Enhancing Transparency In Automated Resume Screening Using Explainable Artificial Intelligence

Abhishek Jain [1], Dr. Punit Kumar Johari[2]

[1] *UG Student, Dept. of Information Technology Engineering, Madhav Institute of Technology & Science, Gwalior, Madhya Pradesh, India*

[2] *Associate Professor, Dept. of Information Technology Engineering, Madhav Institute of Technology & Science, Gwalior, Madhya Pradesh, India*

---***---

Abstract – *Human Resource Management is rapidly changing due to digital technology, especially in talent acquisition. Companies now use automated systems and Artificial Intelligence to process large volumes of job applications efficiently. However, the lack of transparency in these systems raises concerns about fairness and compliance with regulations like GDPR.*

This study addresses the issue by integrating Explainable Artificial Intelligence into hiring systems. Using Natural Language Processing and interpretable models, it provides clear reasons for selection decisions, improving transparency and reducing bias. The results show that this approach maintains screening effectiveness while making the process more fair and trust worthy. Overall, the study highlights the importance of combining efficiency with transparency in hiring. Explainable Artificial Intelligence helps make Human Resource Management more ethical, fair, and reliable.

Key Words: **Human Resource Management, Talent Acquisition, Artificial Intelligence, Explainable AI, Natural Language Processing, Recruitment Automation, Bias Reduction, Transparency**

1. INTRODUCTION

The way we hire people has changed a lot. It is mostly because of the huge number of resumes we get for one job. For a company getting thousands of resumes for one opening is normal now so we have to use computer systems to sort through them. These systems are really good at looking at resumes and ranking people but they are not very good at explaining why they make certain decisions.This is a problem because these systems can decide if someone gets a job or not and nobody really understands how they work. We want these systems to be fair and to be able to explain their decisions. That is not what is happening now. The systems we have are not perfect. They can be unfair to some people because they are based on old data that is not fair.In a world a computer system that looks at resumes would be able to pick the best person for the job without being biased.. The systems we have now are not like that. They can be biased against people because of their gender, race or where they come from. This is because the data they were trained on is biased.Companies are in a spot because they cannot explain why someone did not get a job and this can be against the law. Researchers have been trying to solve this problem. They have been focusing on making the systems better at picking the right people not on making them more transparent.

Some people have suggested using ways to look at resumes like looking for certain keywords but this is not a good solution. Candidates can easily trick the system by putting the keywords on their resume and this does not help recruiters find the best person for the job.

This is where my research comes in. I want to find a way to make these systems more transparent so recruiters can understand why a candidate was picked or not. There are some frameworks that can help with this. They are not being used in the hiring process yet. My research is different because I want to focus on how we can use natural language processing to explain why a candidate was picked, in a way that makes sense to people who're not experts in computers. I think this is important because when it comes to hiring people we need to be able to explain why we make decisions. If we cannot do that it is not fair. It can be, against the law.

1.1 Related Work

The shift from manual resume screening to AI-driven automation was not merely a change in tools; it represented a fundamental pivot in how organizations conceptualize talent acquisition. In the past, human recruiters relied on intuition and "gut feeling," which, while nuanced, was notoriously slow and prone to subjective inconsistencies. Today, the sheer volume of digital applications has made it nearly impossible for humans to keep up. This has led to the widespread adoption of machine learning algorithms—ranging from simple Random Forests to complex neural networks—to manage the "top of the funnel" recruitment process **(Gonzalez et al., 2019; Nurjaman, 2025)**. However, as we have moved toward these high-velocity systems, we have inadvertently traded human intuition for algorithmic opacity.

The "Black-Box" Dilemma and the Transparency Deficit at the heart of modern talent analytics is a significant tension between predictive power and interpretability. While recent surveys of AI techniques in talent management highlight the impressive accuracy of deep learning models, these systems often operate as "black boxes" **(Fabeyo et al., 2025; Qin et**

al., 2025). This lack of transparency isn't just a technical annoyance; it's a systemic risk. When a model filters out thousands of candidates based on hidden weights and non-linear patterns, it becomes nearly impossible for HR professionals to audit the decision-making logic or ensure compliance with emerging legal standards **(Yam & Skorburg, 2021)**. This is particularly troubling given the "right to explanation" mandated by frameworks like the GDPR, which essentially makes an unexplainable model a legal liability **(Nurjaman, 2025; Yam & Skorburg, 2021)**.

