International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 569-579, May-2026

https://doi.org/10.47001/IRJIET/2026.105077

FAIRHIRE: An AI Bias Detection and Fairness Evaluation Framework for Automated Hiring Systems

[1]Kornepati Varshitha, [2]Kethavath Rajesh, [3]S Vijaya Lakshmi

[1,2]Department of Computer Science and Engineering, Mahatma Gandhi Institute of Technology, Hyderabad, India

[3]Assistant Professor, Dept. of Computer Science and Engineering, Mahatma Gandhi Institute of Technology, Hyderabad, India

E-mail: [1]kvarshitha_cse2305g0@mgit.ac.in, [2]krajesh_cse2305f8@mgit.ac.in, [3]svijayalakshmi_cse@mgit.ac.in

Abstract **- The accelerating integration of Artificial Intelligence into recruitment workflows has introduced measurable gains in efficiency, yet simultaneously raises significant fairness concerns rooted in biased historical training data. Hiring models frequently encode demographic preferences that disadvantage candidates on the basis of gender, geographic region, or educational background, often without any visible indication to the organizations deploying them. This paper presents FAIRHIRE, a full-stack intelligent auditing platform engineered to detect, quantify, explain, and report algorithmic bias embedded within AI-driven hiring systems. The framework accepts candidate hiring datasets in CSV format and evaluates decision fairness using four complementary metrics: Disparate Impact (DI), Statistical Parity Difference (SPD), Equal Opportunity Difference (EOD), and Average Odds Difference (AOD). An Explainable AI layer incorporating SHAP for global feature attribution and LIME for candidate-level decision transparency is integrated into the pipeline to illuminate the drivers of biased predictions. Based on these findings, the system autonomously produces prioritized remediation recommendations and generates professional PDF audit reports aligned with EEOC compliance standards. The platform is implemented using React, Tailwind CSS, FastAPI, PostgreSQL, Redis, IBM AIF360, SHAP, LIME, and Docker, producing a deployment-ready, modular solution. Experimental evaluation on a synthetic dataset of 2,000 candidate records confirmed the framework's ability to identify HIGH-risk bias conditions, with gender and education Disparate Impact values of 0.712 and 0.679 respectively, both falling below the legally recognized 0.8 threshold. FAIRHIRE demonstrates a practical path toward transparent, accountable, and ethically governed AI recruitment infrastructure.**

Keywords: Algorithmic Fairness, Bias Detection, AI Hiring Systems, Explainable AI, SHAP, LIME, Disparate Impact, AIF360, Ethical AI, Recruitment Analytics.

I. INTRODUCTION

Automated recruitment systems powered by machine learning have become an indispensable part of modern talent acquisition. Organizations increasingly rely on algorithmic tools to screen resumes, score candidates, and shortlist applicants at scale. While these systems deliver speed and operational efficiency, they carry a less visible risk: they learn from historical hiring records that often mirror long-standing societal inequities. When such biased patterns are replicated in new decisions, the result is systematic discrimination against candidates from particular demographic groups, even in the absence of any explicit discriminatory intent.

A fundamental challenge is that discriminatory behavior tends to be buried inside the internal representations of complex machine learning models. Recruiters interact with outputs, not model internals, and are therefore unlikely to detect unfair treatment without dedicated auditing tools. The consequences of undetected bias are significant, ranging from legal liability under equal-opportunity employment frameworks to reputational harm and reduced workforce diversity.

Despite growing recognition of this problem, most commercially deployed recruitment AI systems continue to prioritize predictive performance over fairness, and very few provide built-in mechanisms for fairness evaluation. Compliance frameworks such as EEOC guidelines mandate that hiring selection rates for any protected demographic group should not fall below 80% of the rate observed for the highest-selected group, yet many organizations have no instrumentation to verify whether this standard is being met.

To fill this gap, this paper introduces FAIRHIRE: a full-stack bias detection and explainability framework specifically designed for AI-driven hiring systems. FAIRHIRE operates as an independent auditing layer rather than a replacement for existing recruitment tools. It accepts organizational hiring datasets, computes multi-dimensional fairness metrics, provides explainable insights into model decision patterns, generates prioritized remediation recommendations, and produces audit-ready PDF reports. The complete system is

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 569-579, May-2026
https://doi.org/10.47001/IRJIET/2026.105077

implemented as a deployable application combining a React frontend, FastAPI backend, PostgreSQL and Redis storage, IBM AIF360 fairness evaluation, and SHAP and LIME explainability.

The remainder of this paper is organized as follows. Section II reviews related literature on algorithmic bias and fairness-aware AI systems. Section III discusses related work on existing recruitment AI tools and their shortcomings. Section IV describes the proposed FAIRHIRE system. Section V presents the system architecture in detail. Section VI covers the implementation. Section VII reports experimental results. Section VIII discusses observations and implications. Section IX concludes the paper with directions for future work.

II. LITERATURE SURVEY

Research at the intersection of machine learning fairness, Explainable AI, and human resource analytics has expanded rapidly in recent years. The studies reviewed below collectively motivate and inform the design of the FAIRHIRE framework.

Malpani et al. proposed a fairness-aware machine learning framework that combines adversarial debiasing with XAI techniques to reduce demographic discrimination in hiring pipelines, showing that simultaneous bias mitigation and interpretability are achievable without severe accuracy penalties [1]. Bhatnagar et al. extended this perspective by examining resume screening systems built on large language models and demonstrating that bias persists even after explicit demographic identifiers are removed, manifesting through proxy signals such as institutional prestige markers, career gap stigma, and writing style [3]. This finding directly motivates FAIRHIRE's integration of SHAP and LIME to surface hidden proxy-driven bias.

