Bias Aware Resume Screening

A. Hemanth Datta[1*]; A. Hrushi Jiyyan[2]; G. Nagi Reddy[3];
Dr. K. Sreekala[4]; Manas Kumar Rath[5]

[1,2]Under Graduate Student, [3,4,5]Assisstant Professor Department of Computer Science & Engineering,
Mahatma Gandhi Institute of Technology ,Hyderabad ,Telangana ,India

Corresponding Author: A. Hemanth Datta[1*]

Publication Date: 2026/06/01

Abstract: Automated resume screening systems are widely used to manage large volumes of job applications; however, many existing approaches rely on rigid keyword matching or accuracy-focused machine learning models that fail to capture contextual skill relevance and may introduce algorithmic bias. This paper proposes a bias-aware automated resume screening framework that utilizes Natural Language Processing (NLP) techniques to analyze contextual relationships between candidate skills, experience, and job requirements. The system mitigates bias by identifying and neutralizing sensitive textual cues related to protected attributes such as gender or ethnicity. To evaluate robustness, screening outcomes are compared before and after the removal of noise and bias-related information. A Decision Stability Metric is introduced to measure the consistency of candidate rankings under controlled textual perturbations. Additionally, confidence calibration techniques are applied to ensure that the predicted probabilities accurately reflect the reliability of model decisions. The framework also improves transparency in automated hiring by enabling more consistent and interpretable candidate evaluation.

Keywords: *Resume Screening, Natural Language Processing (NLP), Algorithmic Bias Detection, Decision Stability Analysis, Confidence Calibration, Explainable Artificial Intelligence.*

How to Cite: A. Hemanth Datta; A. Hrushi Jiyyan; G. Nagi Reddy; Dr. K. Sreekala; Manas Kumar Rath (2026) Bias Aware Resume Screening. *International Journal of Innovative Science and Research Technology*, 11(5), 2547-2553.
https://doi.org/10.38124/ijisrt/26may1138

I. INTRODUCTION

The rapid digitalization of recruitment platforms has significantly increased the number of job applications received by organizations. Human resource departments often struggle to manually evaluate thousands of resumes efficiently while maintaining fairness and consistency in candidate selection. Automated resume screening systems have therefore emerged as important tools for assisting recruiters in filtering and ranking candidates according to job requirements.

Traditional automated screening approaches mainly rely on keyword matching techniques or conventional machine learning models that prioritize accuracy over contextual understanding. While these methods can quickly identify specific keywords in resumes, they often fail to capture the deeper semantic relationships between candidate skills, professional experience, and job requirements. Consequently, highly qualified candidates may be overlooked simply because their resumes use different terminology than the job description.

Another major concern in automated hiring systems is the presence of algorithmic bias. Machine learning models trained on historical hiring data may unintentionally inherit biases related to gender, ethnicity, educational background, or other protected attributes.

Such biases can lead to unfair candidate evaluations and discriminatory outcomes. Moreover, many automated decision systems operate as black boxes, making it difficult for recruiters to understand the reasoning behind screening decisions

Recent advancements in Natural Language Processing (NLP) offer promising solutions to these challenges. NLP techniques enable systems to analyze textual data more intelligently by capturing contextual relationships between words, phrases, and professional attributes within resumes. By leveraging semantic representations, NLP-based models can perform more meaningful candidate–job matching and reduce reliance on rigid keyword- based filtering.

In addition to contextual understanding, modern recruitment systems must also ensure fairness, transparency, and robustness. Bias mitigation strategies aim to detect and neutralize sensitive attributes that may unfairly influence automated decision-making. Furthermore, evaluating the reliability of model predictions under different conditions is

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1138

essential to ensure stable and trustworthy screening outcomes.

This paper proposes a Bias-Aware Resume Screening System that integrates NLP techniques with fairness-aware preprocessing and robustness evaluation mechanisms. The system detects and removes bias-inducing textual cues from resumes before processing them through a classification model. To evaluate the robustness of the screening process, results are analyzed before and after bias removal and controlled noise perturbations.

