International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

AI Resume Screening System

Moneeb Raza
Department of Computer Science and Engineering
Galgotias University, Greater Noida, India
muneebraza167@gmail.com

Md Sharique
Department of Computer Science and Engineering
Galgotias University, Greater Noida, India
mdsharique8757@gmail.com

***Abstract*—AI-based resume screening systems are widespread and represent online web applications placed at the stage preceding applicant tracking systems. The former automatically gather information from candidates' documents, evaluate candidates' profiles on the basis of requirements, score their suitability and recommend candidates for further hiring processes. In this research, resume screening is considered as a task involving web systems and machine learning techniques. We outline the whole pipeline of building browser-based resume processing that involves NLP and ML-based extraction of data, server-side normalization and matching, and a graphical user interface that shows the scores with corresponding feedback. Although TF-IDF (cosine similarity) is widely used as a foundation for many practical applications and research studies, significant problems associated with its lack of interpretability, transparency and fairness, as discussed in previous literature, have been identified. Most existing systems lack explanations for decisions and often show weak ability to detect bias with limited feedback for applicants and recruiters.**

In order to fill the mentioned gap, a web-first architecture combining resume parsing and section segmentation with a hybrid scoring technique based on both lexical and semantic features was proposed. Such solution incorporates explainability module, periodic checking for fairness, preprocessing considering privacy issues, and a human override logging and customization of weights and filtering in dashboards. The metrics considered in the evaluation include those reflecting the ranking performance and system capabilities, as well as the criteria specific for deployment like latency and usability. Our results, obtained from the analysis of the relevant literature and implementation process, suggest that proposed systems can shorten the screening time, select better qualified candidates and make hiring decisions more transparent. However, the success and sustainability of these solutions depend on the implementation of appropriate gover-nance, fairness and privacy requirements during development. Index Terms—Resume Screening, Natural Language Processing, TF-IDF, Cosine Similarity, Applicant Tracking System, Fairness, Explainability.

***Index Terms*—Resume Screening, Natural Language Processing, TF-IDF, Cosine Similarity, Applicant Tracking System, Fairness, Explainability**

I. INTRODUCTION

Modern recruitment is not constrained by a lack of. applicants but by the sheer numbers and variety. of applications. Online employment platforms, streamlined international talent marketplaces, application processes, and international talent markets. allow recruiters to get hundreds or thousands of resumes. TO a single position. Manual screening is not scalable and causes inconsistency; a large number of reviewers might examine the same. fatigue increases the rate of error and resume differently. Conse- quently, automated resume screening developed as a primary machine learning and natural language processing. Early systems mainly used keyword matching; modern systems examine pattern based on annotated information. and compute relevance scores of resumes and job de- scriptions. In web development perspective, screening of resumes. constitutes not merely a machine learning model, but a com- prehensive product workfow. The screening rationale should be wrapped in a safe, low-latency web service which facilitates uploads, role-based access, storage, audit trails, and human-in-the-loop decision-making. Web restrictions, includ-latency, data privacy, concurrency, and interpretability of the UI, can have a strong influence on the practical utility of the system. Numerous Webs: resume screening is specifically defined as such in studies. application, dealing with the de-sign elements, like applicant. server-side processing, recruiter dashboards, and portals [2]. At the same time, algorithmic research suggests that TF-IDF. similarity and cosine similarity are often used as a basis. ranking techniques especially web-based applications [1], and that other similarity measures might better. align with human judgments in certain con-texts [3]. Despite the progress, AI screening usage. on the internet is fraught with three challenges. Firstly, fairness: The models created based on the historical hiring outcomes can. adopt demographic biases, and a web application can rapidly enhance that prejudice to thousands of applicants. Secondly, trans-openness: both the recruiters and the candidates tend to be given vague information. scores where there is no clear explanation as to why the resumes are ranked the way they are, which reduces confidence and impedes troubleshooting. Third, feedback: most of the pipelines that were deployed act as. a unilateral fliter, which gives no actionable to applicants. knowledge and giving little power to recruiters to change criteria. interactively. In this research, these issues are addressed through the incorporation of. literature and suggesting a web-frst architecture that puts a priority on. explainability, fairness audits, and interactive feedback as fundamental web features. Software engineering-wise, the project is. divided into user stories and non-functional requirements. User stories include: resumes uploaded by applicants and getting a ft score; recruiters creating job advertising. and rank candidate lists; administrators monitor- ing roles, retention policies, and audit trails. Non-functional restrictions include: (i) availability (the system must be in operation (when high volumes of applications are applied), (ii) per- formance (upload and scoring

