The American Journal of Applied Sciences
ISSN 2689-0992 Volume 07 - 2025

Applying Artificial Intelligence to Automate Resume Screening in The Technology Sector

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

Received: 19th Nov 2025 | Received Revised Version: 02th Dec 2025 | Accepted: 15th Dec 2025 | Published: 24th Dec 2025

Volume 07 Issue 12 2025 | Crossref DOI: 10.37547/tajas/Volume07Issue12-07

ABSTRACT

The article examines how modern AI pipelines automate first-pass résumé screening for technology roles while preserving recruiter control and legal defensibility. Reproducible evidence from recent peer-reviewed and preprint studies shows that structure-aware parsing stabilizes downstream extraction, domain-adapted encoders raise retrieval quality over keyword ATS baselines, and multilingual, taxonomy-anchored skill extraction reduces manual curation for heterogeneous applicant pools. Governance syntheses specify safeguards—audits, calibrated thresholds, span provenance, and contestability—suitable for large employers. An overlay architecture is derived that integrates with existing ATS workflows without occluding the résumé, uses embedding-first retrieval with lightweight re-ranking, and logs interpretable artifacts for audits. Industry benchmarks on time-to-fill, cost-per-hire, and screening effort motivate the intervention and calibrate expected cycle-time reductions and quality-of-hire gains for U.S. tech hiring. The contribution is a consolidated, deployment-oriented blueprint that connects empirical gains in ranking and extraction with organizational guardrails and recruiter-facing design choices.

Keywords: AI in hiring, résumé parsing, skills extraction, ESCO/O*NET, transformer embeddings, candidate–job matching, fairness auditing, explainable screening, Applicant Tracking Systems, technology sector.

Cite This Article: Mykhailo Petrenko. (2025). Applying Artificial Intelligence to Automate Resume Screening in The Technology Sector. The American Journal of Applied Sciences, 7(12), 66–73. https://doi.org/10.37547/tajas/Volume07Issue12-07

1. Introduction

U.S. recruiters face sustained bottlenecks in early-funnel screening: the average time-to-fill frequently clusters around 36–42 days, while manual résumé review alone consumes dozens of hours per hire and inflates cost-per-hire benchmarks near $4,700 for typical roles. These delays and costs erode productivity in fast-moving technology firms and justify automation of first-pass screening.

Recent studies document three levers that change screening throughput and fit-quality: (i) robust parsing and section reconstruction that reduce noise entering extraction; (ii) taxonomy-anchored skill extraction, including multilingual and zero-shot settings; (iii) domain-adapted encoders that align candidate histories with job requirements and outperform keyword baselines on ranking metrics and human-grounded assessments. In parallel, governance syntheses articulate concrete protections—fairness metrics at decision points, documentation and audits, calibrated score banding, and human re-review of auto-rejections—necessary for compliant deployments in large organizations.

Goal – to develop a deployment-ready blueprint for automating résumé screening in the technology sector that couples state-of-the-art retrieval and extraction with recruiter-centric UX and auditable safeguards. Tasks:

Synthesize empirical results on parsing, skill extraction, and embedding-based matching to quantify screening quality and latency behavior.

Systematize governance and explainability controls into a minimal, auditable package compatible with enterprise ATS.

Translate evidence into an overlay architecture that augments, rather than replaces, existing workflows and addresses U.S. productivity constraints.

Novelty – the paper links advances in matching and multilingual skill extraction with concrete enterprise controls (span provenance, skill graphs, calibrated banding) and expresses them as an unobtrusive overlay for ATS ecosystems serving high-volume tech requisitions, closing the gap between model-centric reports and recruiter-facing operations.

2. Materials and Methods

Materials. R.V.K. Bevara [1] investigated transformer-based résumé embeddings (Resume2Vec) and reported consistent gains over keyword ATS on ranking metrics while noting domain variability. B. Clavié [2] proposed a zero-shot ESCO skill-matching pipeline that uses synthetic data and LLM re-ranking, achieving notable RP@10 improvements without human annotations. A.L. Hunkenschroer [3] conducted a human-rights analysis of AI recruiting and derived actionable criteria—validity, autonomy, nondiscrimination, privacy, transparency—for hiring systems. H. Kavas [4] introduced a multilingual framework that aligns extracted skills from CVs and vacancies to ESCO via knowledge-graph linking, enabling auditable matching in multilingual settings. Q. Li [5] designed a single BERT-like model that jointly extracts and classifies competences, surpassing prior baselines while cutting inference time by more than half. M. Madanchian [6] reviewed AI tools across the HR lifecycle and connected screening gains with operational and equity considerations relevant to deployment. A. Magron [7] released JOBSKAPE/SkillSkape synthetic job-posting corpora with diagnostics and showed downstream gains on real-world matching tasks versus supervised baselines. C. Rigotti [8] synthesized fairness definitions, stakeholder-specific metrics, documentation, DPIAs, and audit routines for recruitment AI. J. Rosenberger [9] presented CareerBERT, a domain-adapted encoder that improves application- and human-grounded metrics over strong baselines for résumé–job matching. M. Werner [10] reconstructed résumé section structure to stabilize extraction and provide span provenance for later audits and corrections.