Additionally, the system introduces a Decision Stability Metric to measure how consistently candidate rankings remain under small input changes. This helps assess the reliability of automated hiring decisions. The framework also incorporates confidence calibration to ensure that the model's predicted probabilities accurately reflect true decision confidence.

By combining contextual NLP analysis, bias mitigation strategies, stability evaluation, and calibrated confidence estimation, the proposed system aims to support fair, transparent, and reliable AI-driven recruitment processes.

II. LITERATURE SURVEY

Y. Buolamwini and T. Gebru (2025) This paper focuses on auditing automated decision-making systems used in high-stakes domains like hiring and law enforcement. It highlights how such systems can exhibit significant bias against certain demographic groups. The authors emphasize the need for transparency and accountability in AI systems. Their work builds on earlier research exposing racial and gender bias in algorithms. [1]

A. Fabris et al. (2025) This survey provides a comprehensive overview of fairness and bias issues in algorithmic hiring systems. It discusses different types of bias, such as historical and representation bias, and their impact on recruitment decisions. The paper also reviews fairness metrics and mitigation strategies. It serves as a foundational reference for designing ethical hiring models. [2]

M. M. A. M. Sony et al. (2025) This study examines the presence of bias in AI-driven Human Resource Management (HRM) systems. It explores how automated tools can unintentionally reinforce workplace inequalities. The authors analyze organizational risks and ethical concerns associated with biased AI. They also suggest guidelines for implementing fair HR technologies. [3]

E. Ip et al. (2025) This paper investigates how gender bias in AI hiring systems affects applicant perceptions. It studies how candidates respond to decisions made by biased algorithms. The findings reveal that perceived fairness significantly influences trust and acceptance. The paper highlights the importance of fairness for user experience in AI systems. [4]

L. Floridi and M. Taddeo (2025) This work discusses the concept of trustworthy artificial intelligence and its practical implementation. It outlines key principles such as fairness, accountability, transparency, and ethics. The authors propose ways to translate these principles into measurable system guarantees. It bridges the gap between theory and real-world AI deployment. [5]

R. Guidotti et al. (2025) This paper focuses on post- hoc verification methods for black-box AI systems. It introduces techniques to interpret and validate decisions made by complex models. The study is particularly relevant for domains like hiring where explainability is crucial. It emphasizes building trust through interpretable AI. [6]

J. García-González et al. (2024) This research explores the role of explainable AI (XAI) in recruitment systems. It discusses methods to make hiring decisions more transparent and understandable. The authors highlight how explainability can reduce bias and improve fairness. The paper also addresses practical challenges in implementing XAI. [7]

C. Rigotti (2024) This paper examines the intersection of fairness, AI, and recruitment practices. It analyzes ethical and legal implications of using AI in hiring. The study discusses fairness frameworks and their application in real-world systems. It emphasizes the importance of responsible AI adoption in HR. [8]

A. Mehrabi et al. (2024) This work studies the robustness and stability of machine learning models under input perturbations. It highlights how small changes in input data can significantly affect predictions. The findings are important for ensuring reliability in sensitive applications like hiring. The paper also discusses ways to improve model stability. [9]

K. Kallus and A. Zhou (2024) This paper focuses on decision stability and sensitivity analysis in machine learning classifiers. It evaluates how consistent model predictions are when inputs slightly vary. The authors propose methods to measure and improve stability. This is crucial for fair and dependable decision-making systems. [10]

Y. Chen et al. (2023) This study evaluates bias in automated hiring systems using empirical analysis. It identifies discriminatory patterns in AI-based recruitment tools. The authors propose evaluation frameworks to detect and measure bias. Their work contributes to improving fairness in hiring algorithms. [11]