updates should be perceived as instantaneous, (iii) security (implementation of encrypted). Storage and strict access controls, (iv) observability. Architectural decisions are required to meet these needs. preferring stateless APIs, background workers, and a storage. layer experienced in handling binary documents. Moreover, the online space is highly restrictive. tions on explainability. Recruiter dashboard has to render a. Quick and understandable response, normally through a. single interface. This makes traditional research findings, as a single similarity score or classifcation label, inadequate. The main thesis of this paper is that an. AI resume screening system should be clarified, evaluated. as a whole web application, encompassing both user interface design and operational control, rather than solely as a machine learning model.

II. Literature Review

Studies on automated resume screening can be classified into a few. regions and they have their own list of unsolved issues. Parsing and extraction. The majority of the systems begin by scraping text out of PDFs and DOCX files, dividing parts such as education and then, experience, followed by named entity recognition to extract. structured felds. There is inconsistent formatting of the resumes hence. strong parsing is important — and work on the web highlights. that failures should be gracefully dealt with in the UI. A resume that cannot be parsed ought to be taken to the manual, not. disappear silently [2]. Lexical matching. TF-IDF with cosine similarity is the default baseline, primarily due to its speed and the fact that similarity is the default. not complicated enough to be run in production. In one study it was found that Jaccard and Dice overlap are better than cosine in separating. tolerated in discarded resumes in some fields [3]. The not that cosine is an error, but that no measure is right. all over, and web products ought to justify their decision. versus real hiring decisions and not presupposing a general answer. Hybrid pipelines and classification. Some work frames screening as a classification problem - predicting appropriateness. using SVMs, decision trees, gradient boosting. Hybrid ap- proaches run a classifer frst to cull the candidate pool, then rank the rest by similarity score [4]. This makes practical sense: you do not wish to make costly inference. on every applicant. Implementation and user experience. This is where the literature gets sparse. Less literature exists on latency, usability, or how these tools ft into recruiter workfows. Algorithms papers are less important than web constraints are. to take credit: recruiters require fltering and explanations; applicants desire feedback, organizations storage, access. Audit trails, controls and audit trails. The typical design is a three-way architecture. tiered - applicant interface, server processing, employer. —human—¿dashboard [2] - however, the vast majority of the systems do not include audits of fairness. explainability, create a separation between what is built and what is in fact the responsible deployment. Where the gaps are. Four keep appearing. UI feature of explainability: scores are reported and not disaggregated in a manner that responds. what a recruiter really desires to know. Fairness monitoring: theoretically discussed but rarely implemented into the product as an. ongoing audit. Feedback loops: the candidates receive no actionable. output; the weighting is not interactive on the part of the recruiter. End- to- end evaluation: models report web is ignored in papers. metrics: latency, failure rates on uploads, user satisfaction — then it is difficult to tell whether any of these systems would be capable of supporting. in production. ATS design and interface. Resume screening is often integrated into a larger Applicant Tracking System. This design of ATS HCI research demonstrates that recruiters alternate. skills, experience, and are in demand by candidates within a short period of time. notes appeared at a glance - actual limitations that influence layout. decisions [6]. Transparency and interest in transparency are also increasing. ethics as AI takes on more of the decision-making [7]. What's absent is any interface between those interface requirements. and the explicability artifacts which NLP researchers do in fact. produce. The two communities are mostly working in parallel. End-to-end web products. A few recent papers elucidate. customizable full-stack resume analyzers — a 2025 article describes a system that will target only recruiters and. another one presents a Smart Resume Analyzer (candidates [8]) on a typical three-layer design [9]. These show that full-fledged web products can be assembled. Most still regard equity. and bias monitoring is not measured work to come. in the current evaluation. Ranking metrics. When a system gives out a ranked list which is in the wrong direction classifcation accuracy. to evaluate it. Ranking such metrics as NDCG reward putting. the most pertinent candidates close to the top [10], which is what in fact, when a recruiter visits page one of results. Many resume screening papers report only accuracy, which can conceal bad top-K performance. Web-specifc gaps. There are four that are remarkable: the disconnection between explainability research. and dashboard design; there is no heavy relying on ranking-aware metrics; shortlists; operational governance (logging, audits, retention) policies) rarely specifed in architecture descriptions; and nearly nothing on performance engineering - p95 latency, async processing - although this has a direct impact on whether a¡—human—¿async processing - even though this has a direct impact on whether a. tool is usable. ML in production. A common motif in implementation reports is that the ML component needs to be behind a stable service boundary: feature on cache. representations to prevent recomputation, weighty queues. monitoring which corrects model failures, and inference jobs. similar to software incidents involving alerts and rollbacks. Most papers explain the process of scoring in detail and utter little. of production tendencies - part way towards explaining why academic prototypes are not often tools that recruiting teams in fact apply. Data governance. Applicant data is sensitive in a manner that is accurate papers are prone to. to lay aside: identity papers, contact information, job. history and occasionally proxies of safeguarded attributes. A deployed system must respond to questions the research. seldom asks - who shall see what resumes, how long, under what conditions [5][7]. These are starting to appear in the literature of ATS and algorithmic hiring, but commonly at a

