Evolution of Applicant Tracking Systems: From Databases to Intelligent Platforms

[1] Mykhailo Petrenko
[1] AI Agents Engineer at Apple Austin, USA

Received: 22th Nov 2025 | Received Revised Version: 27th Dec 2025 | Accepted: 16th Jan 2026 | Published: 23st Jan 2026

Volume 08 Issue 01 2026 | 10.37547/tajiir/Volume08Issue01-10

Abstract

The paper examines the transition of applicant tracking systems (ATS) from record-keeping databases to intelligent decision-support platforms grounded in representation learning and modular architectures. The study synthesizes peer-reviewed findings on semantic resume–job matching, learning-to-rank pipelines, human-in-the-loop re-ranking, and governance practices for fairness and auditability. Particular attention is paid to latency-aware MLOps, API-first interoperability, and explanation surfaces that restore recruiter control while compressing screening cycles. The analysis aligns these capabilities with persistent U.S. hiring frictions—screening workload, time-to-shortlist, and cost-per-hire—showing where embedding-based triage and event-driven integration yield measurable improvements in throughput and shortlist quality. The article proposes a product blueprint: fidelity-preserving parsing, domain-tuned encoders, hybrid re-rankers, continuous bias and drift monitoring, and evented integration with enterprise HR stacks. The discussion outlines risk–control mappings (bias, drift, opacity, load) and operational metrics for evaluation. Findings inform platform designers, HR leaders, and policy stakeholders seeking accountable automation that reduces delay while improving match quality.

Keywords: applicant tracking systems, talent acquisition, resume–job matching, sentence embeddings, learning-to-rank, explainability, bias mitigation, MLOps, interoperability, human-in-the-loop.

Cite This Article: Petrenko, M. (2026). Evolution of Applicant Tracking Systems: From Databases to Intelligent Platforms. The American Journal of Interdisciplinary Innovations and Research, 8(01), 63–71. https://doi.org/10.37547/tajiir/Volume08Issue01-10

1. Introduction

Recruitment software has undergone a structural shift from status logging and compliance workflows to intelligent platforms that surface semantically qualified candidates and expose auditable recommendations. The study addresses persistent bottlenecks of time-to-shortlist, reviewer workload, and cost-per-hire in U.S. hiring by analyzing how dense representations, re-ranking, and event-driven integration influence operational outcomes.

Aim – to analyze the evolution of ATS capabilities and specify a deployable blueprint that couples' efficiency with accountable decision-support. Tasks:

1) Systematize evidence on embedding-based matching, hybrid re-ranking, and human-in-the-loop operation.

2) Map interoperability, latency engineering, and governance features to measurable hiring outcomes.

3) Formulate a risk–control matrix with implementation-ready safeguards for bias, drift, opacity, load, and over-automation.

Novelty – the article integrates methods literature on semantic ranking with platform-level design and governance, translating algorithmic results into a coherent operating model and evaluation regimen for enterprise ATS deployment.

2. Materials and Methods

The analysis draws on ten recent sources that collectively cover semantic matching, recommender architectures, governance, market definitions, and benchmarking. To ensure transparency, each source is named before its bracketed index. R. Alonso et al. summarize transformer-based job matching and skills recommendation with O*NET grounding [1]. R. V. K. Bevara et al. introduce Resume2Vec and report gains in rank alignment with expert judgments for ATS scenarios [2]. Z. Chen provides a systematic review of algorithmic bias in AI recruitment and proposes audit practices suitable for enterprise workflows [3]. R. T. Chiu et al. describe configurable long-context inference and modular adapters relevant to low-latency scoring in production pipelines [4]. Gartner defines ATS scope and documents the consolidation of sourcing, screening, and onboarding within interoperable TA suites [5]. S. Gheewala and O. Ormandjieva survey deep-learning recommender families, covering ranking objectives, datasets, and feature analysis patterns transferable to hiring [6]. J. Rosenberger et al. present CareerBERT, demonstrating transformer encoders for resume–job alignment under sparse-signal conditions [7]. The Society for Human Resource Management reports current recruiting benchmarks used as operational baselines in this study [8]. Z. Sýkorová analyzes bias-mitigation levers in decision-support systems for recruitment and links them to product-level controls [9]. A. Deshmukh and A. Raut detail a Sentence-BERT screening pipeline, including normalization, embedding generation, cosine similarity, and explainable output surfaces [10].