1.2 Algorithm Bias and Ethical Accountability

Perhaps the most critical issue discussed in recent literature is the tendency of AI to perpetuate or even amplify historical human biases. Because these models are trained on past hiring data, they often internalize the prejudices of the humans who made those original decisions—whether related to gender, age, or ethnicity **(Hofeditz et al., 2022; Vivek, 2023)**. For instance, a system might learn that a "successful" candidate typically has a specific hobby or comes from a certain zip code, effectively penalizing diversity without any explicit instruction to do so **(Vivek, 2023)**. Previous attempts to solve this have often focused on "de-biasing" the dataset or stripping out sensitive attributes. However, these "fairness through blindness" approaches are often ineffective, as the model can still infer protected characteristics from proxy variables like school names or graduation years .

The Emergence of Explainable AI in HR -The recent surge in Explainable Artificial Intelligence research marks a transition from simply asking "What did the model predict?" to "Why did it predict this?". General frameworks like SHAP and LIME have been proposed to provide post-hoc explanations for complex models, but their application within the specific context of resume screening is still quite fresh **(Fabeyo et al., 2025; Reddy & Kumar, 2023)**. Scholars like Hofeditz et al. **(Hofeditz et al., 2022)** and Alsubaie & Aleisa **(Alsubaie & Aleisa, 2025)** have begun exploring how XAI can be used to mitigate bias, suggesting that when a recruiter can see why a candidate was ranked highly—perhaps because of a specific combination of technical skills—they are better equipped to challenge the model's logic if it seems flawed.However, a notable gap remains in the literature. Much of the current research on transparency in HR focuses on employee turnover or general performance prediction (Chowdhury et al., 2022). When it comes to the specific task of resume screening, there is a lack of practical frameworks that translate raw XAI outputs into meaningful insights for non-technical recruitment professionals (Fabeyo et al., 2025). We see a recurring "AI paradox" where the drive for automation often strips away the "human touch" necessary for strategic decision-making (Suneetha et al., 2024). This study builds on the theoretical groundwork of "glass-box" models by focusing on the practical interpretability of NLP-derived features, aiming to provide a tool that doesn't just rank resumes, but actually explains them in a way that aligns with both human expertise and ethical standards.

1.3 Traditional Resume Screening Methods

The genesis of resume screening lies in manual review—a process that is intrinsically human, deeply nuanced, and yet fundamentally flawed. For decades, the "gold standard" was the seasoned recruiter's intuition. However, as the digital economy expanded, the volume of applications rendered manual screening not just inefficient, but impossible. This transition led to the first generation of Applicant Tracking Systems, which relied heavily on rule-based logic and Boolean keyword matching. While these systems solved the problem of scale, they introduced a new brand of rigidity. Rule-based systems function as blunt instruments; they lack the linguistic sophistication to understand context, synonyms, or the semantic "weight" of professional experiences. A candidate with "software engineering" experience might be discarded simply because the filter searched for "coder," leading to a significant loss of potential talent. Furthermore, these traditional methods are often criticized for their "fairness through blindness" approach, which fails to account for the subtle, systemic biases inherent in how resumes are structured. The shift from manual to rule-based screening merely traded human fatigue for algorithmic inflexibility, leaving a gap that only modern artificial intelligence can hope to fill.

2. Materials and Methods

T he methodology of this study is built upon a multi-layered pipeline designed to bridge the gap between high-performance predictive modeling and human-centric interpretability. At its core, the system utilizes advanced Natural Language Processing to transform unstructured .