P. Hiremath and S. Deshpande (2023) This paper discusses the use of supervised learning techniques for resume classification. It explores different machine learning models for categorizing resumes. The study highlights the efficiency and accuracy of automated screening systems. It provides a technical foundation for resume-based AI systems. [12]

J. Lee and R. Singh (2023) This research examines ethical risks associated with AI-based resume screening systems. It discusses issues like bias, lack of transparency, and

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1138

privacy concerns. The authors emphasize the need for ethical guidelines and regulations. The paper raises awareness about responsible AI use in hiring. [13]

C. Papakyriakopoulos et al. (2022) This paper focuses on methods for measuring bias in algorithmic hiring systems. It introduces quantitative techniques to assess fairness across different groups. The study highlights challenges in defining and evaluating fairness. It contributes to developing standardized bias measurement frameworks. [14]

T. Zhang et al. (2022) This research explores public perceptions of algorithmic hiring systems. It analyzes how people view fairness, trust, and transparency in AI-driven recruitment. The findings show that trust depends heavily on perceived fairness and explainability. The paper emphasizes the importance of user acceptance in AI adoption. [15]

III. DESIGN OF THE SYSTEM

➢ *Frontend Interface:*
The system provides an interactive web interface that allows recruiters to upload job descriptions and candidate resumes. The interface enables users to view candidate rankings, screening results, and model confidence scores.

➢ *Backend Processing:*
The backend processes resumes using NLP techniques such as tokenization, text normalization, and feature extraction. These processed features are then passed to the machine learning model for candidate ranking.

➢ *Bias Detection Module:*
A preprocessing module scans resume text to detect sensitive attributes such as gender-related terms, personal identifiers, or demographic indicators. These features are removed or neutralized before model evaluation.

➢ *Machine Learning Model:*
The model analyses contextual relationships between resume features and job requirements to produce candidate suitability scores.

➢ *Evaluation Module:*
The system evaluates robustness using a Decision Stability Metric and calibrates prediction confidence using probability calibration techniques.

IV. BLOCK DIAGRAM

The block diagram of the proposed Bias-Aware Resume Screening System illustrates the overall workflow of the system from resume input to candidate evaluation. The process begins when resumes and job descriptions are uploaded through the user interface. The text data then undergoes preprocessing using Natural Language Processing (NLP) techniques, including tokenization, stop- word removal, and normalization to convert raw text into structured information.

Next, the system performs bias detection and mitigation, where sensitive attributes such as gender-related terms or demographic indicators are identified and neutralized to reduce potential bias in screening decisions. The processed text is then passed to the feature extraction module, which captures relevant skills, qualifications, and experience from resumes.

These features are analysed by a machine learning model that evaluates candidate suitability based on job requirements and generates candidate rankings. To ensure reliability, the system also evaluates the robustness of predictions using a Decision Stability Metric and applies confidence calibration to produce reliable probability scores. Finally, the system outputs a ranked list of candidates along with calibrated confidence scores, supporting fair and transparent hiring decisions.