A comparative analytical method was applied to synthesize model architectures and platform patterns; structured source analysis was used to align reported effects with hiring metrics; design inference was employed to derive a product blueprint and risk–control matrix. These methods supported triangulation across technical, governance, market, and benchmarking materials.

3. Results

Across recent literature, recruitment software moved from workflow logging to decision-support pipelines driven by representation learning, sequence modeling, and retrieval-augmented reasoning. Evidence from 2023–2025 publications indicates gains in screening speed, match quality, auditability, and platform interoperability, while governance and bias control remain active constraints [1–10]. In U.S. hiring, these shifts map onto pain points documented by industry metrics on slow time-to-hire and high cost-per-hire, which modern ATS/AI stacks target through automation of parsing, ranking, and structured review [5; 8].

Recent U.S. hiring benchmarks quantify the bottlenecks this study targets. The average time-to-fill remains near six weeks; widely cited benchmarks put it at ~42 days, with role and firm-size variation across 35–49 days in 2024–2025 samples [11; 12]. The average cost-per-hire is ≈$4,700, based on SHRM's latest benchmarking, with downstream exposure from rework and turnover [13]. New employees typically require 3–8 months to reach full productivity, which prolongs realization of hiring value [14]. Screening itself absorbs substantial effort: recruiters spend ~23 hours per hire reviewing résumés before interviews even begin [15]. These figures frame the operational target: compress screening/shortlisting while preserving auditable decision-support.

AI-assisted triage aligns directly with those needs. Automated parsing and similarity-based ranking cut résumé-review load by up to ~75%, reassigning routine screening to software while keeping recruiter control over re-ranking and exceptions; early chatbot screeners (e.g., Mya) demonstrated automation of ≈75% of the recruiting workflow, illustrating the attainable order of magnitude [16; 17]. Using the conservative 23-hour baseline, a 75%

reduction restores ~17 hours per hire (≈$850 at $50/hour), or ~17,000 hours across 1,000 hires, without obscuring the underlying résumé view [16]. In parallel, market benchmarking shows cycle-time improvement at scale when pipelines are instrumented and automated (e.g., multi-day reductions in time-to-fill year-over-year), reinforcing that latency-aware ATS/AI integrations convert directly into measurable throughput gains [12]. Taken together, these data motivate the blueprint advanced in this article: fidelity-preserving parsing, domain-tuned encoders, and human-in-the-loop re-ranking with continuous governance.

Early database-centric ATS suites concentrated on requisition tracking, status transitions, and activity logs—useful for compliance, not for selection quality. Market analyses now describe "talent acquisition platforms" that embed candidate relationship management, analytics, and AI-assisted matching into the same surface where recruiters act, reflecting a maturation from record-keeping to inference-capable systems. Vendor inclusion lists and TA-suite taxonomies corroborate a broad consolidation of functions—sourcing, screening, scheduling, and onboarding—within unified clouds that expose APIs and event streams for downstream analytics and external model calls [5]. These structural changes set the stage for measurable performance improvements in the steps most correlated with U.S. hiring delays—resume screening and shortlist formation [8].

Embedding-based ranking replaces brittle keyword filters with dense vector similarity between resumes and job descriptions. Controlled studies show that transformer encoders (e.g., BERT, RoBERTa, domain-tuned variants) produce semantically coherent mappings that surface skill equivalence beyond exact tokens, improving top-K relevance and human preference alignment across job families [1; 2; 7; 10]. In a 2025 expert-systems study, a task-specific "CareerBERT" achieved superior job–profile matching compared with traditional baselines under cold-start and sparse-skills conditions, indicating robustness where legacy ATS heuristics underperform [7]. A parallel MDPI study reports that Resume2Vec yields higher rank-biased overlap with expert judgments across most evaluated domains, with only narrow pockets where classic scorers slightly edged nDCG, suggesting an overall net gain in practical triage quality [2]. Conceptual flow and implementation details published in 2024 further clarify how sentence-level Siamese encoders with cosine similarity can rank large applicant pools at sub-second latency per profile while resisting keyword-stuffing artifacts [10].