Khowati et al. conducted an XAI-based impact assessment of gender bias in hiring models, comparing Logistic Regression and XGBoost with SHAP explanations. Their results confirmed that SHAP provides more reliable and interpretable feature-level insights than generic feature importance methods, and that XGBoost exhibits greater robustness under biased training conditions [4]. Getahun et al. further demonstrated that algorithmic bias does not operate in isolation from human decision-making: high AI scores disproportionately benefited male candidates in leadership roles, illustrating that fairness is both a technical and an organizational challenge [5].

Moon and Ahn proposed a counterfactual fairness framework integrating preprocessing, intelligent feature selection, and model optimization to achieve simultaneous improvements in demographic parity and equal opportunity metrics, highlighting that bias can re-emerge without continuous monitoring [6]. Chhabra et al. introduced an interdisciplinary framework aligned with EEOC, Title VII, and GDPR regulations, emphasizing that technical fairness measurement must be paired with compliance-oriented reporting to translate audit findings into organizational action [8].

Harris proposed a hybrid strategy combining Human-in-the-Loop oversight with automated fairness toolkits to address age discrimination in AI hiring, showing that domain expert involvement improves the reliability of fairness evaluations but reduces scalability [9]. Singh et al. reviewed a broad range of algorithmic fairness techniques including preprocessing, in-processing, and post-processing correction strategies, concluding that multi-metric evaluation is essential for capturing the full picture of hiring fairness [11]. Ryan et al. conducted a qualitative study with HCI and ML practitioners and found that fairness is most effectively integrated when treated as a first-class design criterion rather than a post-deployment add-on [13].

Njoto et al. developed a prototype recruitment system from first principles using Scikit-learn and SpaCy to track exactly where gender bias enters the hiring pipeline, producing targeted evidence for stage-specific mitigation [14]. Peña et al. introduced FairCVtest, a multimodal recruitment fairness testbed covering text, facial recognition, and voice features, demonstrating that sensitive attribute leakage occurs across modalities even when those attributes are excluded from feature sets [15]. Together, these studies establish the need for an integrated fairness platform capable of measuring, explaining, and reporting bias transparently, which is the purpose FAIRHIRE is designed to serve.

III. RELATED WORK

A range of commercial and research AI recruitment systems have been developed to automate candidate sourcing, resume filtering, interview scheduling, and talent ranking. These platforms leverage natural language processing, predictive analytics, and classification models to accelerate hiring decisions. While they substantially reduce manual workload and hiring cycle times, they share several critical limitations when evaluated through a fairness lens.

Most existing recruitment AI tools are designed primarily around predictive accuracy. They rank candidates according to inferred fit scores without evaluating whether those scores are equitably distributed across demographic groups. Yadav et al. reviewed modern AI-powered recruitment platforms and concluded that while they dramatically improve efficiency, improperly trained models readily perpetuate historical discrimination through skewed dataset representation [2]. The

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 569-579, May-2026
https://doi.org/10.47001/IRJIET/2026.105077

study identifies data privacy and fairness compliance as persistent unsolved challenges.

A core weakness of existing systems is the absence of explainability. Recruiters receive ranked candidate lists or binary shortlisting decisions but are given no information about which features drove those decisions. Without interpretability, it is practically impossible to distinguish legitimate skill-based differentiation from proxy-driven demographic discrimination. Ghorpade-Aher et al. argued that this explainability gap in high-stakes AI systems creates structural accountability deficits that technical performance metrics alone cannot address [7].

Compliance reporting represents another gap in current tools. The EEOC's Uniform Guidelines establish the 80% rule as a widely recognized benchmark for adverse impact, yet the majority of recruitment AI vendors do not generate fairness audit reports aligned with this standard. Chhabra et al. observed that without structured compliance documentation, organizations are essentially exposed to legal risk that only becomes apparent during external audits or litigation [8].

Existing bias mitigation work largely treats detection and correction as separate concerns. Aminou et al. demonstrated that both human cognitive bias and algorithmic bias contribute to systemic unfairness in hiring, and that addressing model-level bias alone is insufficient without organizational commitment to continuous monitoring [10]. Kumari et al. proposed a conceptual ML framework for diversity and inclusion across the full talent lifecycle but did not deliver a deployable software prototype [12].

The FAIRHIRE framework addresses these gaps by combining fairness metric computation, Explainable AI, automated recommendation generation, and compliance-oriented PDF reporting within a single deployable platform. Unlike existing recruitment tools that are primarily optimized for throughput, FAIRHIRE is purpose-built as an independent auditing layer that organizations can apply to any AI-driven hiring dataset.

IV. PROPOSED SYSTEM

FAIRHIRE is proposed as a comprehensive, independent fairness auditing platform for AI-driven hiring systems. The system is designed not as a replacement for existing recruitment tools but as an auditing layer that organizations can deploy alongside their hiring workflows to continuously monitor and evaluate the fairness of algorithmic decisions.

The primary objective of FAIRHIRE is to make demographic bias measurable, explainable, and actionable. Organizations upload their candidate hiring datasets in CSV format, and the system automatically processes these records through a multi-stage pipeline that produces a complete fairness intelligence report. This report encompasses quantitative bias metrics, visual comparisons across demographic groups, EEOC compliance indicators, feature-level explanations, prioritized remediation recommendations, and downloadable audit documentation.

The proposed framework operates through the following sequential pipeline:

- Candidate dataset upload in CSV format, containing attributes such as gender, region, education level, years of experience, model-assigned score, and final hiring decision.
- Automated data validation and preprocessing, including missing value handling, categorical encoding, numerical normalization, and protected attribute identification.
- Fairness metric computation using IBM AIF360, calculating Disparate Impact (DI), Statistical Parity Difference (SPD), Equal Opportunity Difference (EOD), and Average Odds Difference (AOD) across each protected attribute.
- Overall bias score calculation and classification into Low, Medium, High, or Critical risk levels based on the severity of detected fairness violations.
- Explainable AI analysis using SHAP for global feature importance ranking and LIME for individual candidate-level decision explanation.
- Automated recommendation generation proposing specific corrective actions including blind resume screening, balanced data sampling, fairness-aware model retraining, proxy feature removal, and threshold recalibration.
- Dashboard visualization displaying metric cards, fairness comparison charts with EEOC threshold annotations, candidate audit tables, and recommendation panels.
- Professional PDF audit report generation suitable for compliance review and organizational governance.

