RAIS Journal for Social Sciences | Vol. 10, No. 1, 2026
ISSN 2574-0245 (Print) | ISSN 2574-1179 (Online) | DOI:10.5281/zenodo.19661187

False Negatives at Scale: Governing AI-Enabled Applicant Screening Under Digital Transformation

Melissa Drew

Marymount University, Arlington, Virginia, USA
mdrewinbox@gmail.com

Abstract: Digital transformation has reshaped talent acquisition by embedding artificial intelligence (AI) within applicant tracking systems (ATS) to enable high-volume screening. However, these systems may systematically exclude qualified applicants before human evaluation occurs. This paper examines false-negative outcomes in AI-enabled screening as a socio-technical governance failure rather than a purely technical limitation. The objective is to analyze how automated screening produces exclusion errors and to identify governance controls that improve decision quality. Using a qualitative, case-based analytical approach informed by interdisciplinary literature, the study treats automated screening as a multi-stage decision process comprising résumé parsing, matching, and threshold-based filtering. The analysis demonstrates how errors in candidate representation, interpretation, and threshold calibration collectively amplify small inaccuracies into large-scale exclusion outcomes. The findings show that false-negative exclusions are primarily driven by insufficient validation, calibration, and oversight, rather than isolated technical defects. This paper reframes ATS failures as governance breakdowns and outlines a structured approach for aligning system design, organizational processes, and oversight to improve hiring outcomes and accountability.

Keywords: Applicant Tracking Systems (ATS), AI-Enabled Screening, Applicant Screening, Applicant Governance, Hiring Algorithms

Introduction

Digital transformation has fundamentally reshaped talent acquisition by shifting hiring from human-centered evaluation to platform-mediated decision pipelines in which automated systems increasingly determine which applicants receive human attention. In large organizations, applicant tracking systems (ATS) and AI-enabled screening tools now function not only as administrative repositories but as gatekeeping mechanisms that filter, rank, and prioritize candidates before recruiter review. This shift reflects the widespread adoption of automated screening across enterprise hiring environments, where organizations rely on these systems to manage scale and standardize early-stage decision-making. As a result, hiring outcomes are increasingly influenced by how these systems represent, process, and evaluate candidate information, making automated screening a structurally significant component of organizational decision-making.

This reliance has intensified under growing operational pressures to reduce time-to-fill, manage high application volumes, and sustain hiring throughput despite constrained recruiter capacity and persistent shortages in specialized roles. In response, organizations have positioned automated screening as a scalable mechanism for improving efficiency and consistency in early-stage hiring decisions. However, when these systems operate on brittle parsing, rigid keyword matching, and threshold-based filtering, small representation errors can be amplified into large-scale exclusion outcomes across applicant pools. The mechanism is that distorted candidate representations reduce match scores, and threshold gating converts those reductions into categorical rejection decisions before human evaluation occurs. Consequently, automated screening systems can simultaneously increase efficiency and suppress the flow of qualified candidates, creating a structural risk that is often misinterpreted as a labor market constraint rather than a system design and governance issue.

The increasing reliance on AI-enabled screening has elevated hiring automation from a technical implementation issue to a governance and accountability concern. Regulatory scrutiny and stakeholder expectations are expanding as organizations use algorithmic systems to make consequential employment decisions that affect access to opportunity. This matters because automated screening systems embed assumptions about qualification, relevance, and evaluation criteria that are often not transparent to users or affected individuals. The mechanism is that opaque system design, combined with high-volume decision-making, amplifies the impact of small design errors into large-scale exclusion outcomes, raising concerns about fairness, explainability, and organizational accountability. Therefore, automated screening must be understood as a governed decision system rather than a neutral efficiency tool.