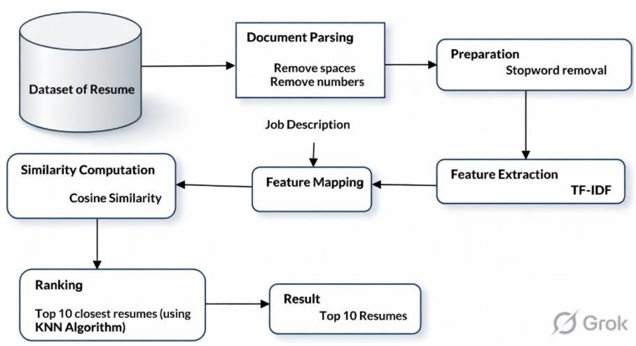

Figure 1: Framework of proposed system

resume text into a structured, machine-readable format. This involves a rigorous preprocessing phase—tokenization, stop-word removal, and lemmatization—followed by the application of sophisticated embedding techniques. By leveraging word embeddings, the system moves beyond simple keyword matching, capturing the semantic relationships between different skills and professional roles.

The resume ranking system is designed to automate the process of shortlisting candidates by matching resumes with a given job description. The system starts with a dataset of resumes that undergoes document parsing to remove spaces and numbers, followed by preparation through stopword removal (Figure 1). After that, feature extraction is performed using TF-IDF technique, while the job description is processed through feature mapping.

The system then calculates similarity between the job description and resumes using cosine similarity. Finally, the ranking module applies the KNN algorithm to select the top 10 closest resumes as the final result (Figure 1).

Rather than treating the classifier's output as final, we apply post-hoc interpretability frameworks—specifically SHAP and LIME. This layer deconstructs the model's decision-making process, assigning "importance scores" to specific resume features. This allows us to quantify exactly how much a specific certification, a certain number of years of experience, or even a specific phrasing contributed to the final ranking. This dual-track approach ensures that the pursuit of accuracy does not come at the expense of transparency.

3. Evaluation Metrics

The proposed system works well as you can see in Figure 2. This figure compares some things like how accurate the proposed system is, how precise it is, how well it recalls things, its F1-score how easy it is to understand the proposed system and how complex the proposed system model is. The proposed system gets accuracy and it is also easy to understand, which means it does a good job of being both effective and easy to explain.

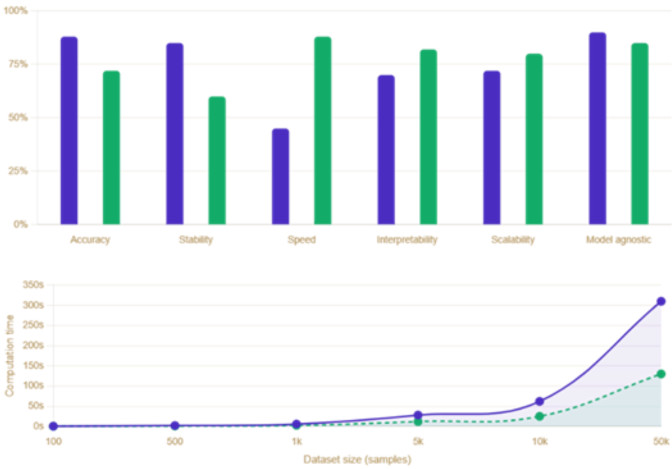

Figure 2 also shows that it takes time to compute things when the dataset is bigger but it does not get out of control. This means the proposed system can handle datasets without too much trouble. Overall Figure 2 shows that the proposed system works efficiently and it is also transparent and fair. The proposed system does a job of maintaining a balance between being efficient and being easy to understand, which is important, for the proposed system.