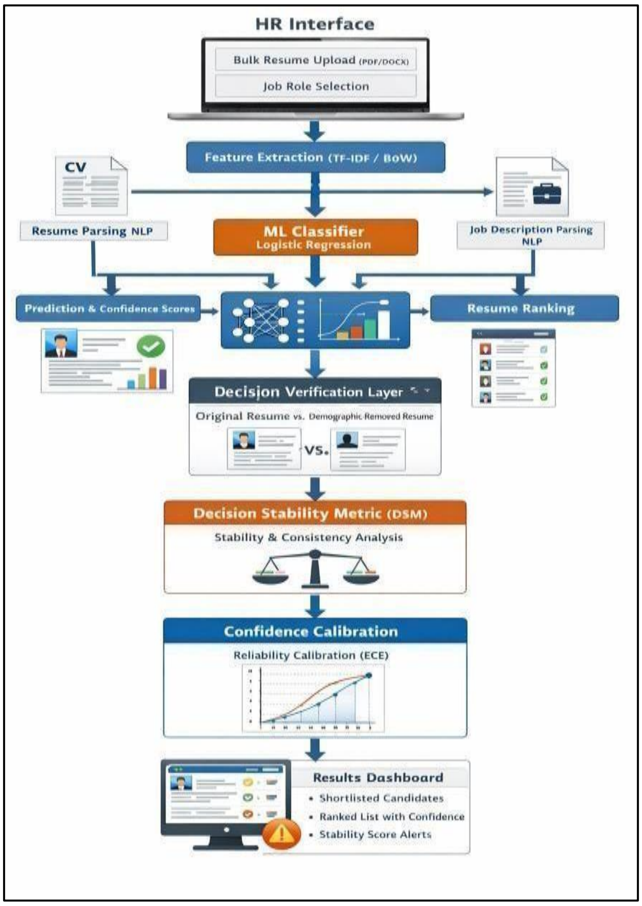

Fig 1 Showing the Block Diagram (Architecture) of Bias Aware Resume Screener

➢ *Flow Diagram of Bias Aware Resume Screening*
The flow chart illustrates the working process of the Bias-Aware Resume Screening System. The process begins when resumes and job descriptions are uploaded into the system. The resumes first undergo text preprocessing.

Next, the system performs bias detection and mitigation by identifying and removing sensitive attributes that may introduce unfair bias. The cleaned data is then passed to the feature extraction module, which identifies important information such as skills, experience, and qualifications.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1138

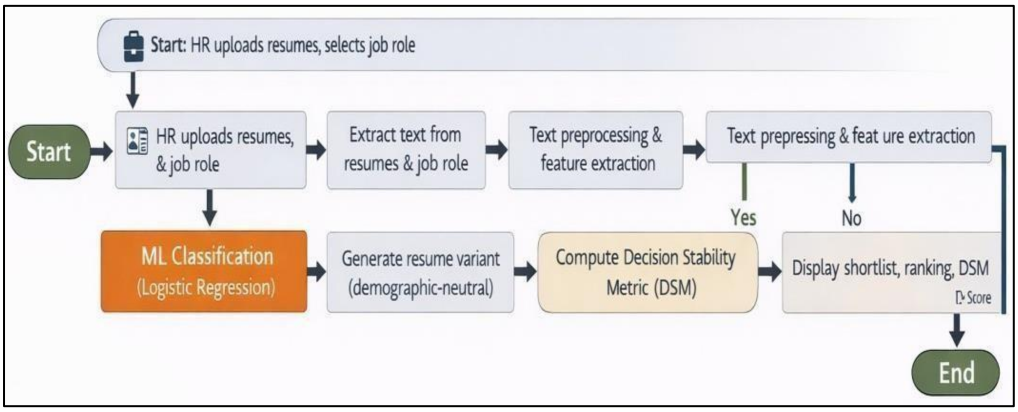

Fig 2 Flow Chart of the Bias Aware Resume Screener

These features are analysed by a machine learning model that evaluates candidate suitability and generates rankings. Finally, the system outputs the ranked list of candidates along with confidence scores, ensuring fair and transparent recruitment decisions.

V. METHODOLOGY

➢ *Resume Input and Data Collection*

The system begins with the collection of candidate resumes through the user interface. Resumes can be uploaded in formats such as PDF or text files. These documents serve as the primary input for the system and are used for further processing and analysis.

➢ *Text Extraction and Preprocessing*

The uploaded resumes undergo text extraction followed by preprocessing using NLP techniques. This stage includes conversion to lowercase, removal of special characters, elimination of numerical values, and normalization of text. The objective is to convert unstructured resume data into a clean and structured format suitable for machine learning processing.