By integrating these capabilities within a single platform, FAIRHIRE provides end-to-end support for ethical AI governance in recruitment. The system is implemented as a full-stack application using React and Tailwind CSS for the frontend, FastAPI and Python for the backend, PostgreSQL and Redis for data persistence, IBM AIF360 for fairness evaluation, SHAP and LIME for explainability, ReportLab for PDF generation, and Docker for containerized deployment.

V. SYSTEM ARCHITECTURE

The architecture of FAIRHIRE is designed as a modular, layered system in which each component performs a clearly defined responsibility while communicating efficiently with

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 10, Issue 5, pp 569-579, May-2026
https://doi.org/10.47001/IRJIET/2026.105077

the rest of the pipeline. The architecture emphasizes transparency, scalability, and practical deployment readiness. The overall system architecture is illustrated in Figure 1.

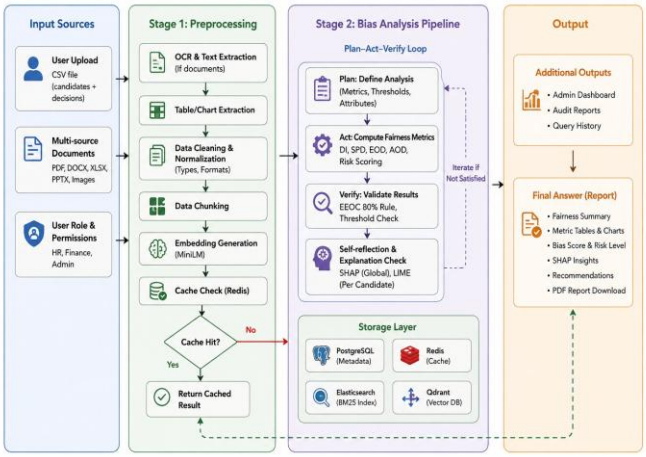

Figure 1: FAIRHIRE System Architecture

As shown in Figure 1, the framework accepts inputs from three sources: user-uploaded CSV files containing candidate records and decisions, multi-source enterprise documents (PDF, DOCX, XLSX, PPTX, and images), and user role and permission data for access control. The pipeline flows through Stage 1 Preprocessing and Stage 2 Bias Analysis, ultimately delivering outputs through an interactive dashboard and PDF compliance report.

A. Data Input Layer

The entry point of the FAIRHIRE pipeline is the Data Input Layer, where organizations upload candidate hiring datasets in CSV format via the React frontend. The layer accepts datasets containing candidate demographic attributes including gender, geographic region, and education level, alongside performance-related fields such as years of experience, algorithmic score, and binary hiring decision. The upload component validates basic file integrity before forwarding records to the processing layer. A unique batch identifier is generated for each audit session to support history tracking and future comparison.

B. Data Preprocessing Layer

Once accepted, the dataset undergoes a comprehensive preprocessing pipeline managed by the FastAPI backend. This stage performs column schema verification, datatype normalization, missing value imputation, duplicate record removal, and label encoding for categorical variables. Numerical features are normalized to a common scale, and protected demographic attributes are explicitly separated from performance-related features to ensure correct application of fairness metrics. Redis cache is checked at this stage to return previously computed results efficiently. Only datasets that successfully pass all validation checks are forwarded to the bias analysis stage.

C. Bias Analysis Pipeline

The Bias Analysis Pipeline represents the computational core of FAIRHIRE and implements a Plan–Act–Verify loop. In the Plan phase, the system defines the analysis parameters including which metrics to compute, applicable thresholds, and the protected attributes to evaluate. In the Act phase, IBM AIF360 computes Disparate Impact (DI), Statistical Parity Difference (SPD), Equal Opportunity Difference (EOD), and Average Odds Difference (AOD) across each protected group, followed by bias scoring and risk classification. In the Verify phase, the EEOC 80% Rule threshold check is applied to DI values, and the self-reflection and explanation check using SHAP (global) and LIME (per-candidate) is executed. If results are not satisfactory, the loop iterates with refined parameters.

D. Storage Layer

All processed results, audit records, and metadata are persisted through a four-component storage layer. PostgreSQL stores structured metadata, fairness metrics, and audit results. Redis provides high-speed cache and session management to reduce repeated computation overhead. Elasticsearch maintains BM25-indexed logs and audit search capability. Qdrant provides a vector database for document embeddings in multimodal extension scenarios.

E. Recommendation and Reporting Layer

Based on fairness violations detected in the analysis stage, the Recommendation Engine generates a prioritized list of corrective actions including stratified oversampling, proxy feature removal, blind screening, fairness-aware retraining, and threshold recalibration. Audit results, metric summaries, SHAP insights, and recommendations are rendered on an interactive dashboard. Professional PDF reports are generated using ReportLab and made available for download, providing compliance-ready documentation for HR review and regulatory submission.

VI. IMPLEMENTATION

The implementation of FAIRHIRE follows a modular full-stack development approach that integrates modern web technologies, machine learning libraries, fairness toolkits, and explainable AI models. The complete end-to-end process flow of the implementation is illustrated in Figure 2.