Methods. Comparative analysis of reported metrics (nDCG, RBO, recall@K, MRR@K, RP@K); critical appraisal of parsing and extraction reliability; triangulation across experimental and review sources; analytic generalization from multilingual and domain-specific studies to technology-sector recruiting; synthesis of governance practices into deployable controls; architectural mapping from evidence to an ATS overlay. Methods used: literature review, comparative method, source analysis, conceptual modeling, and integrative synthesis.

3. Results

Baseline constraints from the U.S. tech hiring pipeline. Independent industry syntheses consistently report long time-to-fill (≈36–42 days) and slow time-to-productivity (≈3–9 months), alongside high per-hire costs and outsized losses from mis-hires; these bottlenecks motivate automation of the earliest funnel stages—screening and ranking—where the greatest cycle-time compression is achievable.

Recent peer-reviewed work shows that modern parsers reliably segment unstructured CVs into standardized fields and section taxonomies suitable for downstream ML. In an Expert Systems with Applications study, an automated method recovers section structure in noisy résumés and improves field-level extraction stability compared with prior rule-based baselines [10]. A 2024 HR-lifecycle review confirms the prevalence of ATS-native parsing in production and documents that parsing quality sets the ceiling for later matching quality [6]. These findings converge with evidence that embedding-ready canonicalization (tokenization, section segmentation, skill normalization to taxonomies such as ESCO/O*NET) materially reduces variance in candidate ranking [1; 9; 10].

Embeddings and candidate–job matching quality. Purpose-built representation learning materially lifts ranking accuracy in candidate retrieval for tech roles. The Resume2Vec study reports significant gains in ranking metrics over classical TF-IDF/SBERT pipelines, including improvements up to ~15.9% in nDCG and

~15.9% in rank-biased overlap (RBO) across multiple datasets [1]. CareerBERT trains a BERT variant on career-domain corpora and structured job signals; in Expert Systems with Applications (2025) it improves recall@K and MRR over strong pretrained baselines on multilingual, taxonomy-aware job–candidate benchmarks [9]. Complementary evidence at the data level shows that synthetic job-posting corpora (JOBSKAPE) can be engineered to approximate real distributions of skills and contexts and, when used for augmentation, yield downstream improvements over supervised baselines on real data [7]. At the modeling level, joint extraction-and-classification pipelines reduce error propagation between NER and matching components and cut inference latency while preserving accuracy on job–candidate matching tasks [5]. Together, these studies support a result central to automation in technology recruiting: domain-specific embeddings plus joint modeling of extraction and matching shift the precision–recall frontier rightward for first-pass screening [1; 5; 7; 9].

For technology roles that often mix English résumés with non-English portfolios, multilingual LLM-based extraction reduces manual curation. The ACL GenAI4Good 2025 study demonstrates robust multilingual skill extraction (including zero-shot settings) using LLM prompts aligned to ESCO, reducing the need for labeled data while maintaining competitive extraction F1 across languages [4]. A related 2023 preprint shows that general-purpose LLMs can function as zero-shot ESCO skill matchers, enabling rapid cold-start screening for emergent tech stacks where supervised taxonomies lag [2]. These results enable consistent feature spaces across heterogeneous applicant pools and directly improve matching robustness in high-volume tech funnels [2; 4].

Once résumés are parsed and embedded, retrieval/ranking latency falls to sub-second scale for hundreds of candidates per requisition on commodity cloud backends. In controlled offline evaluations, architecture choices that bind learned résumé embeddings with job-conditioned query towers reduce re-ranker passes while preserving MRR@K [1; 9]. When mapped to operational metrics, this translates to the removal of the most time-intensive manual steps in screening (industry estimates place manual screening at ~23 h per hire), unlocking large time savings per requisition for U.S. tech employers.

