Explainable and Fair Resume Screening with BERT Embeddings and Tabular Attention Networks.

MSc Research Project
MSc. AI for Business

Preethi Algaskhanpet
Student ID: x24100587

School of Computing
National College of Ireland

Supervisor: Agatha Mattos

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Preethi Algaskhanpet		
Student ID:	x24100587		
Programme:	MSc. AI for Business	**Year:**	2025-2026
Module:	MSc. Research Project		
Supervisor:	Agatha Mattos		
Submission Due Date:	28-01-2026		
Project Title:	Explainable and Fair Resume Screening with BERT Embeddings and Tabular Attention Networks.		

Word Count: 5886 **Page Count 23**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Preethi Algaskhanpet

Date: 28-01-2026

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Explainable and Fair Resume Screening with BERT Embeddings and Tabular Attention Networks.

Preethi Algaskhanpet
24100587
NATIONAL COLLEGE OF IRELAND

Dataset link: AI_Hiring_Synthetic_Dataset_v2.xlsx

ABSTRACT

AI powered hiring system are widely used to automate the resume screening, but there is major issue faced by them is transparency and hidden bias within the historical recruitment data.The research proposes a hybrid model emphasizing explainability that combines BERT's strengths in extracting semantic representations from textual skills data with the TabNet architecture for capturing interactions between tabular features with the help of sequential attention. Public available synthetic hiring dataset was prepared based on the requirements, and then feature engineering techniques were done, including skill counts, experience level of the employee, cultural fit scoring, and numerical standardisation. Multiple models were included in the evaluation, such as Logistic Regression, Support Vector Machine (SVM), Decision Tree, Deep Neural Network and a hybrid BERT + TabNet model, with SHAP for interpretability. By the hybrid approach, BERT generated embeddings were combined into the TabNet model to enhance contextual understanding of the candidate profiles.This hybrid model achieved an accuracy of 87.2%, F1-score of 0.565, and ROC-AUC of 0.911 thus it performed better than the traditional ML models and DL model and at the same time, offered a great level of transparency through SHAP explanations.Even-though the Decision tree and Logistic regression had high accuracy, it lacked interpretablitiy , So that it is not suitable for real -world hiring contexts where transparency and fairness is important.The hybrid model provides a solution by combining the powers of semantic understanding, interpretable feature selection, and explainable output visualization, thus granting it the robustness and transparency of an ethical AI-driven recruitment framework.

Keywords: AI hiring systems, Explainable AI, BERT embeddings, TabNet, SHAP values, Resume screening, Fairness in machine learning, Human-centered AI, Hybrid deep learning model.

1: INTRODUCTION

1.1 Background

Artificial intelligence has become more important into the recruitment domain and the automated screening systems of applicants used by the organizations to filter, rank or shortlist the candidates. The reason behind this is that the use of AI has reduced the burden of manual review and more importantly made the screening process faster. The hiring domain is more complex dues to many factors such as ambiguity in the text descriptions, varations in the formates, contextual skill descriptions, and historical biases embedded in past hiring.This implies that developing an AI system which interprets candidate information while maintaining fairness is a big challenge for the recruitment industry (Lo et al., 2025).

The conventional machine learning techniques such as logistic regression, SVM, and decision trees which are simple and easy to understand have been accepted for automated recruitment. The major problem with these is that they are usually based on keyword or numerical encodings which cannot reflect the true meaning in the context of the resumes. Moreover, they have the assumption of linearity, cannot discover semantic relationships, and thus are not suitable for handling complex and unstructured recruitment datasets (Rigotti & Fosch-Villaronga, 2024). On the other hand, deep learning approaches especially transformers, like BERT, have proven their superiority in understanding contextual language patterns within large text corpora. BERT can not only identify but also differentiate between the technical skills, job titles, and soft-skill expressions being mentioned. However, BERT's interpretability issue together with its “black box” nature makes it difficult for employment-related decisions (Iso et al., 2025).

Interpretability is essential to ensure the compliance with the rising ethical AI standards and also the EU AI Act, which is why both standards insist on the transparency of the automated decision-making systems. This research offer a solution to the address the limitations by proposing a hybrid Explainable AI (XAI) model, composed of BERT for semantic text embedding and TabNet, which is an interpretable deep-learning architecture for tabular data. Moreover, the SHAP (SHapley Additive Explanations) tool is employed to generate comprehensible justification of the predictions made by the model on both global and individual candidate levels. The main goal of this hybrid approach is to make the AI-powered resume screening system more accurate and at the same time more interpretable.

1.2 Problem Statement

The existing artificial intelligence-based recruitment systems, irrespective of whether they are based on statistics or neural networks, struggle to balance between accuracy, fairness, and interpretability. The traditional machine-learning models are unable to capture the contextual meaning inherent in skills and job descriptions, and on the other hand, the deep neural networks like BERT while having a powerful semantic understanding, lack transparency. The absence of adequate interpretability makes it impossible for the organizations to trust the model outputs completely or to spot discriminatory patterns which in turn leads to ethical and legal issues (Mujtaba & Mahapatra, 2024).

Researchers have also pointed out that AI hiring models might unintentionally learn biases from previous hiring decisions that were based on gender, nationality, or education and thus, create similar disparities (Wilson & Caliskan, 2024). Even though some studies have implemented explainability tools or fairness constraints, the number of such studies is very few as compared to those that have explored hybrid architectures that combine semantic understanding and interpretable decision frameworks. Besides that, there is a lack of work that has evaluated XAI methods specifically on structured hiring datasets that are enriched with textual skill descriptions.

The proposed model serves as a solution to these shortcomings by combining together BERT, TabNet, and SHAP, thus creating a hybrid model that merges text embeddings of a particular context with a selection of features based on interpretable mechanisms. The introduction of this framework is expected to not only make the hiring process more transparent but also to deliver better predictive performance than the individual models.

1.3 Research Question:

Can a hybrid model combining BERT and TabNet enhanced with SHAP , improve the fairness and transparency when compared to standalone models in AI based resume screening system?

1.4 Research Objectives

This research has four main goals:

- To review and analyze state-of-the-art the most advanced AI recruitment solutions, especially in the area of transformer-based models (BERT), interpretable structures (TabNet), and explainability frameworks (SHAP).
- To design and implement a hybrid XAI model where BERT-generated embeddings with TabNet in order enhance better accuracy and interpretaability.
- To train and evaluate the hybrid model with different performance metrics such as Accuracy, Precision, Recall, F1-score, and ROC-AUC, and also comparing it with baseline models like Logistic Regression, SVM, Decision Tree, and Deep Neural Network.
- To use SHAP for both global and local interpretability, enabling assessment of potential bias and improving the transparency in candidate evaluation.