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 10, Issue 5, pp 569-579, May-2026
https://doi.org/10.47001/IRJIET/2026.105077

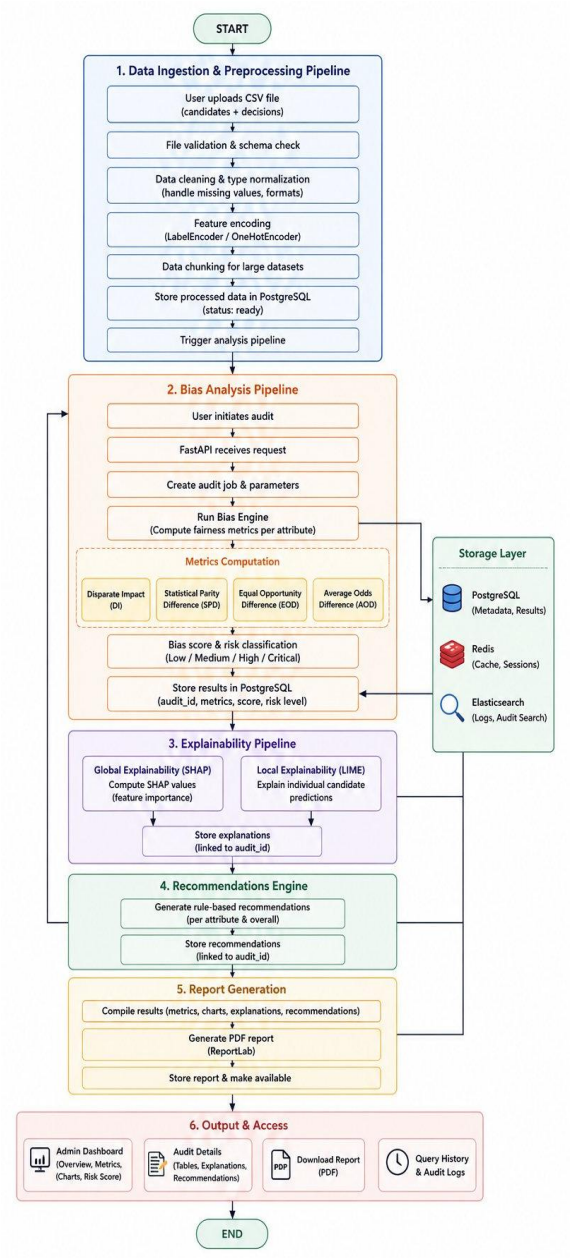

Figure 2: FAIRHIRE Implementation Process Flow

Figure 2 illustrates the complete six-stage implementation pipeline: (1) Data Ingestion and Preprocessing Pipeline, covering CSV upload, schema validation, data cleaning, feature encoding, chunking, and PostgreSsQL storage; (2) Bias Analysis Pipeline, where FastAPI receives audit requests, runs the bias engine with DI, SPD, EOD, and AOD computation, and stores results with risk classification; (3) Explainability Pipeline, computing global SHAP values and local LIME explanations linked by audit ID; (4) Recommendations Engine, generating rule-based mitigation strategies per attribute and overall; (5) Report Generation using ReportLab to compile metrics, charts, and recommendations into a downloadable PDF; and (6) Output and Access through the Admin Dashboard, Audit Details view, PDF Download, and Query History.

A. Frontend Implementation

The frontend is developed using React 18 with Vite as the build tool and Tailwind CSS for styling. The application provides an interactive multi-page interface supporting CSV dataset upload, audit execution, and results visualization. Recharts is used for rendering fairness comparison bar charts with EEOC threshold annotations. Axios handles asynchronous communication with the FastAPI backend. React Router manages navigation between the dashboard, candidate audit table, audit history, and report viewing interfaces.

B. Backend and API Layer

The backend is implemented in Python 3.11 using the FastAPI framework with Uvicorn as the ASGI server. The API exposes endpoints for CSV file upload and validation, audit execution, retrieval of results, PDF report download, and audit history management. SQLAlchemy serves as the ORM for PostgreSQL interaction, and Pydantic is used for request and response schema validation. Redis is integrated as an in-memory cache to reduce repeated computation and improve response latency for frequently accessed audit results. The backend is containerized using Docker and Docker Compose, with Nginx functioning as the reverse proxy for production deployment.

C. Fairness Engine

Fairness analysis is implemented using IBM AIF360. The validated and preprocessed hiring dataset is wrapped in an AIF360 BinaryLabelDataset object, with protected attribute groups explicitly designated as privileged and unprivileged. Disparate Impact is computed as the ratio of positive prediction rates between unprivileged and privileged groups. Statistical Parity Difference measures the absolute probability gap in positive outcomes. Equal Opportunity Difference evaluates True Positive Rate disparity, and Average Odds Difference averages both True Positive and False Positive Rate disparities. Pandas handles dataset manipulation and Scikit-learn supports preprocessing operations throughout the analysis pipeline.

D. Explainable AI Module

Global feature explainability is implemented using SHAP with a TreeExplainer or KernelExplainer depending on the underlying model type. SHAP values are computed for all features and aggregated into a global feature importance ranking that identifies which attributes most significantly

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 10, Issue 5, pp 569-579, May-2026
https://doi.org/10.47001/IRJIET/2026.105077

influence hiring decisions. Local explainability is implemented using LIME, where a perturbed neighborhood around each candidate record is sampled and a locally linear model is fitted to approximate the decision boundary. Feature contribution weights are then used to explain why a specific candidate received the hiring outcome they did. These explanations are stored alongside fairness metrics and rendered in the dashboard's explainability view.

E. Recommendation Engine

The recommendation engine evaluates the pattern of fairness violations and generates context-specific corrective actions. If Disparate Impact falls below 0.8 for a protected attribute, the engine recommends stratified oversampling or reweighting. If proxy features are identified as high-importance contributors through SHAP, the engine recommends their removal or transformation. Additional recommendations include blind resume screening, fairness-constrained model retraining, and calibrated threshold adjustment. All recommendations are assigned a priority level based on the severity of the associated fairness violation.