➢ *Feature Extraction using TF-IDF*

After preprocessing, the cleaned text is transformed into numerical features using the Term Frequency– Inverse Document Frequency (TF-IDF) technique. This method captures the importance of words in each resume relative to the dataset, enabling effective representation of candidate skills and experience.

➢ *Resume Classification using Machine Learning*

The extracted features are passed to a trained Logistic Regression model, which classifies each resume into a specific job category. The model also generates probability scores for each class, indicating the confidence of prediction.

➢ *Candidate Ranking and Match Strength Calculation*

Based on the predicted probabilities, the system calculates a match strength score for each resume. This score is normalized into a user-friendly percentage range to improve interpretability. Candidates are ranked according to their relevance to the selected job role, and shortlisting is performed based on classification results.

➢ *Top Prediction Analysis*

To provide deeper insights, the system identifies the top five predicted job roles for each resume along with their corresponding probabilities. This helps recruiters understand alternative role suitability for candidates.

➢ *Bias Detection using Counterfactual Testing*

To ensure fairness, the system performs counterfactual testing by modifying sensitive attributes such as names in the resume text. The model is re-evaluated on the modified input, and any variation in prediction is analyzed. This helps detect whether the model is influenced by bias.

➢ *Bias Score and Fairness Evaluation*

A bias score is calculated by measuring the difference in prediction probabilities before and after modification. If the variation is below a defined threshold, the prediction is considered fair; otherwise, it indicates potential bias. This ensures ethical and unbiased decision-making.

➢ *Result Generation and Visualization*

Finally, the system generates the ranked list of candidates along with match strength scores, fairness indicators, and bias scores. The results are displayed through an interactive interface, including tables, charts, and downloadable reports, enabling transparent and informed hiring decisions.

VI. IMPLEMENTATION

The implementation of the proposed Bias-Aware Resume Screening System focuses on integrating Natural Language Processing (NLP), machine learning, and fairness evaluation techniques into a functional application. The system is implemented using Python and deployed through a Streamlit- based user interface for interactive resume screening.

➢ *User Interface Development*

The system interface is developed using Streamlit to provide an interactive and user-friendly environment. It allows users to upload multiple resumes, select a target job role, and view screening results. The interface displays ranked candidates, match strength scores, fairness indicators, and visual insights, enabling efficient interaction with the system.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1138

➢ *Resume Text Extraction*

The uploaded resumes are processed to extract textual content. For PDF files, the system uses PDF processing libraries to extract text page by page, while text files are directly read and decoded. This step ensures that all input resumes are converted into a consistent textual format for further processing.

➢ *Text Preprocessing*

The extracted text undergoes preprocessing using NLP techniques. This includes converting text to lowercase, removing special characters, eliminating numerical values, and normalizing whitespace. These steps help clean the data and prepare it for feature extraction and analysis.

➢ Feature Extraction using TF-IDF

The cleaned resume text is transformed into numerical vectors using the TF-IDF technique. This representation captures the importance of words in each resume relative to the dataset, allowing the machine learning model to understand and compare candidate profiles effectively.

➢ *Model Training and Classification*

A Logistic Regression model is used for resume classification. The model is trained on labeled resume data to learn patterns associated with different job roles. During execution, the trained model predicts the job category of each resume and provides probability scores indicating prediction confidence.

➢ *Match Strength Calculation and Ranking*

The system calculates a match strength score based on the model's probability output. This score is normalized into a percentage range to improve interpretability. Candidates whose predicted role matches the selected job role are shortlisted, and all candidates are ranked based on their match strength scores.

➢ *Top Prediction Analysis*

To provide additional insights, the system identifies the top five predicted job roles for each resume along with their corresponding probability scores. This helps recruiters understand alternative role suitability and improves decision-making.

➢ *Bias Detection using Counterfactual Testing*

The system incorporates a fairness mechanism by generating counterfactual versions of resumes. Sensitive attributes such as names are modified, and the model is re-evaluated on the modified text. This helps determine whether predictions are influenced by bias.