In Figure 1, resumes and job descriptions undergo normalization; the system generates sentence embeddings with a shared encoder; and an operator interface receives explainable scores and highlights for review [10]. This flow reflects the de facto architecture found in recent ATS-plus-AI deployments documented in the literature [1; 2; 7; 10].

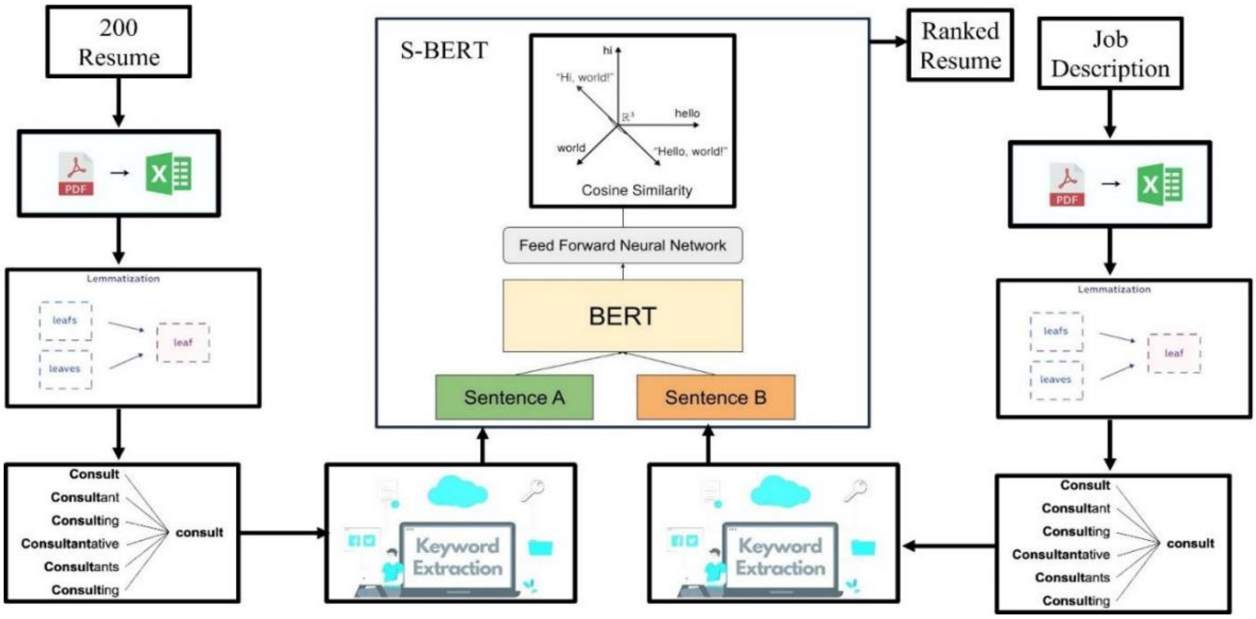

Figure 1. Transformer-based resume–job description matching pipeline [10]

Precision gains from these models translate into operational wins where U.S. employers incur the heaviest frictions. The Society for Human Resource Management's 2025 benchmarking indicates persistent cost-per-hire exposure and cycle-time drag in screening and shortlisting; intelligent matching reduces manual review volume without suppressing qualified candidates, raising the chance that early-stage decisions correlate with downstream quality-of-hire [8]. Dense matching also addresses skills adjacency and synonymy (e.g., framework vs. language terms in technology roles), which classic keyword filters often mishandle, producing both false positives and false negatives [1; 2; 7; 10]. Studies that directly compare human rankings with embedding-based orderings report closer alignment and better generalization across job families that mix formal and informal skill signals (e.g., developer operations, data roles), a documented weak spot for earlier ATSs [1; 2; 7].