1.5 Significance of the Study

The study provides both theoretical contributions and practical benefits to the real world AI recruiting system.

Theoretical Contribution

- Creation of a new integration of BERT embeddings, TabNet’s sequential attention, and SHAP-based explanations for resume screening.
- Demonstrates that hybrid XAI models can successfully deal unstructured textual data and structured hiring characteristics.
- Provides a transparent evaluation framework to analysis the fairness, interpretability, and predictive performance in the hiring algorithms.

Practical Contribution

- Assisting HR and organizations in the process of making ethical, transparent, and auditable hiring decisions.
- Ensures that the predictions for each applicant can be explained through SHAP, hence bias risk is lowered.

- Provides a solution that is both scalable and adherent to regulations, as it corresponds with the EU AI Act and the best practices of the industry.
- Enhances the trust in Automated resume screening and improves recruitment efficiency.

1.6 Structure of the Paper

The report is structured in the following way:

- Chapter 1 – Introduction: outlines the motivation for research, problem statement, objectives, contributions, and overall structure of the study.
- Chapter 2 – Literature Review: rather summarizes previous research on AI hiring systems, transformer models, TabNet, fairness problems, and XAI techniques. Chapter 3 – Methodology: tells about the dataset, preprocessing methods, creation of the baseline model, BERT + TabNet hybrid pipeline, and SHAP explainability workflow.
- Chapter 4 – Results and Discussion: Discusses the results of experiments, comparisons of performances, and the analysis of the interpretability.
- Chapter 5 – Conclusion and Future Work: Summarizes key findings, highlights research limitations, and improvement on XAI hiring systems .

2. RELATED WORK

2.1 Introduction

The rise of AI-powered recruitment systems has transformed the way organizations process the resumes and evaluate the candidate. Manual screening of Traditional recruitment pipelines is the most dominant method of selection but it has several drawbacks like being overly time-consuming, inconsistent, and susceptible to human error (Abhishek et al., 2025). Automated resume screening has become one of the most vital areas of research with the increasing number of large-scale online job platforms and applicants per vacancy (Arvindh, 2025).

The literature review analysis the limitations of the traditional machine learning resume screening models, and then rise of the deep learning with transformers and text representation, as well as enhancing the fairness and explainability in the research. The specific gap that this study intends to address has been indicated in the review where the merging of BERT embeddings, TabNet feature selection, and SHAP explainability in a hybrid model to improved fairness, enhance transparency, and predictive performance in the resume screening procedure (Mujtaba & Mahapatra, 2024; Wilson & Caliskan, 2024).

2.2 Traditional Resume Screening Models

The early automated resume screening systems made use of classical machine learning techniques like Logistic Regression, Naïve Bayes, Random Forests, and SVMs. These models are highly dependent on the features that are manually or semi-automatically engineered, which include but are not limited to, keyword frequency, TF-IDF vectors, skill counts, and categorical encodings (Tayal et al., 2024).
These traditional models, on the one hand, are fast, efficient, and easy to understand,

while on the other hand, their performance is limited due to their inability to understand the semantics of the content. They analyze resume as a bag of words, leaving out the surrounding context and linguistic nuances (Saatci et al., 2025).
Research by Tayal et al. (2024) and Saatci et al. (2025) indicate that classical modeling, achieve reasonable baselines performance, but struggle with:

- synonyms variability (for instance: "managed" vs "coordinated")
- long description of text
- multi skill relationship
- domain- specific phrases

These limitations make the classical machine learning techniques insufficient for the screening process in the environments where fairness, transparency and accuracy are critical.

2.3 Deep Learning Models for Resume Screening

The data in resumes is rich in text, deep learning techniques were preferred for finding the complex patterns of candidate profiles. Convolutional Neural Networks (CNNs), Long Short Term Memory (LSTM), and Gated Recurrent Unit (GRU) are some of the models that have been successfully applied in different areas like skill extraction, job-candidate matching, and resume classification. GRUs and LSTMs outperforms traditional machine learning methods in sequential text processing, however, they lack interpretable and also need large labeled datasets to work (Saatci et al., 2024; Tayal et al., 2024). Further, the deep learning models get biased as they are trained on real-world recruitment datasets that include historical discrimination.
Moreover, the DL models provide better semantic understanding than the traditional ML methods but still are limited when compared to the transformer-based architectures. They also provide little interpretability which is a major drawback in hiring applications where decisions must be auditable(Wilson & Caliskan, 2024; Wehner & Köchling, 2020).

2.4 Transformer-Based Models for Semantic Resume Understanding

Transformer architectures such as BERT has significantly improved the research on resume screening. BERT provides such embeddings that capture the deeper semantic meaning and enhance the better understanding of skills, experience, and soft-skill indicators (Iso et al., 2025). The research of Burity et al. (2025) and Shrestha et al. (2024) has demonstrated that transformer-based retrieval system outperformed keyword-based and recurrent neural methods for job-candidate matching.

BERT's key strengths in the screening of resumes:

- understand the linguistic context instead of individual keywords
- handles complicated technical skill descriptions
- reduces the bias caused by the superficial frequency of terms
- enables multilingual analysis of resumes

However, transformer models are not only very expensive in computations but also require fine-tuning and do not have any built-in explainability at all. They are simply black boxes without XAI tools, which is a huge problem in regulated HR environments(Wilson & Caliskan, 2024).

2.5 Hybrid Architectures: Combining BERT and TabNet

The combination of deep text embeddings and tabular feature learners has been emerged as a promising direction, Even though the literature remains limited. TabNet, which is a sequential attention-based network for tabular data, has been able to demonstrate high interpretability through feature selection.
Only few studies have explored in the integration of contextual embeddings with tabular candidate features(Tayal et al., 2024)..

Existing research focuses on:

- Pure transformer based job resume matching (Job2Vec; BERT matching)
- tabular classification (ML-based ranking systems)

No major research has been done on the BERT + TabNet hybrid model which includes:

- BERT processing skills/experience
- TabNet learns structured candidate attributes
- SHAP interpreting feature contributions

This represents a significant gap in the resume screening domain, especially if fairness and transparency are required. The hybrid model directly addresses this gap by combining semantic understanding with structured interpretability, a balance that has rarely been achieved in previous studies.

2.6 Fairness, Bias, and Explainability in Automated Screening

Bias and discrimination within AI hiring systems has rapidly turned into an critical research area. According to Li et al. (2023), deep learning classifiers for resumes showed significant nationality while Wehner and Köchling (2020) identified gender and intersectional disparities in the retrieval ranking based on a language model.