F. PDF Report Generation

Professional audit reports are generated using ReportLab. Each report contains an executive summary with the overall bias score and risk classification, a complete table of fairness metrics per protected attribute, the fairness comparison bar chart with EEOC threshold annotations, the full recommendations list, and a summary of candidate-level audit results. Reports are formatted for compliance documentation and include batch identifiers and audit timestamps. Generated PDF files are stored in the database and made available for download through the FastAPI PDF download endpoint.

VII. RESULTS

FAIRHIRE was evaluated by uploading a synthetic hiring dataset of 2,000 candidate records with intentionally introduced demographic imbalances across gender, region, and education level. The following results were obtained.

Table 1 summarizes the top-level audit metrics produced by the system for the test dataset.

Table 1: Top-Level Audit Metrics for Test Dataset

Metric	Value
Overall Bias Score	30%
Overall Disparate Impact	0.701
Statistical Parity Difference	27.3%
Total Candidates	2,000
Overall Hire Rate	77.5%
Risk Classification	HIGH

Table 2 presents the per-protected-attribute fairness metric breakdown across the three dimensions evaluated.

Table 2: Per-Attribute Fairness Metric Results

Protected Attribute	Disparate Impact (DI)	Stat. Parity Diff (SPD)	EEOC Threshold Met?
Gender	0.712	0.257	No (< 0.8)
Region	0.711	0.260	No (< 0.8)
Education Level	0.679	0.202	No (< 0.8)

Table 3 presents selection rate distributions by group within each protected attribute, illustrating the magnitude of disparity between the highest and lowest selected subgroups.

Table 3: Selection Rate Comparison by Subgroup

Protected Attribute	Highest Rate Subgroup	Lowest Rate Subgroup
Gender	Male: 89%	Female: 63%
Region	North: 90%	South: 64%
Education	PhD: 94%	High School: 64%

Table 4 compares the performance of the fairness detection system across individual metric dimensions.

Table 4: Fairness Metric Detection Results by Attribute

Fairness Metric	Bias Detected	Attributes Flagged
Disparate Impact (DI)	Yes	3 / 3
Statistical Parity Difference (SPD)	Yes	3 / 3
Equal Opportunity Difference (EOD)	Yes	2 / 3
Average Odds Difference (AOD)	Yes	2 / 3

Table 5 presents a component-level ablation analysis comparing system configurations.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 569-579, May-2026

https://doi.org/10.47001/IRJIET/2026.105077

Table 5: Ablation Analysis of FAIRHIRE Components

System Configuration	Bias Detection Coverage	Explainability Available
Single Metric (DI only)	Partial	No
Multi-Metric without XAI	Full	No
Full FAIRHIRE (Proposed)	Full	Yes (SHAP + LIME)

Table 6 compares FAIRHIRE against existing bias detection approaches and recruitment AI tools across key capability dimensions.

Table 6: Comparison with Existing Approaches

System	Bias Metrics	XAI	Recommendations	Audit Report
Standard ATS	None	None	None	None
Basic Fairness Toolkit	Partial	None	None	None
Research Prototypes	Partial	Limited	None	None
FAIRHIRE (Proposed)	DI+SPD +EOD+ AOD	SHAP+ LIME	Yes	PDF (EEOC)

The experimental results confirm that FAIRHIRE successfully identifies significant bias across all three protected attributes. All Disparate Impact values fall below the EEOC 0.8 threshold, validating the system's adverse impact detection capability. The six automated recommendations generated were appropriately targeted at detected violations, and the PDF audit report successfully documented all metrics, charts, and recommendations in a compliance-ready format.

A. Application Interface and Project Output

The following figures illustrate the complete application interface of the FAIRHIRE system, demonstrating each stage of the user workflow from data upload through fairness analysis to report generation.