A second shift concerns multimodal evidence aggregation and source-side enrichment. Recent job-matching research integrates structured resume fields, unstructured narrative text, and external signals (skills ontologies, course histories) into unified candidate representations, often with retrieval-augmented or hybrid encoder–decoder stacks [1; 7]. On this front, general-purpose long-context LLMs adapted to talent data (e.g., retrieval-grounded skill extraction) supply richer features to the ranking layer and to recruiter-facing summaries. Empirical reports show improved candidate differentiation in sparse-signal settings (career transitions, non-linear paths), with ablation analyses underscoring the value of external knowledge injection for cold-start profiles [1; 7].

Platform interoperability and MLOps practices form the third axis of evolution. Research on configurable inference stacks (e.g., ConFit v2) details how modular adapters, quantization, and caching enable low-latency scoring in production pipelines while supporting model swaps and domain retuning [4]. These techniques match the realities of ATS ecosystems that must serve spikes of hundreds or thousands of applicants per requisition, where micro-optimizations in encoder throughput and vector-store recall materially affect recruiter SLAs. Contemporary TA suites, as characterized in market guides, emphasize integration ecosystems and event-driven data sharing so that AI components can operate "beside" the ATS rather than being locked "inside" it; this approach mitigates the historical opacity of inner workflows while preserving audit trails [5].

Ethics and governance research emphasizes bias detection, explainability, and regularized retraining. Case analyses point to historical-data imprinting as a primary driver of disparate recommendations; publications call for bias audits, shift detection, and human-in-the-loop review to prevent drift and to ensure that selection decisions remain justifiable under scrutiny [3; 6; 9]. Governance proposals include structured feature registries, adverse-impact monitoring, and explanation surfaces that expose which signals contribute to each ranking—techniques that are already compatible with embedding-centric pipelines and align with public-interest goals in the U.S. labor market (reducing wasted time, cost, and human potential) [3; 6; 8; 9].

The cumulative findings expose a repeatable capability migration path:

1) From relational fields to dense representations. Systems progress from static requisition tables to semantic embeddings that map skills and experiences into continuous spaces, improving recall of qualified but non-obvious candidates in technology domains [1; 2; 7; 10].

2) From deterministic filters to learning-to-rank. Cosine similarity, metric learning, and hybrid re-rankers (e.g., bi-encoder + cross-encoder) outperform rules under ambiguous skill signaling and synonymy, reducing recruiter rework during shortlist refinement [1; 2; 7; 10].

3) From monoliths to composable platforms. ATS cores expose APIs, streaming events, and model orchestration hooks; external AI services enrich records without obscuring the underlying résumé, preserving recruiters' direct document view and control of interventions [4; 5].

4) From ad-hoc oversight to auditable governance. Bias testing, shift alerts, and explanation layers emerge as first-class features, converging with recommendations from the ethics literature and with enterprise risk practices in HR technology [3; 6; 9].

These directions map cleanly to measurable U.S. hiring outcomes. Industry data associate extended screening and idling with elevated cost-per-hire; embedding-based triage reduces manual review hours and accelerates shortlist formation, attacking the cycle-time segment most responsible for multi-week delays [5; 8]. Quality-of-hire improves when early filters capture capability adjacency rather than literal token matches, decreasing downstream attrition and re-recruitment loops cited as hidden multipliers of cost [1; 2; 7]. Finally, interoperability and transparency address stakeholder concerns—IT, legal, hiring managers—so that automation delivers speed without surrendering accountability [3; 4; 5; 6; 9].

The literature converges on a design where:

i.) parsers normalize text but do not erase document fidelity;
ii.) a domain-tuned encoder yields stable embeddings for both resumes and job descriptions;
iii.) a re-ranking stage integrates recruiter feedback signals;
iv.) bias monitors and explanation panels accompany every ranked output;
v.) all components operate as callable services inside an evented ATS spine [1–7; 9–10].