The scale of this transformation is illustrated by the widespread adoption of ATS technologies, with approximately 98.4% of U.S. Fortune 500 firms using automated screening tools to manage résumé evaluation and candidate progression (Jobscan, 2025). While these systems are designed to improve efficiency and consistency, rigid keyword-matching logic and résumé-parsing limitations can systematically exclude qualified candidates before human review occurs (Bharadwaj et al., 2013; Vial, 2019; Köchling & Wehner, 2020). As a result, extraction errors and literal matching reduce candidate scores, and threshold-based filtering converts those reductions into categorical exclusion decisions at scale. Consequently, what appears to be a productivity tool can function as a high-volume exclusion mechanism when not properly governed. This dynamic establishes the conditions under which automated screening failures should be examined as systemic rather than incidental.

Digital transformation research emphasizes governance alignment (Bharadwaj et al., 2013; Vial, 2019), while hiring-focused studies highlight risks in automated screening outcomes (Köchling & Wehner, 2020). In hiring contexts, automated screening systems introduce new efficiencies but also pose risks to fairness, transparency, and validity when governance structures are underdeveloped (Köchling & Wehner, 2020; Lacroux & Martin, 2022). When organizations implement automation without redesigning oversight and accountability processes, system outputs are treated as authoritative despite limited validation. The implication is that screening failures should be interpreted as governance breakdowns within digitally transformed operating models rather than as isolated technical errors.

Consider a representative case in which a global biotechnology company underwent a rapid digital transformation and implemented an AI-driven applicant-screening tool in 2023 to reduce recruiting labor costs and increase hiring efficiency. The system was configured to compare résumés against job-description-derived keywords and assign a confidence score, with only candidates exceeding a selected threshold advancing to human review. Although leadership expected faster and more consistent screening, department leaders soon observed a sharp decline in the number of candidates reaching interviews, including for roles with persistent staffing shortages and clearly defined technical requirements. Subsequent process investigation indicated that the screening pipeline was highly sensitive to lexical and formatting variation (e.g., hyphenation differences such as "cross-functional" vs. "cross functional," and résumé formatting artifacts such as bold text), and that keyword decisions were often generated by the application rather than curated by HR subject matter experts. These conditions suggest a brittle screening process in which text parsing, keyword matching, and thresholding may be producing false-negative rejections of qualified candidates. Accordingly, this paper examines the case not as an isolated technology malfunction, but as an example of a broader organizational challenge in digital hiring transformation: how to govern and redesign automated screening systems so that efficiency goals do not undermine selection quality, stakeholder trust, and organizational performance (Vial, 2019; Warner & Wäger, 2019).

While alternative explanations may account for reduced interview flow, they do not fully explain the observed pattern of candidate exclusion. One plausible explanation is that the organization is experiencing a shortage of qualified applicants, particularly in specialized roles, which could naturally reduce the number of candidates advancing to interviews. Another possibility is that job requirements are overly restrictive or misaligned with available talent, leading to lower match rates across applicants. However, these explanations are insufficient because the observed failures occur upstream in the screening pipeline, where candidates are excluded based on parsing sensitivity, lexical variation, and threshold-based filtering before their qualifications can be meaningfully evaluated. This occurs because representation errors and rigid matching logic suppress candidate scores regardless of actual capability, and threshold gating converts those distortions into categorical rejection decisions. Therefore, while labor market conditions and job design may contribute to hiring challenges, the evidence suggests that the primary driver of false-negative outcomes in this case is a governance failure in the configuration, validation, and monitoring of the screening system within the hiring process.

This paper argues that the prevalence of false-negative rejections in AI-enabled applicant screening is not primarily a technical problem, but a socio-technical governance failure, in which digitally transformed hiring pipelines operate without sufficient validation, calibration, and oversight, resulting in misalignment between efficiency objectives and selection quality.