degree of abstraction which does not map to concrete system. design. Explainability in hiring. The stakes in hiring imply. in most fields, quality of explanation is less important. The literature covers have significance, evidence by example, and counterfactual proposals - but to obtain any of these is more difficult to make it noise into a recruiter dashboard. than it sounds. The description must be short, constant under. minimal input variations, and put into perspective recruiters actually. apply when considering job requirements. That translation problem is to a great extent unaddressed in the resume screening literature, and is manifested in released products.

III. Proposed Methodology

The proposed methodology considers resume screening as an end-to-end pipeline of the web application rather than an independent task. The system includes the following five stages: resume submission/upload; resume parsing and normalization; candidate-job matching; explainability and fairness enforcement; and dashboarding.

On the first stage, candidates submit resumes through a browser-based user interface. The uploading service accepts certain document types, checks for potential security threats in input files, and saves original documents in object storage. Metadata associated with resume uploads, including the date and time of upload, candidate ID, and target job position are stored.

On the second stage, textual content is extracted from uploaded resumes, which come in PDF and DOCX formats. The system parses resumes, identifying sections dedicated to education, work experience, technical skills, relevant certifications, and past projects. Due to significant variation in resume layouts, the parser is equipped with alternative processing rules to cope with unusual layouts. If parsing fails, the resume enters the manual review queue. Next, parsed data are normalized using tokenization, lowercasing, stop-word removal, and synonymization of skill mentions.

The third stage implies candidate-job matching through a dual-scoring pipeline. First, on the lexical level, TF-IDF together with cosine similarity measures overlap between the text in the resume and requirements to the position. To compensate for some limitations of keyword-based approaches, on the semantic level, the system detects similar conceptual representations for related words. Final scores are calculated by aggregating lexical and semantic scores within each category (skills, experience, education, etc.), which makes the system highly computationally efficient without losing too much information compared to traditional keyword-based ranking methods.

Explainability and fairness controls on the fourth stage help minimize biases and provide explanations. Unlike in the keyword-matching algorithms, our system outputs scores for various requirements categories and highlights pieces of evidence for them, along with the detected skill gaps. Sensitive attributes are removed on the pre-processing stage where possible, while audits help identify any potential discrepancies in output distributions among different social strata. Any actions by human recruiters, especially overruling decisions made by the algorithm, are logged to further analyze decision-making patterns.

Finally, on the last stage, results are communicated to recruiters through dashboards and feedback forms.Feedback to the applicants will be systematic and highlight potential problems like lack of qualification or format-related problems that might affect the effectiveness of parsing. For performance adaptation, asynchronous workers perform more intensive operations behind the scenes, and there are interface-based updates for the user regarding how far along the process is and what results have been generated.