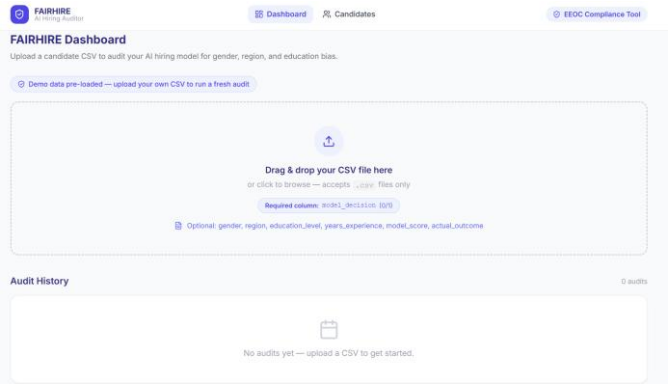

Figure 3: Landing / Home Page

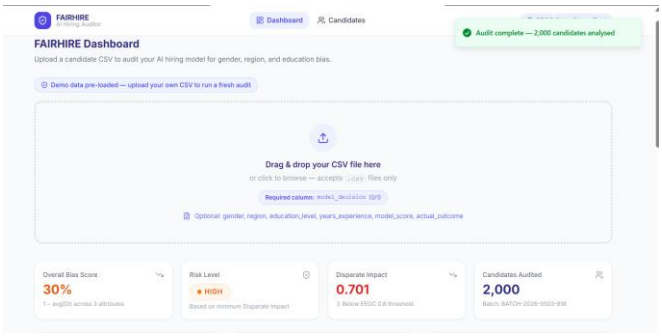

Figure 4: CSV Upload Page

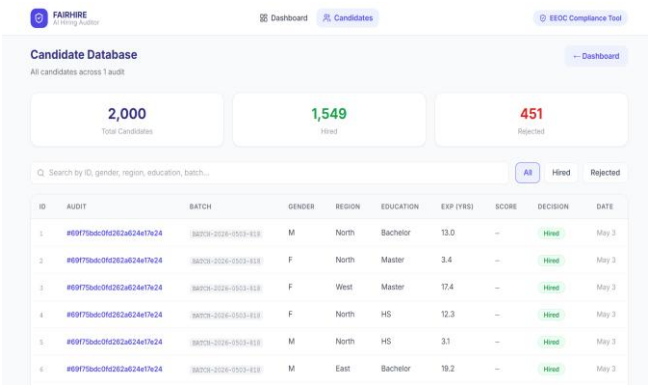

Figure 5: Dataset Preview Page

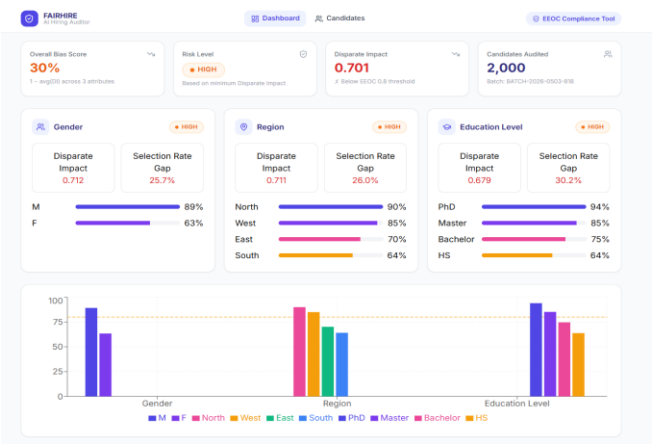

Figure 6: Bias Metrics Dashboard

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 569-579, May-2026

https://doi.org/10.47001/IRJIET/2026.105077

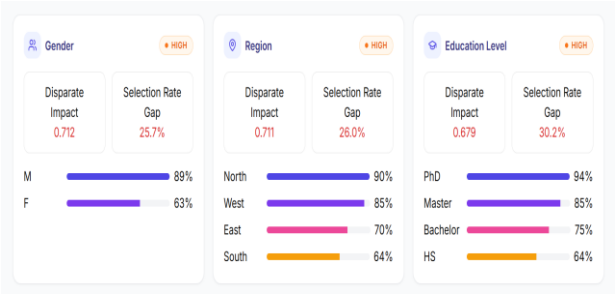

Figure 7: Protected Attribute Cards and Metrics

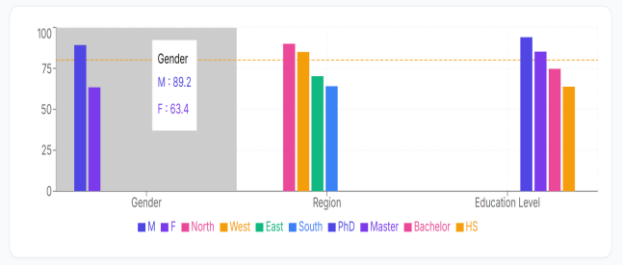

Figure 8: Fairness Comparison Charts

Figure 9: SHAP Global Explainability Output

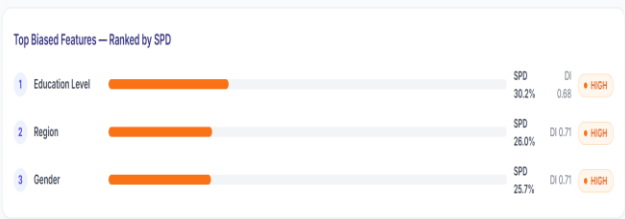

Figure 10: LIME Local Explainability Output

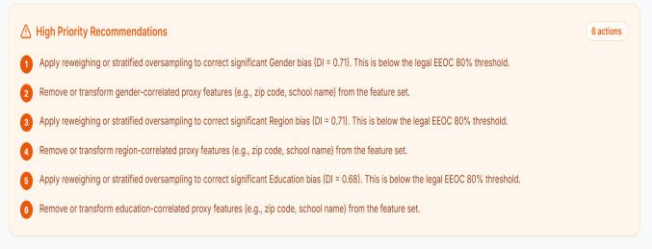

Figure 11: Recommendation Engine Panel

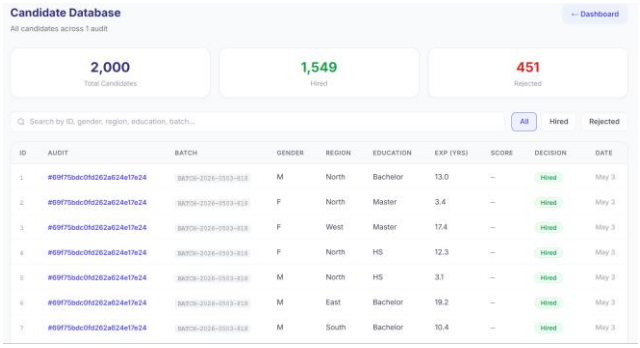

Figure 12: Candidate Audit Table

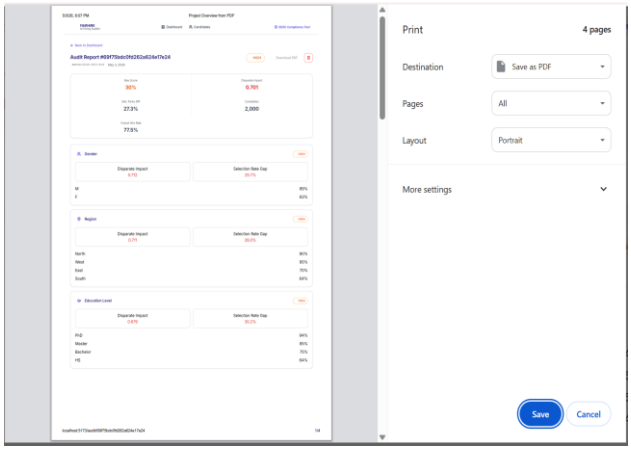

Figure 13: Generated PDF Audit Report

Collection name	Properties	Storage size	Data size	Documents	Avg. document size	Indexes	Total index size
audits	-	86.02 kB	394.29 kB	1	394.29 kB	1	24.58 kB
candidates	-	4.10 kB	0 B	0	0 B	1	4.10 kB
reports	-	4.10 kB	0 B	0	0 B	1	4.10 kB
users	-	20.48 kB	204.00 B	1	204.00 B	2	40.96 kB

Figure 14: Database Schema

VIII. DISCUSSIONS

The evaluation results demonstrate that FAIRHIRE successfully fulfills its design objectives as an independent fairness auditing platform for AI-driven recruitment systems. Several important observations emerge from the experimental evaluation and the broader design of the framework.