➢ *Bias Score Computation and Fairness Evaluation*

A bias score is computed by measuring the difference between prediction probabilities before and after modification. If the difference is below a predefined threshold, the prediction is considered fair; otherwise, it indicates potential bias. This ensures ethical and unbiased screening outcomes.

➢ *Result Visualization and Reporting*

Finally, the system presents the results through an interactive interface. It displays candidate rankings, shortlisted profiles, match strength scores, fairness indicators, and graphical visualizations such as charts. The system also provides an option to download results in CSV format for further analysis.

➢ *Key Functionalities:*

- *Resume Upload and Processing*

The system allows users to upload multiple resumes in PDF or text format through an interactive interface. These resumes are automatically processed and converted into textual data for further analysis, enabling efficient handling of large volumes of candidate information.

- *Text Preprocessing using NLP*

The system applies Natural Language Processing (NLP) techniques to clean and normalize resume content. This includes removal of special characters, conversion to lowercase, and elimination of irrelevant information, ensuring that the data is structured and suitable for machine learning processing.

- *Feature Extraction using TF-IDF*

The processed text is transformed into numerical features using the TF-IDF technique. This helps in identifying important words and patterns within resumes, allowing the system to effectively represent candidate skills and qualifications.

- *Resume Classification*

A machine learning model is used to classify resumes into different job categories based on their content. The model predicts the most suitable role for each candidate and provides probability scores indicating prediction confidence.

- *Candidate Ranking and Match Strength Scoring*

The system calculates a match strength score for each resume based on model predictions. This score is normalized into a percentage range and used to rank candidates according to their relevance to the selected job role, improving interpretability for recruiters.

- *Top Prediction Analysis*

The system provides the top five predicted job roles for each resume along with their probability scores. This helps in understanding alternative role suitability and provides deeper insights into candidate profiles.

- *Bias Detection and Fairness Evaluation*

The system incorporates counterfactual testing to detect bias in predictions. By modifying sensitive attributes such as names and re-evaluating the model, it ensures that decisions are not influenced by biased information.

- *Bias Score Calculation*

A bias score is computed by measuring the difference in prediction probabilities before and after modification of

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1138

sensitive attributes. This score helps quantify the level of bias present in the system's decisions.

- *Visualization and Insights*

The system provides visual insights such as charts and metrics to represent candidate distribution, fairness evaluation, and screening outcomes. These visualizations enhance understanding and support data-driven decision-making.

- *Result Export and Reporting*

The system allows users to download screening results in CSV format for further analysis and documentation. This feature improves usability and supports integration with other recruitment processes.

➢ *Software System Configuration*

- *Development Environment*

✓ Programming Language: Python 3.8 or newer
✓ IDE / Text Editor: Visual Studio Code, PyCharm, or Jupyter Notebook
✓ Operating System: Windows / Linux / macOS
✓ Virtual Environment: Recommended to manage project dependencies and maintain library compatibility.

- *Required Python Libraries*

The proposed system uses several Python libraries for implementing Natural Language Processing, machine learning, and data processing tasks:

✓ NLTK / spaCy: Used for Natural Language Processing tasks such as tokenization, stop-word removal, and text normalization during resume preprocessing.
✓ Scikit-learn: Used for implementing machine learning models for candidate classification and ranking based on resume features.
✓ Pandas: Used for data handling and manipulation of resume datasets and extracted features.
✓ NumPy: Provides numerical computation support for feature processing and model operations.Matplotlib /
✓ Seaborn: Used for visualization of results such as model confidence scores and stability analysis.
✓ SQLite3: Used for storing processed resume data, candidate rankings, and system logs.

- *System Requirements Processor: Intel i5 or higher*

✓ RAM: Minimum 8 GB recommended
✓ Storage: At least 10 GB of available disk space

VII. RESULTS AND DISCUSSION

The proposed Bias-Aware Resume Screening System was evaluated to examine its effectiveness in candidate ranking, bias mitigation, and decision stability. Using NLP techniques, the system extracted relevant features such as skills, qualifications, and experience from resumes and matched them with job requirements.