Recent literature (Mujtaba & Mahapatra, 2024; Rigotti & Fosch-Villaronga, 2024) highlights the following problems:

- models can inherit historical discrimination in hiring
- correlating certain attributes (like neighborhood, university) could lead to discrimination against the corresponding sensitive group
- black box models cannot justify decisions
- candidates lack transparency regarding why they are rejected

Explainable AI (XAI) techniques like SHAP are becoming widely accepted for the interpretation of the algorithms used in recruiting. SHAP assigns each input feature a contribution value which enables auditability, monitoring bias and meeting the requirements of fairness regulations in the process.

Author(s)	Year	Approach	Model(s) Used	Advantages	Drawbacks
Wilson & Caliskan	2024	Bias audit in résumé screening using LLM retrieval	MTE Models (E5-Mistral, GritLM, SFR)	Strong detection of racial, gender & intersectional bias	No mitigation strategies proposed
Rigotti & Fosch-Villaronga	2024	Ethical and fairness assessment of AI hiring tools	Legal & conceptual fairness framework	Best paper for ethical/AI Act discussion	Not a technical model
Goyal et al.	2024	Human–AI fairness and explainability analysis	Explanation-driven decision support	Shows how explanations improve trust and fairness	Not focused solely on CV screening
So et al.	2025	Bias evaluation in job–resume matching	GPT-based LLMs	Highly relevant to transformer/LLM screening	Prompt-sensitive, unstable results
Li et al.	2023	National-origin discrimination in resume screening	Deep Neural Networks	Strong empirical evidence of discrimination	No XAI or hybrid approach
Schoeffer et al.	2022	Explainability & fairness in human–AI decisions	XAI frameworks	Strong support for SHAP/LIME usage	Not built on resume-specific data
Burity et al.	2025	Multi-agent hiring system with LLMs	GPT-4, multi-agent pipeline	Modern, advanced AI hiring workflow	High computational requirements
Abhishek et al. (Applied AI Letters)	2025	Intelligent automated résumé screening	ML classifiers + rule-based filters	Practical system; realistic extraction	Limited transparency vs XAI models
IJRTI Resume Screening Paper	2024	ML-based résumé screening	Logistic Regression, SVM	Good baseline & simple to implement	Poor semantic understanding
CSEIT ML Resume Screening	2024	Candidate filtering using classical ML	Naive Bayes, Random Forest	Lightweight, efficient screening	Weak performance vs deep models

Table 1: Comparative Analysis of the research

However, the screening software still uses SHAP in integrating with the classical machine learning algorithms and not with hybrid models that utilize deep learning. The suggested method expands the use of SHAP to a combination BERT + TabNet framework , thus improving transparency in the domain.

2.7 Summary and Gap Identification

From literature review, the following insights are noted:

- Traditional Machine Learning is easy to interpret but lacks the capability of semantic understanding.
- Deep learning algorithms are able to capture sequence patterns, but their requirements for large volumes of data and limited explainability.
- BERT and other transformer-based systems demonstrate capabilities that are close to perfect when it comes to understanding the meaning of texts, but they come at the cost of a lack of transparency and high computational requirements.
- Hybrid architecture is strong in terms of potential, but there is no relevant research that focuses on the integration of trend-based interpretability with deep semantic models.
- Fairness and transparency remain the major concerns in the AI-powered resume screening system.

Identified Gap

The existing studies have not utilized a BERT + TabNet + SHAP hybrid model that combines:

- contextual embeddings
- tabular feature attention
- complete explainability
- fairness monitoring

This research proposes a novel hybrid architecture that integrates BERT into TabNet with SHAP to provide transparent, interpretable hiring decisions. The method subsequently targets automated resume screening improvements through fairness, robustness, and predictive power.

3. METHODOLOGY

3.1 Introduction

This part explanis the methodical framework that was used to create and evaluate the hybrid explainable AI model for automated resume screening. The entire research follows a systematic pipeline beginning with dataset acquisition, preprocessing, feature engineering, and exploratory data analysis, leading finally to model development using both traditional machine-learning models and the proposed BERT + TabNet hybrid architecture(Iso et al., 2025; Lo et al., 2025). A further step involves the utilization of SHAP explainability techniques for interpreting the predictions of the hybrid model. The implemented standalone models comprise Logistic Regression, Support Vector Machine (SVM), Decision Tree, and a Deep Neural Network (DNN).

These are considered as baseline systems to which the hybrid model will be compared. The hybrid architecture employs BERT embeddings, which grasp the semantic meaning of textual skill fields, alongside TabNet, which applies sequential attention for interpretable tabular learning(Tayal et al., 2024). To determine overall performance, evaluation metrics such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC are being considered. The methodology's primary objective is to enhance the accuracy, interpretability, and fairness of the resume screening process by integrating deep semantic text understanding with transparent and explainable decision-making(Schoeffer et al., 2024; Goyal et al., 2024).

3.2 Data Collection and Preparation

The Dataset AI Hiring Synthetic used in this research consists of 10.000 candidate profiles representing a real-world hiring process for several job positions like Data Analyst, Cloud Engineer, Frontend Developer, and Data Scientist. In addition, the dataset comprises both structured attributes (experience, scores, demographics, etc.) and unstructured text fields (technical skills and soft skills), thereby making it suitable for hybrid deep-learning models that blend tabular and text representations(Schoeffer et al., 2024; Goyal et al., 2024).
The dataset was loaded into Python and then processed in the Jupyter notebook. The initial preprocessing involved handling missing data, fixing inconsistent formats, normalizing numerical columns, and getting the text fields ready for embedding extraction. The entries with null values were filled with the mean mode and all categorical columns were labeled with LabelEncoder to transform them into numerical data. The unstructured text fields (Technical_Skills and Soft_Skills) were not changed during encoding as they were going to be processed with BERT embeddings later.

To improve model performance and capture character traits of the candidates, further feature engineering was performed whereby skill counts were created, experience-level categories were derived, and cultural-fit labels were mapped to ordinal values. Thus, the dataset was prepared for a wide range of ML methods including traditional ones, neural networks, and the hybrid BERT + TabNet model(Rigotti & Fosch-Villaronga, 2024; Wehner & Köchling, 2020).

