systems that use ATS modules based on LLMs, which evaluate the contextual relevance of a candidate's experience relative to the job description rather than relying solely on terminology. In both parts of the table, the left column displays the cosine similarity between the original resume and the job description. In contrast, the right column shows the same relationship after the resume has been optimised. Each row indicates a resume-job description pair.

In all cases, we observe an increase in similarity scores after optimisation in both token and latent spaces. To gauge the system's effectiveness, we calculated the average relative increase in alignment score for all evaluated samples. We found that the average improvement was approximately 85%, indicating an overall improvement in alignment between optimised resumes and their associated job descriptions. A change that reflects improvements in both lexical and conceptual instances that allow the resume to align better with the expectations of automated screening systems and human recruiters. These metrics confirm that the resume optimisation workflow increases alignment not just by copying and pasting prominent keywords, but by modifying content to meaningfully and contextually align concepts, structure, and language with relevant job descriptions.

5. Discussion

Although tremendous strides have been made with the inclusion of Large Language Models (LLMs) and Applicant Tracking Systems (ATS) into recruitment processes, important research gaps remain. For example, many existing systems remain non-adaptable, and transparency is still inadequate. A significant gap exists in the lack of explanation regarding how LLMs rank and interpret resumes [9]. Underscore that there is a need for a context-sensitive and explainable framework for recruiting that could explain the reasoning behind screening decisions and do so in a recruiter-friendly manner. In addition to the gaps mentioned above, current systems may still lack effective ethical frameworks. Navarro identifies a fairness gap in resume screening tools and proposes a design for JustScreen to mitigate bias and enhance transparency, utilising principles of ethical artificial intelligence. Anand and Giri [1] also note that while LLMs can support higher positive ATS scores, they may sometimes disregard critical resume authenticity, thereby misleading recruiters. Despite the widespread availability of increasingly sophisticated embeddings and scoring models, perhaps the most significant barrier is the lack of large, real-world, and diverse training datasets.

In the Baghbanzadeh [7] study of graph neural networks combined with LLMs, he found that while this combined approach generated strong results, the models were trained on and then tested on tightly constrained academic datasets [7]. In addition to the usability gap noted by Somavarapu and Sharma [31] with LLM-enabled platforms, while they emphasise high levels of accuracy with screening, they also note that usability and real-time feedback systems are often underdeveloped, which affects adoption within enterprise environments. As a final note, Manikran Pedige [21] notes in their evaluation of AI recruitment systems, goes back to the need for learning models - tied to changing labour markets, as well as applicable vocabulary or professionalism from industry, inadvertently still lacking with most static LLM systems. These usability and model gaps suggest a clear direction forward for more effective national capacity building, with the basic foundations of better datasets, explainability, bias prevention, and adaptive models tied to industry characteristics, for future LLM-based resume screening solutions.

6. Conclusion

The advancements in recent years in resume evaluation technologies represent steady progress towards greater efficiency, fairness, and contextualization in recruitment. With resumes, we have transitioned from paper to rules-based ATS systems, partly to attract better candidates and partly to support somewhat manual improvements. Meanwhile, traditional machine learning systems achieve better efficiency but leave gaps in fairness. With LLMs, we are seeing the biggest jump in experience and innovation, as they are not only automating tasks but also enhancing them in the process. Still, their understanding of language is highly contextualised, lending itself to a more nuanced understanding of a resume and its positioning in relation to any job description. This study investigated these developments and introduced a new, fully automated resume-generation system based on an LLM. Our pipeline, which consists of structured resume retrieval, job-specific adaptation, semantic evaluation, and iterative optimisation, demonstrates how language models can be operationalised to produce contingent, personalised, and ATS-optimised resumes.

The data we collected confirmed the system's efficacy: resumes improved on average by 85% in alignment score, while nearly doubling their compatibility with job descriptions in both the token and semantic spaces. This demonstrates the technical utility of the methodology, but its practical capabilities to help increase candidates' chances of being selected. Looking ahead, the continued development of transparent, ethical, and adaptable systems will be paramount. By coordinating language modelling technology, user-centred design, and measurable impact, we will move closer to intelligent, inclusive, and empowering recruitment tools.

Acknowledgement: The authors acknowledge the invaluable assistance of their supervisors and the resources provided by Vellore Institute of Technology. They also thank all collaborators for their dedicated involvement in this work.

Data Availability Statement: The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Funding Statement: The authors did not receive any specific grant or financial assistance from funding agencies, institutions, or organisations for the conduct or publication of this research.