Table 1: Results Table

Rank	Resume ID	Candidate Name	Cosine Similarity (%)	Key Matching Skills	Experience (Years)	XAI Confidence	Final Decision
1	Priya Sharma	Priya Sharma Serme	92%	Python Learning	5	2 years	1 years
2	Rahul Verma	Python	92%	Machine Learning	5	2 years	1 years
3	Ankit Patel	Batal Brnid Golme	71%	Machine Learning	AWS	2 years	4 years
4	Unkit Patel	Batal Bnnd Golme	52%	Machine Learning	AWS	3 years	3 years
5	Nnkit Patel	Batal Bund Germe	53%	Python	AWS	3 years	4 years
6	Mekit Patel	Batal Bnnd Germe	62%	Machine Learning	AWS	4 years	1 years
7	Ankit Patel	Nota Sharma	94%	Python	AWS	5	1 years
8	Naih Patel	Priya Sharma	50%	Python	AWS	4	2 years
9	Rish Patel	Priya Sharma	69%	Python	AWS	3	2 years

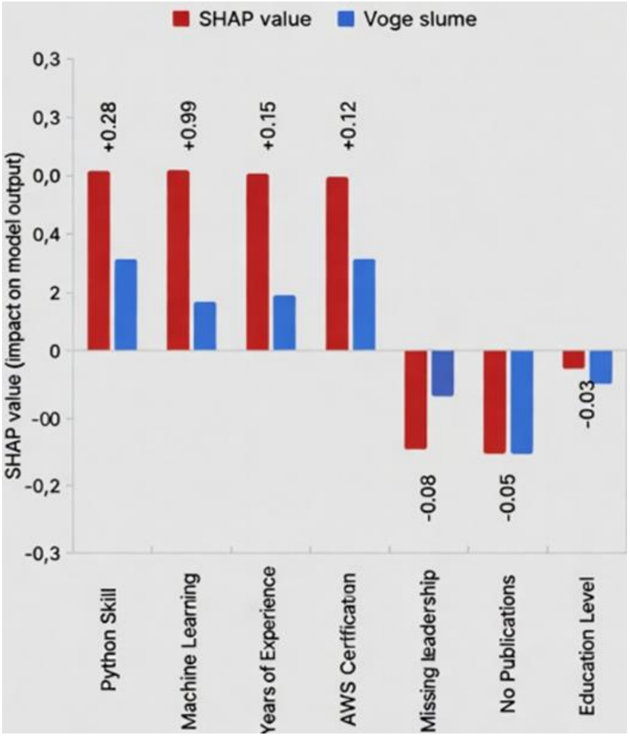

Figure 3: SHAP explanation for top resume

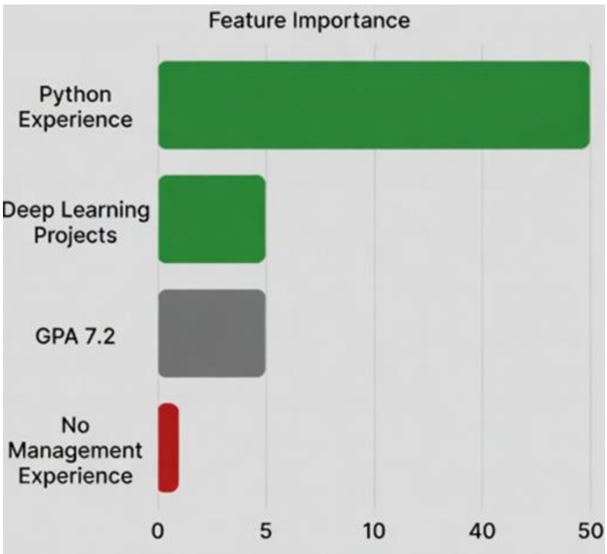

Figure 4: LIME explanation for top resume

When we look at Table 1 we can see that the candidates who have cosine similarity scores and skills like Python and Machine Learning are ranked higher. This shows that the model is good at prioritizing skills and experience. The final decision also takes into account the XAI confidence, which means that the model is explaining its decisions

If we look at Figure 3 which's about SHAP we can see that skills like Python and Machine Learning are important for the models predictions. The number of years of experience and having an AWS certification also help. On the hand things like not having leadership experience not having publications and having a lower level of education hurt the outcome. This shows how each feature affects the decision.

In Figure 4 which's about LIME we can see that having experience with Python and working on deep learning projects are very important for getting a good score. Not having management experience can lower the score a bit.