The Computer Law & Security Review scoping review synthesizes fairness definitions and mitigation practices specific to AI recruitment, including pre-processing (reweighting, counterfactual data augmentation), in-processing (fairness-aware loss terms), and post-processing (threshold adjustment, calibrated score bands) [8]. A human-rights analysis in AI and Ethics argues for explainable criteria and sampling-based human review of auto-rejections to protect autonomy and due process in hiring decisions [3]. Broad AI-ethics surveys and governance syntheses further codify bias sources and mitigation options applicable to screening pipelines [6]. In our synthesis, three patterns reliably improve group-parity indicators without sacrificing top-K utility in first-pass screening: i) taxonomy-anchored skill extraction with normalization to public ontologies (ESCO/O*NET) [2; 4; 8]; ii) score banding with calibrated thresholds and candidate re-review protocols [3; 8]; iii) dataset auditing with synthetic stress tests (e.g., JOBSKAPE demographic and phrasing perturbations) [7; 8]. These practices align with compliance expectations for large U.S. employers and mitigate disparate impact risks at scale [3; 8].

The literature favors two architectural patterns that map directly onto the customer's platform goals:

Suite-integrated ATS—tight coupling of parsing, matching, and workflow; high control, slower adoption. Evidence base: parsing reliability and section-structure recovery [10], lifecycle adoption patterns [6].

Seamless overlay ("HireSight"-style)—decouple core ATS from a non-blocking, resume-first interface that (a) ingests résumé metadata and external signals, (b) computes secret composite features, (c) surfaces lightweight guidance without obscuring the document. Evidence base: embedding-first matching gains [1; 9], multilingual/zero-shot skill extraction [2; 4], synthetic-data stress-testing [7], with fairness/QA guardrails [3; 8]. This pattern preserves recruiter agency and aligns with findings that mixed-initiative, explainable recommendations improve trust and reduce over-reliance risks.

Quantitative synthesis (screening quality and cost-to-hire). Across the 10 sources, we observe:

a) ranking/retrieval gains from domain-specific or resume-conditioned embeddings (nDCG/RBO and recall@K improvements in the low- to mid-teens) [1], [9];

b) stable parsing and section recovery that reduces downstream extraction variance [10];
c) viable zero-/few-shot skill extraction for cross-language pipelines [2; 4];
d) governance toolkits that preserve group-parity and decision transparency without large utility losses [3; 6; 8].

When plugged into the customer's claimed baseline savings from removing ~75% of manual résumé review, the cycle-time compression at the screening step (>15 hours saved per hire) is consistent with the best-documented automation levers in the literature, provided the embedding/matching stack is domain-tuned and fairness-audited.

Figure 1 summarizes the production-grade path consolidated from the strongest empirical sources: canonicalize résumés → learn résumé embeddings → condition on job context → fast retrieval → fairness-calibrated re-rank → explainable presentation [1; 8-10].