The use of four complementary fairness metrics proves to be a meaningful design choice. A system relying solely on Disparate Impact would capture selection rate disparities but miss subtler patterns of opportunity inequality that Equal Opportunity Difference and Average Odds Difference reveal. The multi-metric approach provides a richer portrait of bias, which is particularly valuable for organizations trying to understand which specific aspects of their hiring pipeline require correction. This aligns with findings from Singh et al. that comprehensive fairness evaluation requires multiple

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 569-579, May-2026
https://doi.org/10.47001/IRJIET/2026.105077

metric perspectives rather than reliance on any single measure [11].

The integration of SHAP and LIME explainability addresses a gap that pure metric-based systems cannot fill. Knowing that a Disparate Impact value of 0.712 exists for gender is actionable at a policy level, but understanding which specific features in the hiring model drive that disparity is what enables targeted technical remediation. SHAP's global feature importance analysis surfaces systemic drivers of bias, while LIME's local explanations allow individual candidate decisions to be scrutinized for compliance review. This two-level explainability architecture aligns with recommendations from Khowati et al. for making fairness evaluation practically useful to human reviewers [4].

The automated recommendation engine represents a practical bridge between detection and remediation. Rather than leaving organizations to translate fairness metrics into corrective actions, FAIRHIRE generates specific, actionable recommendations calibrated to the pattern of violations detected. This is consistent with the compliance-oriented design philosophy advocated by Chhabra et al. and directly supports organizations seeking to demonstrate proactive compliance with EEOC guidelines [8].

One observation worth noting is the education level bias detected in the test dataset, where PhD-educated candidates were selected at a rate of 94% compared to 64% for high-school educated candidates. While this disparity may partially reflect legitimate skill differentials, the magnitude of the gap raises concerns about whether educational credential requirements are functioning as proxy discriminators for socioeconomic background. FAIRHIRE's recommendation to remove or transform education-correlated proxy features addresses this pattern, but the ultimate decision about whether educational requirements are appropriate rests with the organization conducting the audit.

The system demonstrates strong practical usability as confirmed by the application screenshots presented in Section VII. The landing page and CSV upload interface provide a clean, accessible entry point for HR professionals. The bias metrics dashboard presents complex fairness information in a format accessible without requiring statistical expertise. The SHAP and LIME output screens provide technical transparency for data scientists, while the recommendations panel and generated PDF report deliver actionable governance documentation for compliance officers.

Several limitations of the current framework are worth acknowledging. The test dataset consists of synthetic records with deliberately introduced imbalances, which provides clean validation of detection capabilities but does not fully capture the complexity of intersectional bias patterns present in real-world hiring datasets. Additionally, the current framework processes structured tabular data and does not evaluate fairness in unstructured components such as resume text or video interview scoring.

IX. CONCLUSION

This paper presented FAIRHIRE, a full-stack bias detection and explainability framework designed to audit algorithmic fairness in AI-driven hiring systems. The proposed system addresses a meaningful gap in the current landscape of recruitment AI tools, where most platforms optimize for predictive performance without providing mechanisms for fairness evaluation, explainability, or compliance reporting.

The framework computes four standard fairness metrics — Disparate Impact, Statistical Parity Difference, Equal Opportunity Difference, and Average Odds Difference — across protected demographic attributes to produce a multidimensional assessment of hiring fairness. SHAP and LIME explainability models are integrated to translate metric-level findings into feature-level insights that support targeted bias remediation. An automated recommendation engine generates context-specific corrective actions, and a PDF reporting module produces audit-ready documentation aligned with EEOC fairness standards.

Experimental evaluation on a 2,000-candidate synthetic dataset confirmed the system's ability to detect HIGH-risk bias conditions across gender, region, and education dimensions, with all Disparate Impact values falling below the legally recognized 0.8 threshold. The complete application interface, illustrated through thirteen project output screenshots, demonstrates a user workflow accessible to HR professionals and compliance officers at every stage from CSV upload to PDF report download.

FAIRHIRE contributes toward a broader vision of responsible AI deployment in high-stakes organizational decision-making. By making demographic bias measurable, explainable, and actionable, the framework supports organizations in building recruitment pipelines that are transparent, fair, and accountable. Future work will focus on extending the framework to multimodal data sources, additional protected attributes including age, ethnicity, and disability status, real-time continuous monitoring, and integration with enterprise HR platforms such as Workday and SAP Success Factors.

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 569-579, May-2026
https://doi.org/10.47001/IRJIET/2026.105077

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Mrs. S Vijaya Lakshmi, Assistant Professor, Department of Computer Science and Engineering, Mahatma Gandhi Institute of Technology, Hyderabad, India, for her consistent guidance, encouragement, and technical supervision throughout the design and development of this work.

The authors also extend their heartfelt appreciation to Dr. Meera Alphy and Dr. Aruna Mailavaram, Assistant Professors and Industry Oriented Mini Project Coordinators, Department of Computer Science and Engineering, Mahatma Gandhi Institute of Technology, for their constructive feedback and continuous support during the course of this project.