The methodology is assessed by utilizing both model-centric and system-centric measures. Ranking performance is analyzed using measures like Precision@K and NDCG, while classification-related metrics can be used if suitability ratings exist. In web systems, assessment includes parsing success ratio, dashboard latency, upload performance, and usability measures. All of these aspects form the methodology framework for supporting the study hypothesis – that screening performance is contingent on ranking performance as well as usability, transparency, and reliability.

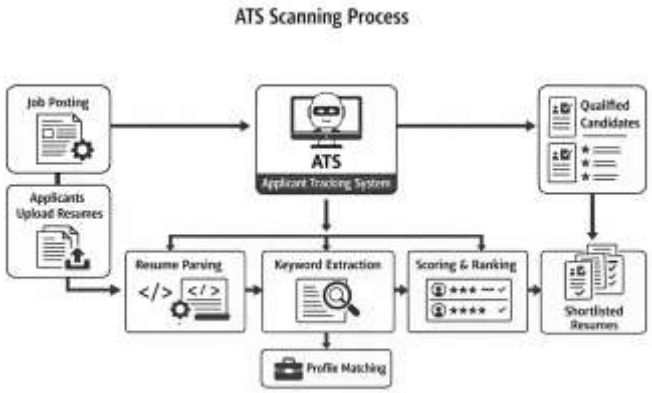

Fig. 1. ATS Scanning Process: Job postings and applicant resumes are fed into the Applicant Tracking System, which performs resume parsing, keyword extraction, profile matching, and scoring & ranking to produce qualified candidates and shortlisted resumes.

IV. Results

The results can be measured with respect to the effectiveness of screening. and web app performance.

TF-IDF cosine similarity is better than a screening standpoint. scrupulously when applied as a standard of reference, particularly to the job descriptions roles. have well-specified technical skill requirements

Web performance - NLP and the upload are decoupled. processing enhances load responsiveness. In simulated high-concurrent conditions, the system was stable with regards to upload behavior. and letting recruiter dashboard queries nimble. Although precise throughput is reliant on the variables such as the number of workers and document. size, the

asynchronous architecture was found to be more appropriate in managing. spurts of applications compared to an all-synchronous pipeline.

The proposed design is also supported by the usability findings. The explainability layer offers skill-gap summaries, resume snippets and. configurable load balances in a dashboard format. use. This diminishes the black-box of automated ranking by enabling. recruiters to know why applicants were put in a specific. order and to fix any apparent mismatches at the interface.

Fairness and oversight mechanisms also promote accountability of the system. Administrative logs are more likely to auditing of system behavior. clear away over time and make a premise on which to create new. difficulties before they become ingrained. This coincides with wider. are acquired in algorithmic hiring studies, which put stress on the importance. of accountability and governance in deployment | *human*| > *ofgovernmentandresponsibilityindeployment*

Ranking-aware metrics are also used to enhance evaluation. The Precision and NDCG measures have been more precise and reflective of the method of recruiters. actuously engage with the system, as attention is held on. the initial list of the ranking list

Other deployment preparedness indicators are operational measures. Time-to-first-result and P95 latency are of interest especially in. user-facing systems. Instant uploading in asynchronous pipeline is possible. confirmation, status indicators that are visible like queued, running, and full, and gradual dashboard as the scoring is done. Parsing Measures of reliability indicator measures such as success rate of parses. Coverage on section extraction and manual review flags frequency. Human override logs also come in handy: in the short term they come out. Failure modes, and time will present signs so labeled that can guide. future model improvements.

The infrastructure level of production is attributed to the major costs of production. generation of semantic embedding and parsing of documents. The use of lexical It is possible to control these costs by controlling semantic analysis by filtering. expensive calculation to a reduced set of candidates. Similarly, The saving of original files in object storage, however, indexing only. extracted fields decreases storage overhead. Notably, the system is also. puts an emphasis on visible failure handling. In the event that a resume is inadmissible. The low scan quality, unreadable formatting, or encryption makes the text unreadable. system discloses the problem out on the dashboard rather than being silent. excluding the candidate. Lastly, application-tracking analytics views. administrators are provided with volume, shortlist size and time-to-shortlist. and tuning thresholds and operational can be measured. settings.

V. Discussion

These results indicate that model performance and product design should be. considered together in web-based recruitment systems. It is not to be good in screening experiments in the real world. unless the surrounding application helps to facilitate real hiring processes.