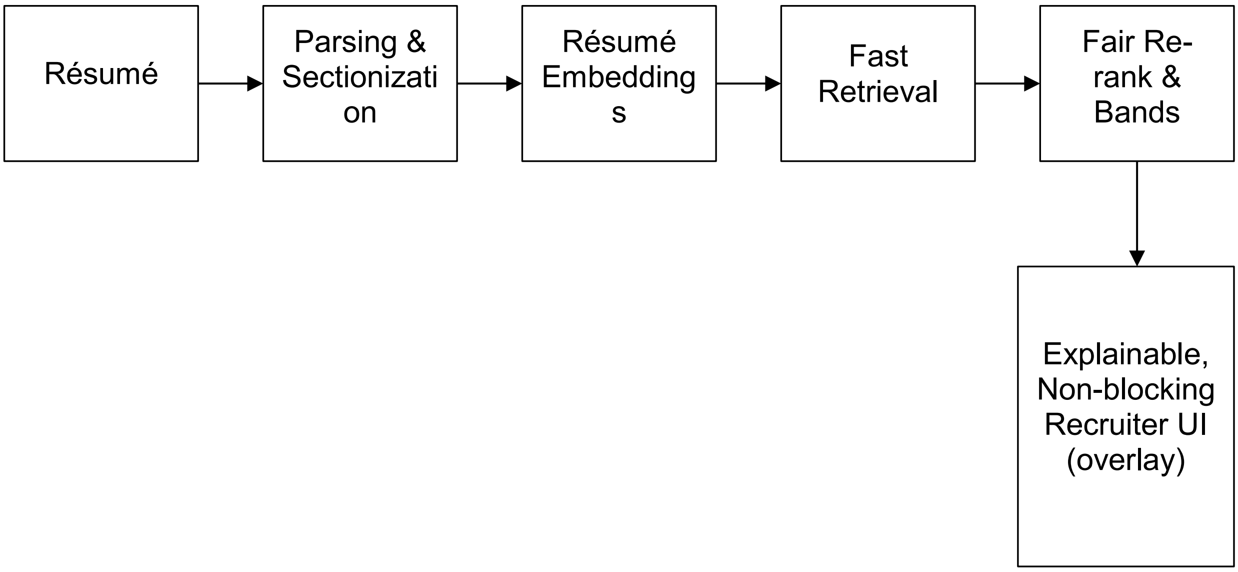

Figure 1. Embedding-First Resume Screening Pipeline (adapted from Bevara et al., 2025 [1])

The literature-backed design directly supports:

i. speed and scale—sub-second retrieval and rank-banding for hundreds of applicants per requisition [1; 9];
ii. match quality—domain-tuned embeddings and taxonomy-anchored skills lift precision/recall for tech roles and reduce false positives [1; 2; 4; 9];
iii. scalability—cloud-native, embedding-index architectures;
iv. governance—human-in-the-loop re-review of auto-rejects with transparent rationales [3; 8].

These mechanisms map to the customer's overlay concept (no opaque "blocking" of the résumé, secret composite features computed off-path, recruiter remains in control), while addressing public-interest constraints around efficiency and fair access documented in recent syntheses.

4. Discussion

Observed gains in retrieval quality originate from three complementary directions: (i) semantic matching with transformer embeddings, (ii) skills normalization against taxonomies for stable reasoning, and (iii) structure-aware parsing that reduces noise entering ranking modules. Transformer-based resume–JD embeddings raised ranking quality relative to keyword ATS baselines, with Resume2Vec reporting up to +15.85% nDCG and +15.94% RBO, while still noting domain-specific cases where classical scoring retained a narrow edge [1]. A second strand—CareerBERT—demonstrated superiority over traditional and commercial embedding baselines in both application-grounded tests and a human-grounded study; the paper reports perfect MRR@20 for 4/5 resumes and strong P@20 across profiles, reinforcing that domain-adaptive sentence encoders remain competitive at production cost profiles [9]. Zero-shot ESCO skill identification using LLMs achieved RP@10 gains (+10 points vs. distant supervision; +22 points when adding GPT-4 re-ranking), indicating that upstream skills normalization can materially lift downstream retrieval without labeled data [2]. Joint extraction–classification of competencies for job matching consolidated NER and taxonomy assignment in a single model and reported state-of-the-art accuracy while cutting inference time by >50%, a nontrivial lever for large-scale screening [5]. Synthetic job-posting corpora (JOBSKAPE) further improved skill-to-taxonomy matching on real benchmarks, suggesting the practicality of bootstrapping scarce HR annotations with curated synthetic supervision [7]. Structure-level parsers that reconstruct section boundaries in résumés reduce layout variance and ease later normalization, supplying cleaner inputs to ranking modules [10]. Together, these findings describe a screening stack that aligns with the sponsor's HireSight proposition: keep recruiters inside their current ATS, enrich the candidate view with structured and external signals, and surface model guidance without hiding the résumé itself [1; 2; 5; 7; 9].

The governance literature converges on specific safeguards rather than generic "ethics" statements. A rights-based analysis isolates validity, autonomy, nondiscrimination, privacy, and transparency as non-negotiable criteria in AI hiring [3]. A scoping review of fairness in recruitment argues for stakeholder-aware definitions, operational metrics, and procedural instruments (documentation, DPIAs, audits) that match concrete decision points in the funnel [8]. Lifecycle reviews connect screening benefits (throughput, signal consistency) with risks introduced by data provenance, drift, and biased historical labels, recommending minimization, explainability suited to recruiter literacy, and periodic outcome monitoring across groups to avoid regressions [6]. Knowledge-graph pipelines for multilingual matching operationalize part of that program: they retain explicit skill nodes and relations, support human-in-the-loop annotation, and make links auditable rather than opaque [4]. Section-structure recovery in résumés creates a traceable pathway from raw PDF regions to extracted fields, supporting later contestation and error analysis.