Conflicts of Interest Statement: The authors declare that they have no known competing financial or non-financial interests that could have influenced the work reported in this study.

Ethics and Consent Statement: This study was conducted in accordance with established ethical standards, and informed consent was obtained from all participants prior to their involvement.

References

1. S. Anand and A. K. Giri, "Optimizing resume design for ATS compatibility: A large language model approach," *Int. J. Smart Sustain. Intell. Comput.,* vol. 1, no. 2, pp. 49–57, 2024.
2. V. Manish, Y. Manchala, Y. Vijayalata, S. B. Chopra, and K. Y. Reddy, "Optimizing resume parsing processes by leveraging large language models," in Proc. *2024 IEEE Region 10 Symposium (TENSYMP),* New Delhi, India, 2024.
3. V. Vijayalakshmi, A. Ananya, and M. U. A. Sharanya, "Optimization of HR recruitment process using large language model (LLM)," in Proc. *2024 1st Int. Conf. Innovations Commun., Elect. Comput. Eng. (ICICEC),* Davangere, India, 2024.
4. K. He and J. Lau, "Optimizing resume authenticity and ATS compatibility with LLM feedback integration," *Cal Poly SURP Poster,* 2024. Available: https://digitalcommons.calpoly.edu/ceng_surp/72/ [Accessed by 12/11/2024].

5. E. Kaygin, “Comparative analysis of ML and LLM in resume parsing: A paradigm shift in talent acquisition,” 2023. Available:https://www.researchgate.net/publication/378698002_Comparative_Analysis_of_ML_Machine_Learning_and_LLM_Large_Language_Models_in_Resume_Parsing_A_Paradigm_Shift_in_Talent_Acquisition [Accessed by 15/09/2024].
6. P. Dhobale, V. Bhoir, A. Vyavhare, P. Yelkar, and P. A. Dharmadhikari, “ResuMatcher: An intelligent resume ranking system,” in Proc. *2025 3rd Int. Conf. Intell. Data Commun. Technol. Internet Things (IDCIoT),* Bengaluru, India, 2025.
7. A. Baghbanzadeh, “Job-resume compatibility scoring using graph neural networks and large language models,” M.S. thesis, *Univ. of Windsor,* Windsor, Canada, 2025.
8. G. Navarro, “Fair and ethical resume screening: Enhancing ATS with JustScreen the ResumeScreeningApp,” *J. Inf. Technol. Cybersecur. Artif. Intell.,* vol. 2, no. 1, pp. 1–7, 2025.
9. F. P. Lo, J. Qiu, Z. Wang, H. Yu, Y. Chen, G. Zhang, and B. Lo, “AI hiring with LLMs: A context-aware and explainable multi-agent framework for resume screening,” in *Proc. Comput. Vis. Pattern Recognit. Conf. (CVPR),* Nashville, TN, United States of America, 2025.
10. S. Bhatnagar, S. Shetty, N. Arora, V. Sachdev, and A. Bahrini, “Beyond traditional biases in AI hiring: Exposing the hidden systemic challenges in resume screening,” in Proc. *2025 Systems Inf. Eng. Design Symp. (SIEDS),* Charlottesville, VA, United States of America, 2025.
11. R. V. Bevara, N. R. Mannuru, S. P. Karedla, B. Lund, T. Xiao, H. Pasem, S. C. Dronavalli, and S. Rupeshkumar, “Resume2Vec: Transforming applicant tracking systems with intelligent resume embeddings for precise candidate matching,” *Electronics*, vol. 14, no. 4, p. 794, 2025.
12. I. Subedi, “Candidate profile evaluation – A RAG approach with synthetic data generation for tech jobs,” M.S. thesis, *Techn. Univ. of Munich,* Munich, Germany, 2025.
13. L. Gagua, “E-recruitment and applicant tracking system: New age of technology based applicant screening. Threat or opportunity?” *Metropolia Ammattikorkeakoulu,* 2015. Available: https://urn.fi/URN:NBN:fi:amk-2015111016169 [Accessed by 22/09/2024].
14. A. Rawat, S. Malik, S. Rawat, D. Kumar, and P. Kumar, “A systematic literature review (SLR) on the beginning of resume parsing in HR recruitment process and SMART advancements in chronological order,” *Research Square,* 2021. Available: https://assets-eu.researchsquare.com/files/rs-570370/v1/9da1a6e1-437f-4f6d-a021-743ea3ee268e.pdf [Accessed by 05/09/2024].
15. A. Singh, C. Rose, K. Visweswariah, V. Chenthamarakshan, and N. Kambhatla, “PROSPECT: A system for screening candidates for recruitment,” in Proc. *19th ACM Int. Conf. Inf. Knowl. Manage. (CIKM),* Ontario, Canada, 2010.
16. P. R. Chavan, Y. Chandurkar, A. Tidake, G. Lavankar, S. Gaikwad, and R. Chavan, “Enhancing recruitment efficiency: An advanced applicant tracking system (ATS),” *Ind. Manag. Adv.,* vol. 2, no. 1, pp. 1–9, 2024.
17. S. Z. Sadiq, J. A. Ayub, G. R. Narsayya, M. A. Ayyas, and K. T. Tahir, “Intelligent hiring with resume parser and ranking using natural language processing and machine learning,” *Int. J. Innov. Res. Comput. Commun. Eng.,* vol. 4, no. 4, pp. 7437–7444, 2016.
18. P. K. Roy, S. S. Chowdhary, and R. Bhatia, “A machine learning approach for automation of resume recommendation system,” *Procedia Comput. Sci.,* vol. 167, no. 1, pp. 2318–2327, 2020.
19. X. Tian, R. Pavur, H. Han, and L. Zhang, “A machine learning-based human resources recruitment system for business process management: Using LSA, BERT and SVM,” *Bus. Process Manag. J.,* vol. 29, no. 1, pp. 202–222, 2023.
20. A. K. Sinha, M. A. K Akhtar, and A. Kumar, “Resume screening using natural language processing and machine learning: A systematic review,” in D. Swain, P. K. Pattnaik, and T. Athawale, Eds., Machine Learning and Information Processing, *Springer*, Singapore, 2021.
21. D. U. Manikran Pedige, “Leveraging large language models to transform recruitment in human resource management: An evaluation of AI-driven approaches,” M.S. thesis, Governance of Digitalization, *Åbo Akademi University*, Turku, Finland, 2025.
22. M. Rithani and R. Venkatakrishnan, “Empirical evaluation of large language models in resume classification,” in Proc. *2024 4th Int. Conf. Adv. Elect., Comput., Commun. Sustain. Technol. (ICAECT),* Bhilai, India, 2024.
23. A. Giri, P. Das, C. Harsha, and H. K. Suman, “Revolutionizing recruitment with large language models: A multimodal AI framework integrating video, social media, and traditional screening for efficient talent acquisition,” in Proc. *2024 1st Int. Conf. Data, Comput. Commun. (ICDCC),* Sehore, India, 2024.
24. K. Wilson and A. Caliskan, “Gender, race, and intersectional bias in resume screening via language model retrieval,” in Proc. *AAAI/ACM Conf. AI Ethics Soc.,* vol. 7, no. 1, pp. 1578–1590, 2024.
25. P. Skondras, P. Zervas, and G. Tzimas, “Generating synthetic resume data with large language models for enhanced job description classification,” *Future Internet,* vol. 15, no. 11, p. 363, 2023.
26. M. Maree and W. A. Shehada, “Optimizing curriculum vitae concordance: A comparative examination of classical machine learning algorithms and large language model architectures,” *AI*, vol. 5, no. 3, pp. 1377–1390, 2024.