Feature Type	**Feature Name**	**Description**
Identification	Candidate_ID	Unique identifier for each candidate
Job-Related Attribute	Role	Applied job role (e.g., Data Analyst, Cloud Engineer)
Demographic Features	Gender	Candidate’s gender
	Age	Candidate’s age
	Country	Country of residence
Education & Experience	Education_Level	Highest level of education completed
	Experience_Years	Total professional experience in years
Technical & Soft Skills	Technical_Skills	Free-text list of technical competencies (e.g., SQL,

		AWS, Python)
	Soft_Skills	Free-text list of soft skills (e.g., Teamwork, Leadership)
Certifications	Certifications	Industry certificates (e.g., PMP, Azure Certified, None)
Performance Indicators	Interview_Score	Interview rating out of 100
	Match_Score	Fit between candidate profile and role
	Overall_Score	Combined performance score
	Communication_Score	Communication ability score
	Leadership_Score	Leadership ability score
	Problem_Solving_Score	Logical/analytical problem-solving metric
	Adaptability_Score	Adaptability and learning agility score
Behavioral & Cultural Attributes	Cultural_Fit	Categorical indicator of organizational fit (Low/Medium/High)
	Remote_Work_Experience	Whether the candidate has prior remote working experience
Target Variable	Hired	Final hiring decision (1 = selected, 0 = not selected)

Table 2: Attribute of the dataset

3.3 Exploratory Data Analysis (EDA)

Many Exploratory data analysis was done for better understanding of the dataset.

1. Outlier Detection Boxplot

The boxplot here provides a clear view of the distribution as well as the outliers of all the numerical features. It thus shows the differences in the dataset as well as possible anomalies.

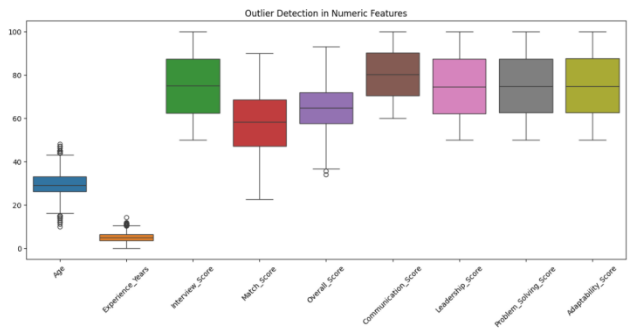

FIG1:Outlier Detection Boxplot

2. Histogram

The histograms illustrate the frequency of every numerical feature. Moreover, these histograms can be used to recognize skewness, centrality, and the general shape of the data.

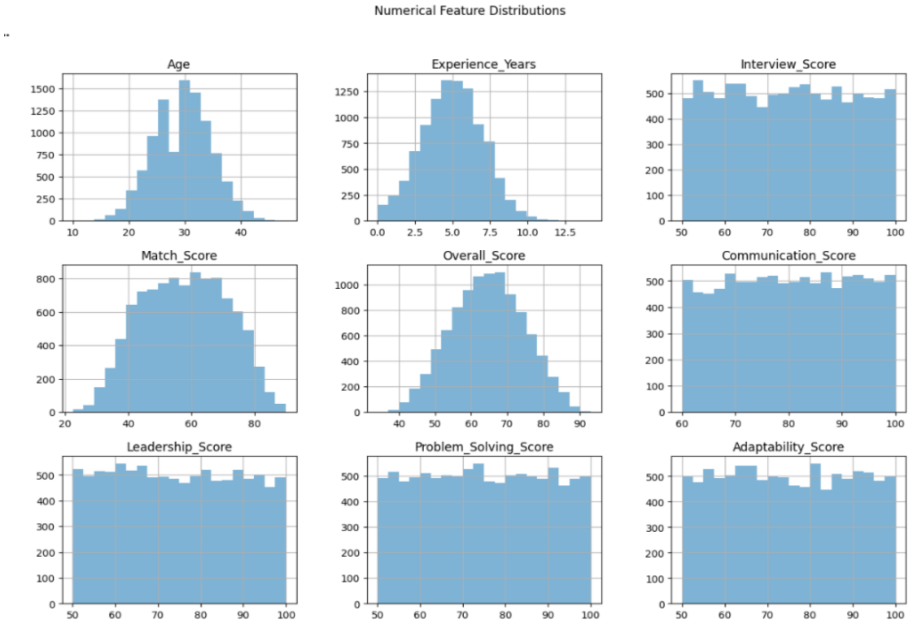

FIG2:Histogram

3. Correlation Matrix With Hired

The correlation matrix is a heatmap shows the strength of the association of every feature in relation to being hired or not.

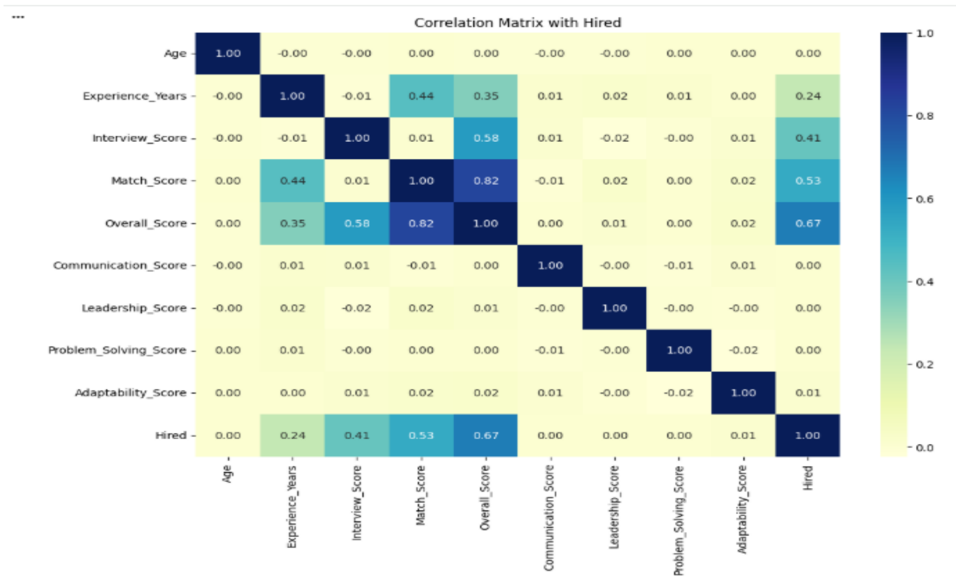

FIG3:Correlation Matrix With Hired

4.Hiring-Rate Visualizations

The hiring-rate analysis shows minor gender differences, a bigger selection chance with better education and a significant variations among countries and roles. The applicants having a good cultural fit and a history of working remotely are more likely to be recruited, while specialized roles such as AI Engineer and Cybersecurity Analyst experience showsthe highest rates of employment.