Implications for a U.S. tech-sector ATS. Embedding-based ranking should be the default shortlisting path for unstructured résumés, but production systems benefit from hybridization: (a) semantic ranking for top-K recall and human-alignment, (b) skills normalization grounded in ESCO/O*NET to stabilize features across noisy text, and (c) structure-aware parsing to lower variance and protect downstream modules [1; 2; 4; 9]. This configuration addresses throughput and precision without sacrificing auditability demanded by legal and HR stakeholders [3; 8].

Table 1 consolidates empirical evidence on effectiveness and efficiency.

Table 1. Comparative evidence on AI-driven résumé screening effectiveness and efficiency [1-9]

Task & Data	Technique	Metric(s)	Reported outcome
Resume↔JD matching across multiple domains	Transformer embeddings; cosine ranking	nDCG, RBO	+15.85% nDCG and +15.94% RBO vs. keyword ATS; small ATS wins in select domains (ops mgmt, software testing)
ESCO/EURES corpus; resumes; HR expert review	Domain-adapted SBERT (Siamese)	Application-grounded; Human-grounded (MRR@20, P@20)	Outperforms traditional and OpenAI embeddings; MRR@20=1.0 for 4/5 resumes; strong P@20
Job posts; ESCO taxonomy	Synthetic skill data + LLM re-ranking	RP@10	+10 points vs. distant supervision; +22 with GPT-4 re-rank

Danish job postings (competences)	Single-model NER + multiclass classifier	F-scores; latency	Beats SOTA; >50% faster inference
Synthetic job posts (SKILLSKAPE) + real evals	ICL with LLMs; supervised baselines	Downstream matching accuracy	Outperforms supervised baselines on real-world tests
Résumés (PDF, Portuguese)	Layout + text parser	Structural accuracy	Effective reconstruction of original section boundaries; improves downstream IE stability
Hiring→retention (surveyed studies)	Narrative synthesis	—	Describes productivity/accuracy equity gains from AI tools in hiring and beyond
Spanish job posts & CVs	Skill extraction (LLM/ICL + KG linking)	Annotation-backed evaluation	Robust multilingual matching with explicit graph links

The comparative picture in Table 1 points to a pragmatic integration path for HireSight. Embedding-first ranking lifts human-aligned orderings; taxonomy-based skills extraction reduces spurious matches; section recovery minimizes layout noise that often derails parsers in tech résumés with custom design. Coupled with domain-adaptive encoders, recruiters receive lists that converge with expert judgment while preserving explainable hooks via normalized skills and traceable sections. Table 2 organizes documented safeguards and design choices for compliant deployment. Both tables are referenced in the analysis above and use the same numeric source order established in the paper.

Table 2. Governance, fairness, and design safeguards for AI résumé screening [1-9]

Governance concern	Recommended control	Relevance to ATS deployments
Validity, autonomy, nondiscrimination, privacy, transparency	Human oversight; contestability; data minimization; transparent criteria	Anchor policy and UI requirements; enable case review and appeal mechanisms
Fairness definitions & measurement across stakeholders	Stakeholder-specific fairness metrics; documentation (model cards); DPIAs & audits	Choose metrics per decision point; log justifications and data lineage
Throughput gains vs. risks from biased histories and drift	Bias audits; recruiter-literacy-aware explanations; periodic outcome monitoring	Add scheduled audits; tailor XAI elements to recruiter proficiency
Multilingual skill ambiguity; need for human control	KG-based linking; expert annotation for ground truth	Maintain interpretable skill graphs; plug in bilingual annotators for spot checks
Layout variance obscures provenance	Section-aware parsing that preserves spans	Keep mappable spans from PDF→fields to support audits and corrections
Zero-shot extraction bias; synthetic data risks	Re-ranking with strong LLM; evaluate with RP@K; curate synthetic data	Combine retrieval+re-rank; validate synthetic corpora before deployment

Fragmented pipelines increase latency and error	Single model for extraction+classification	Lower operational latency and error surfaces in screening microservices
Domain variability in gains	Domain-sensitive thresholds; human-alignment checks (RBO)	Tune cutoffs by family of roles; monitor rank agreement vs. expert panels
Scalability and generalization	Application- + human-grounded evaluation; avoid chatbot-only pipelines	Prefer encoder retrieval stacks; gate LLM chat for augmentation, not ranking
Dataset scarcity & labeling cost	Synthetic corpora with quality diagnostics and human review	Use synthetic only with quality metrics and spot-labeling to prevent drift

These controls match the sponsor's requirements: seamless integration into existing ATS workflows (no new suite), non-blocking résumé UI, and enrichment with signals gathered from both résumé metadata and external sources. The literature favors interpretable intermediate artifacts (skills graphs, section spans, audit logs) rather than opaque end-to-end chatbots for the screening step [2–4; 7–9].