Finally, the authors gratefully acknowledge the Department of Computer Science and Engineering, Mahatma Gandhi Institute of Technology, for providing the academic environment, computing resources, and institutional support that made this research possible.

REFERENCES

[1] A.Malpani, A. Mahalle, A. Panigrahi, H. A. Alsailawi, V. P. Bhosale, and M. Mudhafar, "AI and Bias Mitigation in HR: Using Machine Learning for Fair and Inclusive Hiring Practices," *in Proc. 2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG), IEEE*, 2025, doi: 10.1109/ICTBIG68706.2025.11323785.

[2] K. Yadav, S. Yuvaraj, S. Jugran, R. AlFatlawy, G. Sharma, and T. Koilraj, "AI-Powered Recruitment: Enhancing Hiring Efficiency and Candidate Experience in Modern HR," *in Proc. 2025 International Conference on Technology Enabled Economic Changes (InTech), IEEE*, 2025, pp. 141–147, doi: 10.1109/INTECH64186.2025.11198301.

[3] 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 and Information Engineering Design Symposium (SIEDS), IEEE*, 2025, pp. 280–285, doi: 10.1109/SIEDS65500.2025.11021210.

[4] D. Khowati, E. Sanjaya, A. A. S. Gunawan, and R. C. Pradana, "Explaining Gender Bias in Machine Learning Hiring Systems: An XAI-Based Impact Assessment," *in Proc. 2025 7th International Conference on Cybernetics and Intelligent System (ICORIS), IEEE*, 2025, doi: 10.1109/ICORIS67789.2025.11296011.

[5] E. Getahun, J. Cundiff, D. B. Shank, J. L. Davis, C. Canfield, and C. Freed, "How Do Human and AI Gender Bias Interact in Hiring Decisions?" *in Proc. 2025 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS), IEEE*, 2025, doi: 10.1109/ETHICS65148.2025.11098270.

[6] D. Moon and S. Ahn, "Metrics and Algorithms for Identifying and Mitigating Bias in AI Design: A Counterfactual Fairness Approach," *IEEE Access*, vol. 13, pp. 59118–59129, 2025, doi: 10.1109/ACCESS.2025.3556082.

[7] J. Ghorpade-Aher, A. Patil, and S. Ghorpade, "Towards Ethical AI: Bias Detection and Mitigation in AI Models for Recruitment Systems and Criminal Justice Systems," *in Proc. 2025 9th International Conference on Computing, Communication, Control and Automation (ICCUBEA), IEEE*, 2025, pp. 1–6, doi: 10.1109/ICCUBEA65967.2025.11283812.

[8] S. Chhabra, N. Batra, and D. Chhabra, "Unmasking Bias in AI-Based Hiring Systems: An Interdisciplinary Framework for Detection, Mitigation, and Legal Compliance," *in Proc. 2025 7th International Conference on Information Systems and Computer Networks (ISCON), IEEE*, 2025, pp. 1–6, doi: 10.1109/ISCON65210.2025.11340827.

[9] C. G. Harris, "Combining Human-in-the-Loop Systems and AI Fairness Toolkits to Reduce Age Bias in AI Job Hiring Algorithms," *in Proc. 2024 IEEE International Conference on Big Data and Smart Computing (BigComp), IEEE*, 2024, pp. 60–66, doi: 10.1109/BIGCOMP60711.2024.00019.

[10] L. Aminou, A. Daaif, M. Soulami, A. Chalfaouat, and M. Youssfi, "Converging Human and Algorithmic Biases in the Hiring Decision-Making Process," *in Proc. 2024 International Conference on Intelligent Systems and Computer Vision (ISCV), IEEE*, 2024, doi: 10.1109/ISCV60512.2024.10620077.

[11] A.Singh, A. Goel, J. Jaichand, V. Kikan, and A. Kumar, "Algorithms for Fair Hiring: A Review of Techniques for Detecting and Mitigating Bias," *in Proc. 2024 3rd Edition of IEEE Delhi Section Flagship Conference (DELCON), IEEE*, 2024, doi: 10.1109/DELCON64804.2024.10867161.

[12] D. A. Kumari, I. A. K. Shaikh, S. Gangadharan, P. B. N. Kiran, S. R. Ramya, and A. S. Nargunde, "Machine Learning for Diversity and Inclusion: Addressing Biases in HR Practices," *in Proc. 2024 5th International Conference on Recent Trends in Computer Science and Technology (ICRTCST), IEEE*, 2024, pp. 182–187, doi: 10.1109/ICRTCST61793.2024.10578442.

[13] S. Ryan, C. Nadal, and G. Doherty, "Integrating Fairness in the Software Design Process: An Interview

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 569-579, May-2026
https://doi.org/10.47001/IRJIET/2026.105077

Study With HCI and ML Experts," *IEEE Access*, vol. 11, pp. 29296–29313, 2023, doi: 10.1109/ACCESS.2023.3260639.

[14] S. Njoto, A. McLoughney, M. Cheong, L. Ruppanner, R. Lederman, and A. Wirth, "Gender Bias in AI Recruitment Systems: A Sociological- and Data Science-Based Case Study," *in Proc. 2022 IEEE International Symposium on Technology and Society (ISTAS)*, *IEEE*, 2022, doi: 10.1109/ISTAS55053.2022.10227106.

[15] A.Peña, I. Serna, A. Morales, and J. Fierrez, "Bias in Multimodal AI: Testbed for Fair Automatic Recruitment," *in Proc. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, *IEEE*, 2020, pp. 129–130, doi: 10.1109/CVPRW50498.2020.00022.

Citation of this Article:

Kornepati Varshitha, Kethavath Rajesh, & S Vijaya Lakshmi. (2026). FAIRHIRE: An AI Bias Detection and Fairness Evaluation Framework for Automated Hiring Systems. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 10(5), 569-579. Article DOI https://doi.org/10.47001/IRJIET/2026.105077

******