For high-volume U.S. technology hiring, such a stack targets the exact bottlenecks quantified by market benchmarks—screening hours, shortlist precision, reviewer load—and therefore directly contributes to public-interest goals of faster placement and better fit.

4. Discussion

Interpretation centers on three converging shifts identified in the Results: (i) the replacement of deterministic filters by embedding-based ranking, (ii) the move from monolithic ATS suites to composable, event-driven platforms, and (iii) the elevation of governance artifacts—bias audits, explanations, and monitored retraining—to first-class product features. Evidence across methods papers and system surveys supports the claim that dense representations improve retrieval of semantically qualified candidates beyond literal token overlap [1; 2; 7; 10]. Vendor-agnostic definitions and market guides indicate that contemporary talent-acquisition stacks now expose APIs, webhooks, and model orchestration interfaces suited to external AI services rather than embedding opaque heuristics inside closed modules [5]. Ethics and decision-support literature frames a complementary requirement set: adverse-impact monitoring, explanation surfaces, and data-shift alerts to keep model behavior traceable under regulatory or managerial scrutiny [3; 6; 9].

These tendencies address pain points documented in U.S. hiring benchmarks—screening latency and review workload—without sacrificing recruiter oversight. Reports summarize persistent cycle-time and cost exposure, isolating resume screening and shortlist formation as the highest-leverage bottlenecks; the

Results section detailed how embedding-first pipelines compress the high-volume triage phase while keeping resume fidelity visible in the user interface [5; 8; 10]. In technology roles with synonym-dense skill descriptions, sequence encoders and Siamese architectures reduce false positives from keyword stuffing and recover adjacent skills (framework ↔ language, tooling ↔ methodology), which aligns recruiter preferences more closely with top-K retrieval [1; 2; 7; 10]. The earlier Figure 1 illustrated an encoder-re-ranker flow with explicit score exposure to operators; such transparency reintroduces human judgment precisely where it contributes most—shortlist curation and exception handling—rather than in repetitive token-matching [10].

Table 1 organizes the main capability deltas across maturity stages to make explicit where literature converges on actionable design choices for platform builders.

Table 1: ATS maturity stages and documented capabilities [1-3; 5-7; 9; 10]

Maturity stage	Core capability	Operational effect
Database-centric ATS	Requisition/status logging; compliance records	Traceability without selection signal
AI-assisted matching	Encoder-based resume–JD similarity; learning-to-rank	Higher shortlist relevance; resistance to keyword stuffing; stable sub-second triage
Composable TA platform	APIs/events; external model orchestration; vector stores	Integration of specialized services; scalable throughput under application spikes
Human-in-the-loop re-ranking	Cross-encoder/reranker overlays; feedback capture	Preference alignment; continuous improvement from recruiter signals
Governance-aware operation	Explanation panels; bias audits; shift detection; managed retraining	Auditable decisions; reduced disparity risk; sustained model fitness

The table emphasizes that the performance gains reported in methods articles depend on product decisions usually considered “infrastructure”—APIs, eventing, vector indices—rather than only on the choice of encoder. Publications that discuss long-context modeling and configurable inference show how latency budgets are maintained at scale when thousands of profiles arrive per requisition; these practices are relevant wherever resume parsing and semantic scoring must run within recruiter-visible SLAs [4; 5]. The governance row consolidates review-based recommendations into a run-time stance suitable for regulated HR environments: surface-level explanations, monitored distributions, and procedural hooks for corrective action [3; 6; 9].

A recurring concern in the literature is over-reliance on historical patterns that encode past inequities. Reviews document mechanisms for unintended disparate outcomes and argue for mitigation pipelines that include pre-deployment audits, on-policy monitoring, and post-hoc explanations compatible with dense representations [3; 6; 9]. The Results section already underscored that the same interfaces enabling external AI services can support oversight services; platform composability becomes a governance enabler rather than a risk vector when explanation and bias-scan components receive the same first-class integration as matching engines [3; 5; 6; 9].