Screening results were compared before and after bias removal to analyse fairness. The removal of sensitive attributes reduced the influence of demographic cues on candidate rankings, resulting in more balanced screening outcomes.The system's robustness was evaluated using the Decision Stability Metric, which showed minimal changes in rankings under small input perturbation.

Table 1Shows the Comparision of Bias Aware Resume Screener with Existing Applications (HireVue,Pymetrics ,Eightfold AI)

Feature	Bias-Aware Resume Screening System	HireVue	Pymetrics	Eightfold AI
Resume Analysis	Contextual NLP-based resume understanding	Limited resume analysis	Behavioral assessment focus	AI-based resume matching
Skill Extraction	Advanced NLP skill extraction	Interview-focused review	Cognitive trait analysis	Keyword matching
Bias Mitigation	Removes sensitive attributes	Limited transparency	Indirect fairness checks	Minimal mitigation
Context Understanding	Captures semantic relationships	Limited context	Behavior-based focus	Keyword-based
Decision Stability	Uses stability metrics	Not supported	Not supported	Not supported
Confidence Calibration	Calibrated match strength	Basic scoring	Behavioral scoring	Suitability scoring
Transparency	Interpretable outputs	Black-box	Limited explainability	Limited explainability
Fairness Evaluation	Counterfactual + bias score	No fairness evaluation	Partial fairness	Limited reporting
Ranking Mechanism	Match strength ranking	Interview-based	Behavior matching	AI ranking
Output Visualization	Charts + export options	Limited visuals	Basic reports	Dashboard analytics

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1138

REFERENCES

- [1]. Y. Buolamwini and T. Gebru, "Auditing automated decision systems in high-stakes domains," IEEE Trans. Technol. Soc., 2025.
- [2]. A. Fabris et al., "Fairness and bias in algorithmic hiring," ACM Computing Surveys, 2025.
- [3]. M. M. A. M. Sony et al., "Bias in AI-driven HRM systems," Journal of Business Research, 2025.
- [4]. E. Ip et al., "Fair AI in hiring: Gender bias effects on applicant perceptions," Human–Computer Interaction, 2025.
- [5]. L. Floridi and M. Taddeo, "Trustworthy artificial intelligence: From principles to measurable guarantees," IEEE Computer, vol. 58, no. 4, pp. 72–81, 2025.
- [6]. R. Guidotti, A. Monreale, and F. Giannotti, "Post-hoc verification of black-box decision systems," Knowledge- Based Systems, Elsevier, 2025.
- [7]. J. García-González et al., "Explainable AI in recruitment systems," Information Processing & Management, 2024.
- [8]. C. Rigotti, "Fairness, AI & recruitment," Information Systems Journal, 2024.
- [9]. A. Mehrabi, F. Morstatter, and A. Galstyan, "Robustness and stability of machine learning under input perturbations," IEEE Trans. Neural Netw. Learn. Syst., 2024.
- [10]. K. Kallus and A. Zhou, "Decision stability and sensitivity analysis for machine learning classifiers," IEEE Trans. Knowl. Data Eng., 2024.
- [11]. Y. Chen et al., "Bias evaluation in automated hiring systems," Knowledge-Based Systems, 2023.
- [12]. P. Hiremath and S. Deshpande, "Resume classification using supervised learning," International Journal of Information Management, 2023.
- [13]. J. Lee and R. Singh, "Ethical risks of AI-based resume screening," AI & Society, 2023.
- [14]. C. Papakyriakopoulos et al., "Measuring bias in algorithmic hiring systems," Data Mining and Knowledge Discovery, 2022.
- [15]. T. Zhang et al., "Public perceptions of algorithmic hiring," Technological Forecasting and Social Change, 2022.