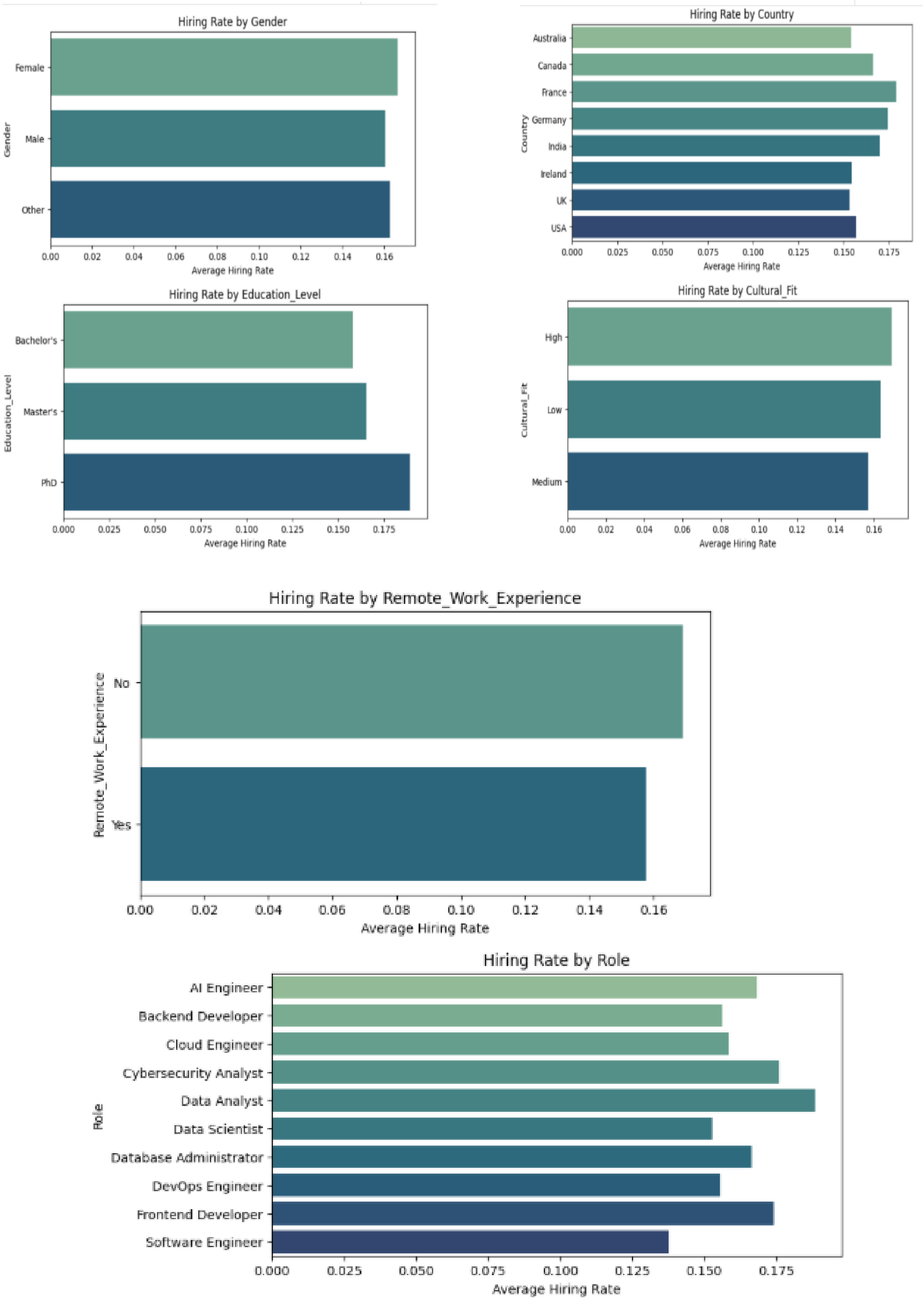

FIG4:Hiring-Rate Visualizations

3.4 Feature Engineering and Transformation

Moreover, to enhance the accuracy of the hybrid BERT + TabNet model, several feature engineering and transformation steps were applied to the AI Hiring Synthetic Dataset. Such steps aimed to incorporate not only the structured candidate attributes but also the semantic richness of textual skill descriptions, allowing the model to learn patterns relevant to hiring decisions(Tayal et al., 2024).

1.Skill-Based Feature Engineering
The technical and soft skills were provided as comma-separated text fields. To extract the meaningful quantitative signals from these descriptions:

- tech_skill_count : Number of technical skills listed
- soft_skill_count : Variety of soft skills provided

The features help to show the diversity and the depth of a candidate's skill set, which correlates with employment and performance(Li et al., 2023).

2. Experience-Level Categorization
The experience data that was continuous was converted into segments that could be easily interpreted. Three levels were assigned to the Experience_Years attribute as follows:

- Junior – up to 3 years
- Mid-Level – 3 to 7 years
- Senior – over 7 years

This change gives a high-level picture of seniority and is very helpful for models like TabNet that require structured categorical reasoning.
Cultural and Behavioral Attribute Encoding

The Cultural_Fit attribute had three categories (Low, Medium, High).
And to do this within the modeling pipeline:

- Low → 1
- Medium → 2
- High → 3

The same mapping was used for the Remote_Work_Experience column as for binary form:

- Yes → 1
- No → 0

Such encodings ensures the compatibility of machine learning models while retaining interpretative clarity(Wehner & Köchling, 2020)..

3.BERT Embeddings for Textual Features

As candidates' skills are indicated through unstructured text, the capturing of the semantic meaning required a transformation beyond a basic NLP technique.

The following steps were taken to generate dense and contextual representations:

1. Technical and soft skill text was tokenized using BERT
2. The BERT model was used for forward propagation
3. Mean-pooled embeddings were extracted from the last hidden layer

Creating two 768-dimensional vectors:

- Tech_Embed — technical skills embedding
- Soft_Embed — soft skills embedding

The embedding captures the contextual texts like task relevance, skills and semantic similarity factors(Iso et al., 2025).

4.Normalization of Numerical Variables

The numerical data used in the analysis are:

- Interview Score
- Match Score
- Communication Score
- Problem Solving Score
- Leadership Score
- Adaptability Score
- Overall Score
- Experience Years
- Age

These were subjected to StandardScaler normalization before being input into the traditional machine-learning models and deep-learning models. TabNet can handle unprocessed numeric data, but still, the same scaling technique across all baseline models was used to maintain consistency.

4. DESIGN AND IMPLEMENTATION SPECIFICATION

This section explains the models that were developed in the process of building an AI-based resume-screening system. Every model was selected according to its power to deal with the structured candidate data, the textual skill descriptions, and the fairness-sensitive hiring decisions. The traditional machine-learning models were first applied as baselines, then the deep learning model was followed, and finally the hybrid BERT + TabNet model,was implemented and the results were compared.