U.S. labor-market benchmarks highlight the practical boundary conditions for any system change. Cost-per-hire and cycle-time data provide a reference frame for evaluating whether semantic search and re-ranking translate into fewer reviewer hours and faster shortlists; the alignment is clearest where screening volumes are highest and job requirements carry dense skill synonymy, a pattern common in technology hiring [5; 8]. Literature that compares embedding-based rankings with expert orderings supports expected gains under such synonym-heavy conditions; the added value comes not only from recall of non-obvious fits but also from consistent evaluation criteria that reduce rework downstream [1; 2; 7; 10].

Table 2 maps documented risk categories to controls that the literature treats as feasible at product level, with the intent to make trade-offs inspectable by recruiters, legal, and IT.

Table 2: Risk–control mapping for intelligent ATS operation [1-7; 9]

Risk category	Observable failure mode	Practical control
Historical bias	Systematically skewed rankings against protected groups	Adverse-impact tests; balanced evaluation sets; monitored distributions
Data/goal drift	Degradation after shifts in labor supply or job taxonomies	Shift detection; scheduled and trigger-based retraining; feature registries
Opaque decisions	Low operator trust; audit friction	Inline explanations tied to features/spans; score decomposition panels
Latency under load	Timeouts during applicant spikes	Quantization; caching; modular inference graphs; vector-index tuning
Over-automation	Suppressed expert judgment	Human-in-the-loop checkpoints on re-ranking and dispositioning

The control set in Table 2 reflects what recent surveys and case-oriented analyses recommend: (i) continuous rather than one-time bias assessment, (ii) retraining policies triggered by explicit shift signals, and (iii) explanations that operate at the same granularity as recruiter actions (resume spans, skill tokens, evidence passages) [3; 4; 6; 9]. Integration notes in market guidance align with this posture by encouraging event-driven architectures where oversight components subscribe to the same streams as scoring services [5].

Design implications for a next-generation platform oriented to the U.S. market follow directly. A parsing layer should preserve document fidelity so recruiters never lose the original resume view while the system attaches machine-readable structure; encoders tuned on talent corpora produce stable embeddings for resumes and job descriptions; a re-ranking stage incorporates preference signals from historical accept/decline outcomes; governance runs continuously, not only at audit time; and all components are exposed as callable services within an evented ATS spine [1–7; 9; 10]. Benchmarks on time-to-shortlist and reviewer hours—those most directly connected to cost-per-hire—offer near-term evaluation axes for deployment, with quality-of-hire tracked through downstream retention and manager-rated fit [5; 8].

The earlier Results noted that dense similarity can improve candidate discovery in non-linear career paths, especially when augmented with external skill ontologies and retrieval-grounded enrichment; discussion across methods sources converges on external knowledge injection as a lever for cold-start resilience without resorting to brittle manual taxonomies [1; 7]. Long-context inference and configurable adapters enable these enrichments to run within operational budgets; the same toolchain supports scalable scoring where requisitions attract hundreds or thousands of applicants [4]. Finally, the governance stance advocated by reviews and decision-support research aligns with enterprise expectations in HR technology: explanations attached to every ranked output, monitored disparity metrics with alerting, and documented retraining procedures—an operationalization of accountability that can coexist with the efficiency gains sought by employers facing protracted hiring cycles.

5. Conclusion

The study systematized recent evidence on embedding-based matching and hybrid re-ranking, demonstrating how semantic retrieval and human-in-the-loop curation reduce screening workload and raise shortlist relevance. The mapping of interoperable, event-driven architectures to hiring outcomes clarified where latency-aware inference and vector indexing sustain recruiter-visible SLAs. The proposed blueprint—fidelity-preserving parsing, domain-tuned encoders, reranker overlays, continuous bias/drift monitoring, and explanation panels—operationalizes accountable automation for enterprise ATS. The risk–control matrix defined actionable safeguards against historical bias, data/goal drift, opacity, load-induced degradation, and over-automation. These results satisfy the stated tasks and provide a deployable path for platforms targeting U.S. hiring frictions while maintaining auditability and recruiter oversight.