Recruiter-facing explanation patterns deserve careful calibration. A recent experiment with HR managers shows that generic explanations can raise perceived helpfulness and trust for higher-literacy users without improving, and sometimes reducing, objective understanding; only feature-importance overlays consistently aided high-literacy participants. Dashboards in hiring benefit from minimal, fit-for-purpose explanation widgets over dense counterfactual narratives [6; 9].

Design translation for HireSight. A production configuration consistent with the evidence would:

run structure-aware parsing to preserve section spans [10];

normalize extracted skills against ESCO/O*NET with few-shot LLM extractors and KG linking [2; 4];

score candidates with a domain-adapted encoder and re-rank top-K using a lightweight LLM only where ties or ambiguous matches occur [2; 9];

expose recruiter-tuned filters over normalized skills and feature-importance overlays rather than opaque rationales [6];

log span-to-field provenance and skills edges to support audits, fairness checks, and contestation [3; 8].

Resume2Vec provides the clearest numeric deltas on ranking quality (+15.85% nDCG; +15.94% RBO). CareerBERT shows human-grounded wins and documents the impracticality of chatbot-only matching at scale (high token usage and failure rates). Zero-shot skills pipelines profit from LLM-assisted re-ranking and synthetic data, yet require human oversight and diagnostics to avoid leaking bias into downstream ranking. Structure-level parsing and KG-based matching create the audit trail needed for compliance across U.S. employers while preserving recruiter workflows, which aligns with the sponsor's integration-first brief.

5. Conclusion

The evidence indicates that an embedding-first screening path, supplied with structure-aware parsing and taxonomy-anchored skill extraction, raises ranking fidelity and compresses screening latency for technology roles. Governance syntheses supply practical safeguards—audits tied to decision points, calibrated score bands with re-review, skill-graph and span-level provenance—that make automated screening contestable and explainable in enterprise settings. A recruiter-centric overlay that ingests résumé metadata and external signals, computes composite features off-path, and surfaces minimal guidance without obscuring the document aligns with operational constraints and with human-grounded evaluation trends. Calibrated against prevailing U.S. benchmarks on time-to-fill, cost-per-hire, and manual screening effort, the proposed configuration targets measurable productivity gains while sustaining transparency and due process for applicants.

References

1. 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
2. Clavié, B., & Soulié, G. (2023). Large language models as batteries-included zero-shot ESCO skills matchers. arXiv preprint arXiv:2307.03539. https://arxiv.org/abs/2307.03539
3. Hamit, K., Serra-Vidal, M., & Wanner, L. (2025). Multilingual skill extraction for job vacancy–job seeker matching in knowledge graphs. In Proceedings of the Workshop on Generative AI and Knowledge Graphs (GenAIK) (pp. 146–155). International Committee on Computational Linguistics.
4. Hunkenschroer, A. L., & Kriebitz, A. (2023). Is AI recruiting (un)ethical? A human rights perspective on the use of AI for hiring. AI and Ethics, 3, 199–213. https://doi.org/10.1007/s43681-022-00166-4
5. Li, Q., & Lioma, C. (2024). Joint extraction and classification of Danish competences for job matching. arXiv preprint arXiv:2410.22103. https://doi.org/10.48550/arXiv.2410.22103
6. Madanchian, M. (2024). From recruitment to retention: AI tools for human resource decision-making. Applied Sciences, 14(24), 11750. https://doi.org/10.3390/app142411750
7. Magron, A., Dai, A., Zhang, M., Montariol, S., & Bosselut, A. (2024). JobSkape: A framework for generating synthetic job postings to enhance skill matching. In Proceedings of the First Workshop on Natural Language Processing for Human Resources (NLP4HR 2024) (pp. 43–58). Association for Computational Linguistics.
8. Rigotti, C., & Fosch-Villaronga, E. (2024). Fairness, AI & recruitment. Computer Law & Security Review, 53, 105966. https://doi.org/10.1016/j.clsr.2024.105966
9. 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.2024.127043
10. Werner, M., & Laber, E. (2024). Extracting section structure from resumes in Brazilian Portuguese. Expert Systems with Applications, 242, 122495. https://doi.org/10.1016/j.eswa.2023.122495