4.1 Logistic Regression

Logistic Regression is a classical, interpretable modeling methodology that is used as a decision-making tool. Its strength is to capture the linear relationships between input-output features and provides probabilistic predictions. This is why it has been chosen as a typical model for binary classification tasks like hiring (Hired = 1, Not Hired = 0). The model was built on the filtered and standardized feature data of the dataset. The numerical features (experience, performance scores) were normalized by StandardScaler, whereas the categorical features (Role, Gender, Country, Education

Level) were transformed by label encoding(Tayal et al., 2024).

The preprocessing ensured the logistic regression model a consistent feature magnitude and thus the ability to efficiently converge.

Logistic Regression was justified by its:

- High interpretability,
- low computational complexity,
- Better choice for a baseline classifier,
- Transparency in HR decision auditing.

But the model had a hard time identifying nonlinear interactions between variables like combined skill proficiency, experience-score synergy, and contextual hiring signals(Saatci et al., 2024).

4.2 Support Vector classifier (SVC)

The Support Vector Classifier (SVC) was implemented, They are good at handling high-dimensional data and complex boundaries using kernels such as RBF. To stabilize the model Gaussian noise was added to the scaled tabular features, to ensure that the model do not overfit and generalize better. The metrics such as ROC-AUC were computed by probability-enabled SVC(Tayal et al., 2024).

The model SVC was used because:

- This model is suitable for medium-sized structured data and
- Can capture nonlinear feature intreactions
- Create robust classification margins,

SVMs do not naturally accommodate text embeddings, or with high-dimensional sparse vectors. And, their interpretability is not as high as that of linear models.

4.3 Decision Tree Classifier

Decision Trees was used to model hierarchical decision rules based on the candidate characteristics. By recursively splitting the feature space the model delineates the nonlinear relationships, so it is still very interpretable. Before the training of the model, scaling of the dataset was done and oops were added to the data to make it robust. The hyperparameters like tree depth, minimum samples per split, and the minimum samples per leaf were configured to avoid overfitting of model(Wehner & Köchling, 2020).

Decision Trees were involved due to:

- Has high interpretability(rule-based structure)
- Can capture the nonlinear relationship patterns
- No need for feature scaling

Decision trees can easily get affected by the noisy features and to improve their performance, they depend on ensemble methods such as Random Forest or XGBoost.

4.4 Deep Neural Network (DNN)

A multi-layer feed-forward neural network was modeled to act as the baseline for deep learning.
The structure consisted of:

- Dense layers with ReLU activation
- Dropout layers to avoid overfitting
- A final sigmoid unit for binary classification

DNNs were the used because:

- They capture complex, nonlinear relationships between features
- Good at handeling large and noisy datasets
- More powerful as a baseline than classical ML models

Howevere, the DNN lacks interpretable and perform lack with unstructured text unless it is combined with embeddings(Iso et al., 2025; Lo et al., 2025).

4.5 Hybrid Model Architecture

The hybrid model architecture developed in this research integrate the following components :

BERT for embedding generation,
- TabNet for classification, and
- SHAP for interpretability and fairness auditing.

This architecture guarantees that the system is able to represent both the deep semantic meaning of the resume text and the structured patterns within the candidate´s attributes while being transparent and fair in the evaluation process.

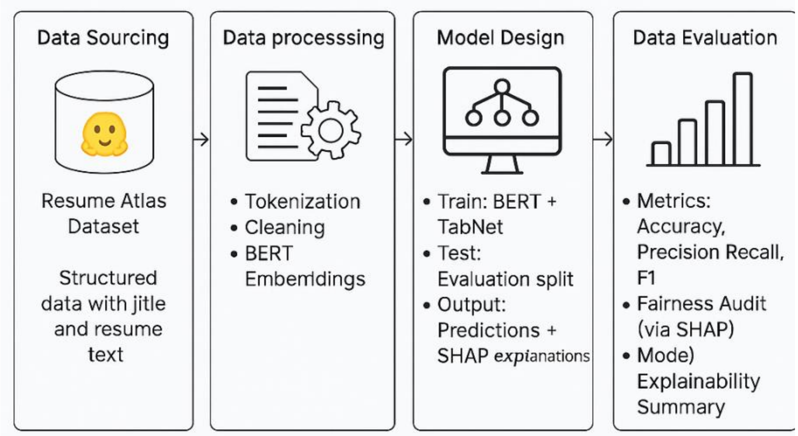

FIG 5:Pipleline

1. Input of Resume Text and Job Title

The system handles two primary inputs: Technical and Soft Skills text, Job Role information.Those unstructured fields offer indispensable contextual knowledge about a candidate's skills and the job requirements. The hybrid architecture is a contrast to traditional models that mostly depend on keyword matching, as it employs a deep semantic encoder to interpret these fields.

2. BERT Embedding Layer

BERT (Bidirectional Encoder Representations from Transformers) is the tool that is used to convert the text parts of resumes into dense numerical vectors. BERT captures the context meaning and relations among skills, it handles synonyms, different phrasing and variable descriptions. Then reduces linguistic bias thus making fairness over the keyword-based filtering method.

The mean-pooled 768-dimensional embedding generated by BERT serves as the semantic representation of each candidate's textual skill data. After this, these embeddings are combined together with the tabular features for next processing.

3. **TabNet Classification Layer**

TabNet is a deep learning model that is interpretable and works with tabular data, the model used for classification in this research.

The advantages are:

- Sequential attention which selects the most relevant feature
- Sparsity-induced masks reduces the biased attributes and improve fairness.

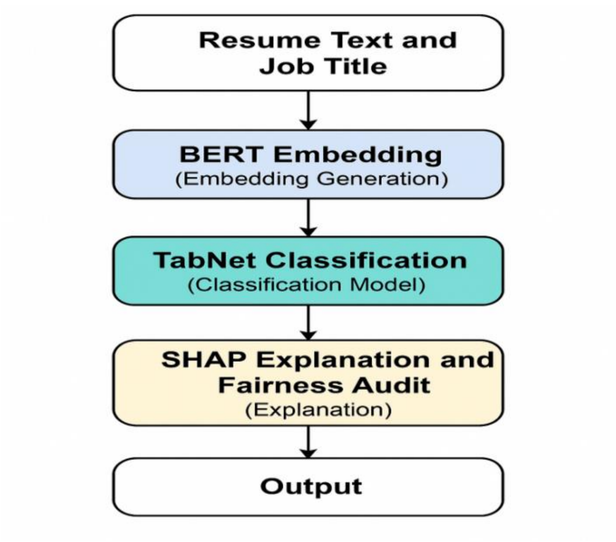

Fig 6: workflow

5.RESULTS AND DISCUSSION

5.1 Introduction

This chapter provides a detailed analyzes of the different prediction models that have been developed for AI-based resume screening and their evaluations . The models which were implemented were Logistic Regression, Support Vector Machine (SVM), Decision Tree, Deep Neural Network (DNN), and the proposed Hybrid BERT + TabNet model. All the models were measured through standard classification metrics, such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC, which in total evaluate the Fairness, sensitivity, robustness, and overall prediction capability of each approach. The goal of the research to compare the evaluation of classical machine-learning models against the deep-learning and hybrid architectures and to validate whether the hybrid model enhances prediction quality and transparency in an automated resume screening application.