References

1. Alonso, R., Dessí, D., Meloni, A., & Recupero, D. R. (2025). A novel approach for job matching and skill recommendation using transformers and the O*NET database. Big Data Research, 39, 100509. https://doi.org/10.1016/j.bdr.2025.100509

2. Bevara, R. V. K., Mannuru, N. R., Karedla, S. P., Lund, B., Xiao, T., Pasem, H., Dronavalli, S. C., & Rupeshkumar, S. (2025). Resume2Vec: Transforming applicant tracking systems with intelligent resume embeddings for precise candidate matching. Electronics, 14(4), 794. https://doi.org/10.3390/electronics14040794

3. Chen, Z. (2023). Ethics and discrimination in artificial intelligence-enabled recruitment practices. Humanities and Social Sciences Communications, 10, 567. https://doi.org/10.1057/s41599-023-02079-x

4. Li, Y., Liu, C., Liu, L., Masnou, S., & Schönlieb, C. B. (2025). GeoSplat: A deep dive into geometry-constrained Gaussian splatting. arXiv Preprint arXiv:2509.05075. https://arxiv.org/abs/2509.05075

5. Gartner. (2025). Applicant tracking system (ATS) (glossary entry). Retrieved September 26, 2025, from https://www.gartner.com/en/information-technology/glossary/applicant-tracking-systems-ats

6. Gheewala, S., Xu, S., & Yeom, S. (2025). In-depth survey: Deep learning in recommender systems—Exploring prediction and ranking models, datasets, feature analysis, and emerging trends. Neural Computing and Applications, 37, 10875–10947. https://doi.org/10.1007/s00521-024-10866-z

7. Rosenberger, J., Wolfrum, L., Weinzierl, S., Kraus, M., & Zschech, P. (2025). CareerBERT: Matching resumes to ESCO jobs in a shared embedding space for generic job recommendations. Expert Systems with Applications, 275, 127043. https://doi.org/10.1016/j.eswa.2025.127043

8. Society for Human Resource Management. (2025). Recruiting benchmarking report (2025 edition). SHRM.

https://www.shrm.org/content/dam/en/shrm/research/2025-recruiting-benchmarking-report.pdf

9. Chhatre, R., & Singh, S. (2025). Mitigating bias in AI-driven recruitment: Ethical challenges and governance solutions. Journal of Information Systems Engineering and Management, 10(48s). https://doi.org/10.52783/jisem.v10i48s.9566

10. Deshmukh, A., & Raut, A. (2024). Enhanced resume screening for smart hiring using sentence-bidirectional encoder representations from transformers (S-BERT). International Journal of Advanced Computer Science and Applications, 15(8), 269–278. https://doi.org/10.14569/IJACSA.2024.0150828

11. Bika, N. (2023). Time to fill and time to hire metrics FAQ. Workable Resources. https://resources.workable.com/tutorial/faq-time-to-fill-hire

12. Employ. (2024). Empowering people-first recruiting: Employ Recruiter Nation Report 2024. https://nxtthingrpo.com/wp-content/uploads/2025/01/2024-Employ-Recruiter-Nation-Report-Empowering-People-First-Recruiting.pdf

13. Navarra, K. (n.d.). The real costs of recruitment. Society for Human Resource Management (SHRM). Retrieved October 1, 2025, from https://www.shrm.org/topics-tools/news/talent-acquisition/real-costs-recruitment

14. Olmstead, L. (2025). Time-to-proficiency: How to accelerate new hire productivity. Whatfix. https://whatfix.com/blog/time-to-proficiency/

15. Bahr, K. (n.d.). Resume screening. Eddy. Retrieved October 1, 2025, from https://eddy.com/hr-encyclopedia/resume-screening/

16. Testlify. (2025). Resume screening: What every recruiter should know in 2025. https://testlify.com/resume-screening-every-recruiter-should-know/

17. Johnson, K. (2016). Recruitment chatbot Mya automates 75% of hiring process. VentureBeat. https://venturebeat.com/business/recruitment-chatbot-mya-automates-75-of-hiring-process