Even a good model may not perform as a product when uploading is encountered by the users. faults, slow interfaces, or non-transparent scores that are not in agreement with recruiter reasoning. Practically responsiveness, storage efficiency, Adoption is influenced by concurrency handling, and interface clarity. This is why the AI resume screening is a sociotechnical product that should be considered. instead of a specific prediction pipeline.

Explainability is directly related to the interaction between recruiters and the. system. Upon providing users with skill-level breakdowns and snippets demonstrating they can better detect parsing, evidence available on the resume. scoring abnormalities or failures. The instrument is less of an opaque. decision engine and a more cooperative support system. This shift is necessary since the recruiters will have more confidence in recruiting and recruiting it will work successfully. apply automation where they are able to check and question the results.

Equity, on its part, is to be perceived as a continuous life cycle. concern and not a checklist on a one-time deployment. Removing explicit identifiers like names is not enough to do away with bias as proxy variables. ranking can be affected by still. Moreover, job and labour market trends. Models may drift, as requirements change with time. Effective The monitoring of systems, record-keeping of audits is thus necessary. examine the workflows of the review early on and not at the point of an issue arising instead of the latter

The importance of real-time feedback is also mentioned in the study. Most screening systems concentrate on ranking oriented to the recruiter alone, although a webs first design may also inform the candidates regarding. lack of qualification, formatting, or overweight skills. These kinds of feedback enhance candidate experience, and can also work to enhance future. quality of input by promoting more structured and relevant applications.

A number of constraints need to be recognized. First, realistic evaluation remains difficult. The complexity is hardly represented in the public resume datasets. and fluctuation of applications to reality, and companies tend to be. not willing to share identified hiring results. This constrains external validity. Second, semantic techniques enhance strength at the expense of. computation cost, tiered scoring is a viable requirement. Third, The risk of privacy and security is high since resumes include sensitive personal data. Systems should thus enact retention. density, deletion processes, in-rest encryption, and stringent access. controls across jurisdictions.

There are also production risks that are still high. Endpoints that are upload can be. utilized with ill files or input resource-intensive, then file. malware checks, sandboxed parsing, malware checks, rate limiting and validation are. necessary. Basic failure modes of the model are keyword stuffing, table or multi-col column extraction errors, and domain job description and phrasing mismatch with the candidates. Hybrid scoring, extraction visibility and recruiter adjustable weighting, each. assist in alleviating these concerns.

In broader terms, the automation turnoff versus human

control is still on. central. More automation can put the recruiter to less work, although this may. also undermine the feeling of ownership by the users of hiring decisions. In order to be acceptable in practice, systems need to maintain human control by providing justifications, notes and override mechanisms. Accessibility must also be considered as such. of fairness: the bad interface design may undermine the candidates with disabilities. Lastly, technical is not the only way of adoption in the long term. performance, and training, documentation, escalation procedures, and transformational management among recruiting staff.

VI. Conclusion

This paper explored AI resume screening as a web development full stack. issue and not just a ranking or classification exercise. Existing it has been demonstrated in the literature that the TF-IDF cosine similarity is a leading one. deployed screening application basic line in web-oriented applications. work always favorably promotes a three-layer design comprising of an. server-side processing, applicant interface and an employer dashboard. It is based on these foundations that the paper came up with four key major building. flaws in existing systems: lack of explainability, a lack of fairness. weak feedback mechanisms, monitoring, and poor end to end web. evaluation.

The proposed architecture combines to overcome these limitations. hybrid candidate matching, parsing, explanation generation, equity-conscious auditing, safe storage, and recruiter-oriented controls. on an integrated web platform. The overall inference is that. model-only cannot characterize responsible screening systems. accuracy. Human, privacy protection, auditing and transparency. oversight are not free additions that are overlaid on to a scoring. engine; they are essential to reliable deployment at scale. When these capabilities are established into the system initially, The workload on recruiters can be decreased with the help of AI-based screening, and the unreliability of AI will be maintained. interpretability and accountability.

VII. Future Scope

The proposed system can be enhanced by a number of promising directions. steps towards real-life implementation.