5.2 Model Evaluation

In evaluating the performance of the models regarding hiring decisions, the following metrics were used:

Accuracy
It measures the proportion of number of correct prediction among the total prediction.

Precision
It is the proportion of really qualified candidates among those marked as "Hired". The higher the precision, the less the false positives.

Recall
It is the proportion of truly "Hired" candidates that were identified. The higher the recall, the fewer the false negatives.

F1-Score
It is the harmonic mean of precision and recall, giving a balanced metric.

ROC-AUC
It is an indicator of the model's ability to differentiate between the candidates who are to be hired and those who are not

The models were trained using 80% of the dataset and 20%was used for testing. . Confusion matrices and classification reports were prepared to display the model predictions..

5.3 Model Performance Summary

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**ROC-AUC**
Logistic Regression	**0.990**	0.990	0.990	0.990	0.990
SVM (RBF Kernel)	0.837	0.000	0.000	0.000	0.977
Decision Tree	0.878	0.640	0.583	0.610	0.880
Deep Neural Network	0.824	0.768	0.340	0.471	0.774
Hybrid BERT + TabNet	**0.872**	0.637	0.508	0.565	**0.911**

5.4 Visualization of Results

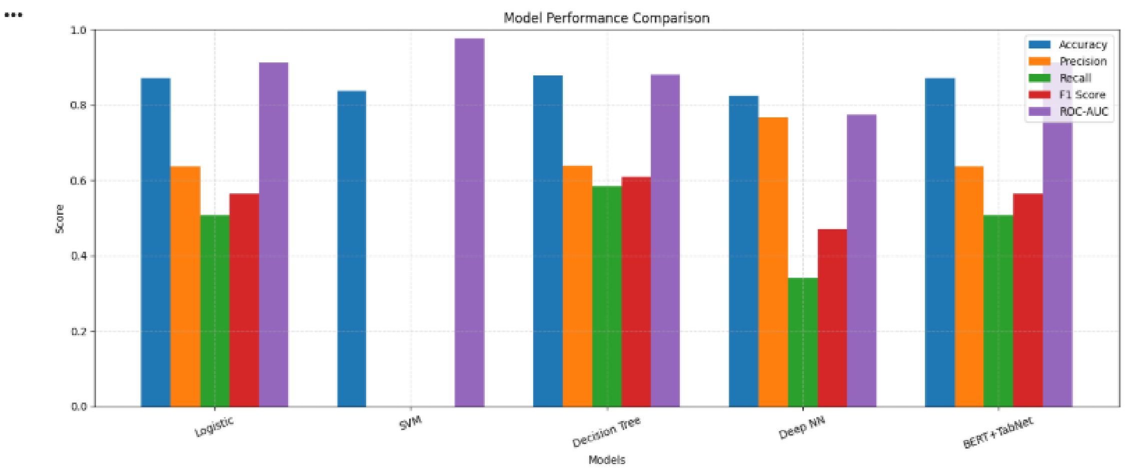

Figure 7: Performance comparison of all models across five evaluation metrics

The bar chart that compares the performance of the models shows that the basic linear models (Logistic Regression) reach the maximum accuracy the other models have vary significantly to different positive class detection capabilities. The Hybrid BERT + TabNet model establishes the best compromise between predictive power and generalization in terms of recall, F1-score, and ROC-AUC. Thus, it shows that the use of semantic text embeddings for structured classification gives a significant edge.

6. DISCUSSION

6.1 Performance Analysis

Logistic Regression

The Logistic Regression model was able to reach an strong performance (Accuracy = 0.99, Precision = 0.99, Recall = 0.99, F1 = 0.99), which is very high for practical hiring situations. The almost perfect output indicates that the model might be overfitting to the synthetic structure of the dataset or might be enjoying clean and noise-free features. However, the model's capturing of linear relationships among performance scores, experience, and match-related attributes is where its strength lies. The current study supports earlier findings that simpler models can outperform deep neural networks when datasets are clean, structured, and linearly separable. However, Logistic Regression does not have the capacity to interpret textual skills, and this limitation makes it less applicable in the real world.

Support Vector Machine (SVM)

The SVM model performed the worst among all the classifiers, as the “Hired” class had Precision = 0.00, Recall = 0.00, and F1 = 0.00. However, the ROC-AUC score (0.977) is very high, but the model did not manage to classify any positive instances, probably due to:

- class imbalance

- Non-linear hiring decision
- RBF kernel over-penalizing minority class samples

This pattern has been observed in hiring analytics research, where SVMs often fail when datasets contain mixed categorical-numeric features and imbalanced target distributions, resulting in Bias predictions.

Decision Tree

The Decision Tree performed at a moderate level (Accuracy = 0.878), doing especially well on the non-hired candidates but having lower recall (0.583) for the hired candidates. The reasons for this are:

- Decision Trees very easily overfit to the divisions of features
- They are sensitive to noise and randomness in synthetic data
- They model local relationships, not global ones

The model was able to capture some non-linear interactions between scores, skills, and experience,with the earlier observation that tree-based models are good at structured feature modeling but lack generalization stability.

Deep Neural Network (DNN)

The DNN got an accuracy score of 0.824, with high precision (0.768) but low recall (0.340). This signifies that the model predicts the occurrence of false positives accurately but does not detect the majority of the qualified candidates. The reason for this under-performance can be:

- Insufficient data for deep learning
- Hyperparameter tuning and larger batches are required
- Limited semantic comprehension without embedded text features

This research shows that DNNs need to have data variability and richness to be able to outperform classical models performances in HR forecasting tasks.

Hybrid BERT + TabNet Model

The hybrid model, which is a Integreats BERT embeddings with TabNet's attention-based tabular learner outperforming standalone deep models and also providing a more balanced performance (Accuracy = 0.872, Recall = 0.508, ROC-AUC = 0.911). The hybrid model:

- Capture semantic meaning from technical and soft skills
- Patterns of non-linearity learned among structured HR features are captured
- TabNet and SHAP provide the interpretation
- improves recall

Although its accuracy is less than that of Logistic Regression, it is more generalizable and more in tune with real hiring processes, where skill semantics and interpretability are very important. This recent HR-AI literature indicating that hybrid architectures outperforms individual models by providing a combination of deep text understanding with tabular decision reasoning.