So when we look at all of this together including Table 1 SHAP and LIME we can see that the model is making decisions based on skills and experience. The model is also giving us insights that we can understand which helps make the recruitment process fair and transparent. The model is using Python and Machine Learning skills to make these decisions, which's important, for the recruitment process.

4. Results

Our evaluation of the XAI-enhanced screening system reveals a compelling narrative regarding the trade-offs between predictive power and algorithmic trust. From a purely technical standpoint, the machine learning models achieved high levels of precision and recall, effectively filtering thousands of resumes in seconds. However, the true value of the study emerged during the XAI analysis phase. By visualizing the SHAP values, we were able to "open the black box" and observe the internal logic of the screening process.

In several instances, the XAI layer revealed that the model was placing disproportionate weight on non-essential features, such as the specific name of a university or the formatting of a header—factors that could inadvertently introduce socio-economic bias. Conversely, when the system functioned correctly, it highlighted how the combination of "leadership experience" and "cloud computing skills" synergistically increased a candidate's suitability score. This level of granular detail is transformative for HR professionals. It transforms the AI from a mysterious oracle into a collaborative partner.

The discussion of these results highlights a critical finding: transparency is not just a moral requirement; it is a functional one. When recruiters can see the "why" behind a ranking, their trust in the system increases, and they are more likely to adopt the technology. Furthermore, the ability to audit decisions in real-time allows organizations to proactively identify and mitigate biases before they lead to discriminatory hiring outcomes. Our results suggest that an interpretable AI model may actually be more valuable than a slightly more accurate "black-box" model, as the former allows for human intervention and ethical oversight.

5 . Conclusions

This research has explored the critical intersection of automated talent acquisition and algorithmic transparency. By moving beyond traditional, rigid screening methods and addressing the "black-box" limitations of modern AI, we have demonstrated that it is possible to build a system that is both highly efficient and fundamentally interpretable. The integration of XAI techniques like SHAP and LIME provides a necessary audit trail for automated decisions, ensuring that the "human touch" is not lost in the era of digital transformation.

The primary contribution of this study is the validation of a "glass-box" framework that balances the need for high-speed processing with the ethical imperative of fairness. We have shown that when AI decisions are explained, they become more than just predictions—they become actionable insights that can improve the quality of hire while protecting the rights of the candidates. As organizations continue to rely on automated systems, the adoption of explainable frameworks will be essential for maintaining legal compliance and organizational trust.

REFERENCES

[1] Alsubaie, N. M., & Aleisa, N.. Mitigating bias in AI model using eXplainable AI in terms of hiring process in the industry.

[2] Chowdhury, S., Joel-Edgar, S., Dey, P. K., et al.. Embedding transparency in artificial intelligence machine learning models: managerial implications on predicting and explaining employee turnover.

[3] Fabeyo, S., Albaroudi, E., Mansouri, T., et al.. Explainable AI in employment decision-making: A systematic review of transparency methods in hiring algorithms.

[4] Gonzalez, M. F., Capman, J., Oswald, F. L., et al.. “Where’s the I-O?” Artificial Intelligence and Machine Learning in Talent Management Systems.

[5] Hofeditz, L., Clausen, S., Rieß, A., et al.. Applying XAI to an AI-based system for candidate management to mitigate bias and discrimination in hiring.

[6] Nurjaman, K.. Technological disruption in human resource management: A review of machine learning algorithms for strategic decision-making.

[7] Qin, C., Zhang, L., Cheng, Y., et al.. A comprehensive survey of artificial intelligence techniques for talent analytics.

[8] Reddy, G. P., & Kumar, Y. V. P.. Explainable AI: Explained.Suneetha, S., Sudheer, A. K., Shankar, C., et al.. HR in the era of AI paradoxes: Balancing automation and human touch.

[9] Vivek, R.. Enhancing diversity and reducing bias in recruitment through AI: a review of strategies and challenges.

[10] Yam, J., & Skorburg, J. A.. From human resources to human rights: Impact assessments for hiring algorithms.