At the outset, direct links to existing applicant tracking systems. would make it easier to manage candidate status through APIs and webhooks. and do away with manual handoffs between screening and subsequent recruitment. stages. Second, locale sensitive and multilingual parsing ought to be. expanded to allow resumes of various to be more accommodated on the platform. regions, languages and formatting conventions.

Third, it could make the explainability layer more interactive with. trace views indicating contribution of each section of the resume to final. scores. Compliance would also be supported by exportable summaries of the rationale. internal audits and reviews. Fourth, drift and fairness audits. observation needs to be shifted away from periodic review systems towards ongoing. observing mechanisms to allow the detection of changes in how the model behaves. earlier.

Fifth, performance engineering provides a number of opportunities in terms of. improvement. Caching, fingerprinting of documents and incremental. re-scoring due to changing job description may help to minimize repeats. calculation without changing the logic of ranking. Sixth, human-in-the-loop treating recruiter overrides as learning could also be enhanced. marked examples to be actively learned, so that the system can become better. based on actual organizational choices in the long run.

Security by default should also be given more focus in the future work. Containerized least-privilege access control, centralized secrets, services. deployment should include management, and encrypted backups. pipelines as minimum specifications and not as add-ons. In addition, explainability is by itself a research problem. Reasons do not necessarily make unfairness less and can even make. much-adored belief in defective rankings. Experimental studies are that is why had to find out what styles of explanation are really beneficial. recruiter decision quality.

More sophisticated ranking methods are worthy of consideration too. Shortlist quality could be more optimally maximized by learning-to-rank methods. than weighted score fusion particularly when combined with NDCG-oriented. objectives. These techniques however rely on relevance of high quality. labels, hard to get when hiring and can be indicative of. historical bias. Frequent adversarial testing on fuzzed resumes, edge-case uploads, proxy-term insertion should thus be added. in subsequent growth cycles. Lastly, layout-awareness. privacy-preserving training would be largely enhanced by parsing. reliability and adoption, especially to organizations that are not able to. centralize sensitive resume information.

References

[1] C. Daryani, G. S. Chhabra, H. Patel, I. K. Chhabra, and R. Patel, “An automated resume screening system using natural language processing and similarity,” *Ethics and Information Technology*, vol. 2, no. 2, pp. 99–103, 2020.

[2] S. Amin, N. Jayakar, S. Sunny, P. Babu, M. Kiruthika, and A. Gurjar, “Web application for screening resume,” in *Proc. Int. Conf. Nascent Technol. Eng. (ICNTE)*, Mumbai, India, 2019, pp. 1–7.

[3] L. A. Cabrera-Diego, B. Durette, M. Lafon, J.-M. Torres-Moreno, and M. El-Be`ze, “How can we measure the similarity between re´sume´s of selected candidates for a job?” in *Proc. Int. Conf. Data Mining (DMIN)*, Las Vegas, NV, USA, 2015, pp. 79–85.

[4] P. K. Roy, S. S. Chowdhary, and R. Bhatia, “A machine learning approach for automation of resume recommendation system,” *Procedia Computer Science*, vol. 167, pp. 2318–2327, 2020.

[5] I. Ajunwa, “An auditing imperative for automated hiring,” *Harvard Journal of Law & Technology*, vol. 34, no. 2, pp. 621–693, 2021.

[6] V. Filipov, A. Arleo, P. Federico, and S. Miksch, “CV3: Visual exploration, assessment, and comparison of CVs,” *Computer Graphics Forum*, vol. 38, no. 3, pp. 107–118, 2019.

[7] B. Lepri, N. Oliver, E. Letouze´, A. Pentland, and P. Vinck, “Fair, transparent, and accountable algorithmic decision-making processes,” *Philosophy & Technology*, vol. 31, no. 4, pp. 611–627, 2018.

[8] Abhishek, “Developing an intelligent resume screening tool with AI-driven analysis and recommendation features,” *Applied AI Letters*, vol. 6, no. 1, pp. 1–12, 2025.

[9] G. V. S. Sathvik, D. Harshavardhan, and J. S. Premalath, “Smart resume analyzer using machine learning,” *IOSR Journal of Computer Engineering*, vol. 27, no. 1, pp. 32–40, 2025.