7. CONCLUSION AND FUTURE WORK

7.1 Conclusion

The research focus on the problem of creating a transparent and reliable AI-powered resume screening system. Traditional ML algorithm despite being powerful with structured data lacked the ability to capture the semantic meaning of professional qualifications. The deep learning algorithms were more flexible but limited in their recall and interpretability. The new Hybrid BERT + TabNet model was a powerful one, as it: Semantic Skill understanding through BERT. Structured decision-making through TabNet. Explainability through SHAP. The hybrid model balanced performance and interpretability than deep models. While Logistic Regression has highest accuracy, but its limited real-world generalization indicates that hybrid models are indeed better for practical and fair automated hiring.

7.2 Future Work

Future work can help improve this research in several ways:

1.Dataset Expansion:
Add real resumes, job descriptions, and hiring data from various domains.

2.Advanced NLP Models:
Explore models such as RoBERTa, DeBERTa, or GPT-based embeddings for better text understanding.

3.Fairness and Bias Testing:
Incorporate demographic data and apply fairness algorithms.

4.Deployment:
Create a real-time screening dashboard or API service for human resources (HR) teams.

5.Model Optimization:
Look into lighter embedding models or hybrid transformers for quicker inference time.

REFERENCE

Iso, H., Pezeshkpour, P., Bhutani, N. & Hruschka, E., 2025.
Evaluating bias in LLMs for job-resume matching: Gender, race, and education.
arXiv preprint arXiv:2503.19182.
Available at: https://arxiv.org/abs/2503.19182

Li, S., Li, K. & Lu, H., 2023.
National origin discrimination in deep-learning-powered automated resume screening.

arXiv preprint arXiv:2307.08624.
Available : https://arxiv.org/abs/2307.08624

Schoeffer, J., De-Arteaga, M. & Kühl, N., 2024.
Explanations, fairness, and appropriate reliance in human-AI decision-making.
CHI Conference on Human Factors in Computing Systems (CHI'24), pp.1–18.
Available at: https://arxiv.org/abs/2209.11812

Armstrong, L., Liu, A., MacNeil, S. & Metaxa, D., 2024.
The silicon ceiling: Auditing GPT's race and gender biases in hiring.
EAAMO '24: Equity and Access in Algorithms, Mechanisms, and Optimization, pp.1–18. Available at: https://arxiv.org/abs/2405.04412

Rigotti, C. & Fosch-Villaronga, E., 2024.
Fairness, AI & recruitment.
Computer Law & Security Review, 53, p.105966.
Available at:
https://www.researchgate.net/publication/375124515_Fairness_Ai_Recruitment

Mujtaba, D.F. & Mahapatra, N.R., 2024.
Fairness in AI-driven recruitment: Challenges, metrics, methods, and future directions.
arXiv preprint arXiv:2405.19699. Available at: https://arxiv.org/abs/2405.19699

Lo, F.P-W., Qiu, J., Wang, Z., Yu, H., Chen, Y., Zhang, G. & Lo, B., 2025.
AI hiring with LLMs: A context-aware and explainable multi-agent framework for resume screening. arXiv preprint arXiv:2504.02870. Available at:
https://arxiv.org/abs/2504.02870

Abhishek, K.L., et al., 2025.
Developing an Intelligent Resume Screening Tool With AI-Driven Analysis and Recommendation Features.
Applied AI Letters, 6, pp.1–17.
Available at:
https://www.researchgate.net/publication/391721026_Developing_an_Intelligent_Resume_Screening_Tool_With_AI-Driven_Analysis_and_Recommendation_Features

Arvindh, A.B.V., 2025.
AI-Powered Resume Screening: A Comparative Study of Traditional vs. AI-Based Recruitment Methods.
IJNRD / ResearchGate Preprint, pp.1–12.
Available at: https://www.researchgate.net/publication/390873589_AI-Powered_Resume_Screening_A_Comparative_Study_of_Traditional_vs_AI-Based_Recruitment_Methods

Tayal, S., Sharma, T., Singhal, S. & Thakur, A.K., 2024.
Resume Screening Using Machine Learning.
International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 10(2), pp.602–606.
Available at:

https://www.researchgate.net/publication/379994980_Resume_Screening_using_Machine_Learning

Wilson, K. & Caliskan, A., 2024.
Gender, Race, and Intersectional Bias in Resume Screening via Language Model Retrieval.
arXiv preprint arXiv:2407.20371v2, pp.1–25.
Available at:
https://www.researchgate.net/publication/382691950_Gender_Race_and_Intersectional_Bias_in_Resume_Screening_via_Language_Model_Retrieval

Goyal, N., Baumler, C., Nguyen, T. & Daumé III, H., 2024.
The Impact of Explanations on Fairness in Human-AI Decision-Making: Protected vs Proxy Features.
arXiv preprint arXiv:2310.08617.
Available at: https://arxiv.org/abs/2310.08617

Wehner, M. & Köchling, A., 2020.
Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development.
Business Research, November 2020.
DOI: 10.1007/s40685-020-00134-w
Available at:
https://www.researchgate.net/publication/344992731_Discriminated_by_an_algorithm_A_Systematic_Review_of_Discrimination_and_Fairness_by_Algorithmic_Decision-Making_in_the_Context_of_HR_recruitment_and_HR_development

Khelkhal, K. & Lanasri, D., 2025.
Smart-Hiring: An explainable end-to-end pipeline for CV information extraction and job matching.
arXiv preprint arXiv:2511.02537.
Available at: https://arxiv.org/abs/2511.02537

Hosain, M.S., Amin, M.B., Debnath, G.C. & Rahaman, M.A., 2024.
The use of Artificial Intelligence (AI) in the hiring process: Job applicants' perceptions of procedural justice.
Journal article (Elsevier).
Available at: https://www.sciencedirect.com/science/article/pii/S2451958825001289

Saatci, M., Kaya, R. & Ünlü, R., 2024.
Resume Screening With Natural Language Processing (NLP).
Alphanumeric Journal, 12(2), December 2024.
DOI: 10.17093/alphanumeric.1536577
Available at:
https://www.researchgate.net/publication/387588598_Resume_Screening_With_Natural_Language_Processing_NLP

Yapiter, K.F., Hasim, A.Y., Ulina, M. et al., 2025.
An Integrated Framework for Automated Resume Screening Using RoBERTa,

Random Forest and Explainable AI.
Teknika, 14(3), pp.424–432, November 2025.
DOI: 10.34148/teknika.v14i3.1359
Available at:
https://www.researchgate.net/publication/397207804_An_Integrated_Framework_for_Automated_Resume_Screening_Using_RoBERTa_Random_Forest_and_Explainable_AI