27. S. Vaishampayan, H. Leary, Y. B. Alebachew, L. Hickman, B. A. Stevenor, W. Beck, and C. Brown, “Human and LLM-based resume matching: An observational study,” in *Findings of the Assoc. for Comput. Linguistics: NAACL 2025*, New Mexico, Mexico, 2025.
28. J. Kurek, T. Latkowski, M. Bukowski, B. Świderski, M. Łępicki, G. Baranik, B. Nowak, R. Zakowicz, and Ł. Dobrakowski, “Zero-shot recommendation AI models for efficient job–candidate matching in recruitment process,” *Appl. Sci.,* vol. 14, no. 6, p. 2601, 2024.
29. M. Y. Bocharova and E. V. Malakhov, “ResJobFit – end-to-end artificial neural networks based technology for job-resume matching,” *Appl. Asp. Inf. Technol.,* vol. 7, no. 4, pp. 378–391, 2024.’
30. G. Pathak and D. Pandey, “AI agents in recruitment: A multi-agent system for interview, evaluation, and candidate scoring,” *SSRN Preprint*, 2025. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5242372 [Accessed by 12/05/2025].
31. S. Somavarapu and P. Sharma, “Revolutionizing talent acquisition: Leveraging large language models for personalized candidate screening and hiring,” *Int. J. Res. Manag. Pharm.,* vol. 14, no. 1, pp. 69–91, 2025.