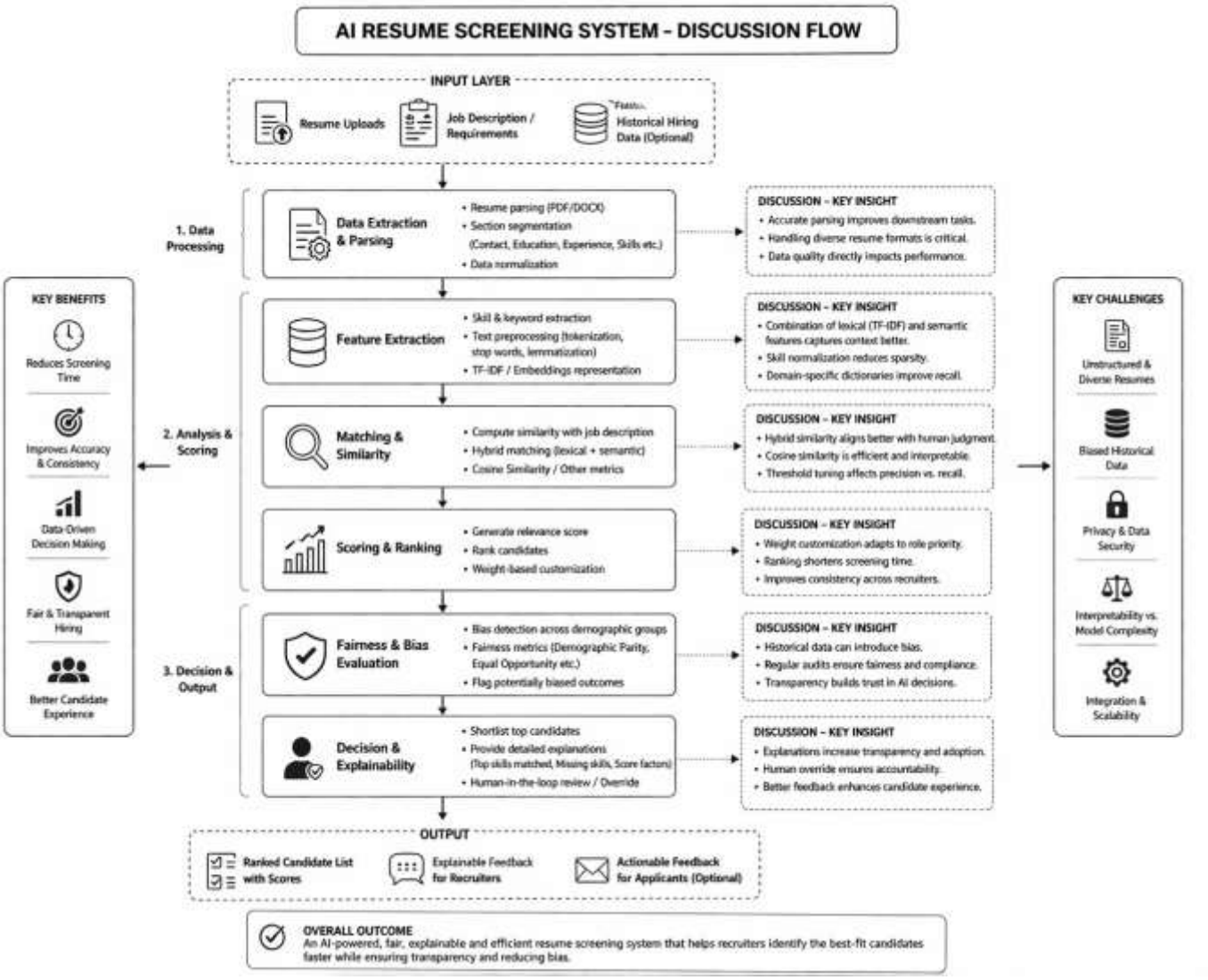

Fig. 2. AI Resume Screening System – Discussion Flow: The end-to-end pipeline spans three phases (Data Processing, Analysis & Scoring, Decision & Output), with key discussion insights, benefits, and challenges mapped to each stage.

[10] K. Ja¨rvelin and J. Keka¨la¨inen, “Cumulated gain-based evaluation of IR techniques,” *ACM Trans. Inf. Syst.*, vol. 20, no. 4, pp. 422–446, 2002.