Problem Statement

The central problem is that the organization's automated applicant screening process functions as a socio-technical governance failure rather than a reliable decision-support tool, thereby excluding qualified candidates before human review. The system compares résumé content with AI-generated keywords, assigns a confidence score, and applies a fixed threshold to determine progression, yet errors in résumé parsing and information extraction can distort the system's representation of candidate qualifications (Yu et al., 2005). The mechanism is that incomplete or inaccurate representations reduce match scores, and threshold-based filtering converts those reductions into categorical rejection decisions at scale. Therefore, the issue is not primarily a limitation of matching algorithms but a failure to validate, calibrate, and oversee their operation within the hiring pipeline.

This governance failure creates a misalignment between the organization's efficiency objectives and its ability to maintain selection quality, resulting in measurable business consequences, including reduced interview pipelines, delayed staffing, and increased recruiter intervention. Although the system was implemented to reduce workload and accelerate hiring, threshold-based decision systems can systematically increase false-negative outcomes when upstream signals are imperfect (Hardt et al., 2016; Pleiss et al., 2017). The distorted scores are treated as valid signals, and high-volume filtering suppresses qualified candidate flow, particularly in roles with existing shortages. Consequently, the absence of governance controls, including validation processes, role-specific calibration, and ongoing monitoring, prevents the organization from detecting and correcting these errors, turning an efficiency tool into a constraint on organizational performance.

Purpose Statement

The purpose of this paper is to examine the organization's automated applicant screening process as a socio-technical pipeline and to analyze how insufficient validation, calibration, and oversight produce false-negative candidate exclusions. The study further aims to identify evidence-based interventions to restore alignment between efficiency objectives and selection quality by improving screening design and governance.

Significance Statement

This issue is significant because it illustrates how AI-enabled hiring systems can introduce organization-wide risk when efficiency-driven automation is implemented without sufficient governance over decision processes and outcomes. Digital transformation research consistently shows that technology adoption alone does not produce performance gains; organizations must align system design with governance structures, monitoring mechanisms, and decision accountability to ensure intended outcomes are achieved (Sebastian et al., 2017; Warner & Wäger, 2019). In the context of applicant screening, automated systems increasingly function as gatekeeping mechanisms that determine access to human evaluation, making their design and oversight consequential for both operational performance and organizational legitimacy.

The significance of this problem is amplified by the scale at which automated screening operates. When parsing limitations, lexical matching constraints, and threshold-based filtering are embedded in high-volume hiring pipelines, small representation errors can translate into large-scale false-negative exclusions. Selection research demonstrates that the quality of hiring decisions directly affects organizational capability and performance outcomes, making selection validity a critical concern in any screening process (Schmidt & Hunter, 1998). However, when efficiency objectives dominate system configuration without appropriate validation and calibration, organizations risk suppressing the flow of qualified candidates, particularly in roles with existing labor shortages, thereby creating a misalignment between operational efficiency and hiring effectiveness.

Additionally, research on procedural justice and algorithmic decision-making shows that opaque or inconsistent screening outcomes can undermine stakeholder trust and organizational legitimacy (Gilliland, 1993; Cropanzano et al., 2003; Köchling & Wehner, 2020). As organizations increasingly rely on AI-enabled systems to make early-stage hiring decisions, the inability to explain or justify candidate exclusion decisions introduces reputational and governance risk. Therefore, this issue is significant not only as a technical or operational challenge but as a broader governance problem within digitally transformed hiring systems, where the absence of validation, calibration, and oversight mechanisms can convert efficiency tools into constraints on organizational performance and trust.

Literature review

False-negative exclusions in AI-enabled applicant screening arise from interacting failures across the screening pipeline, rather than from isolated technical defects, and require an integrated analysis that connects system design, organizational processes, and governance structures. Prior research on digital transformation and algorithmic decision-making demonstrates that performance outcomes in technology-enabled systems are shaped not only by technical configurations but by how those systems are embedded within organizational workflows, accountability structures, and user practices (Bharadwaj et al., 2013; Sebastian et al., 2017; Vial, 2019; Köchling & Wehner, 2020). This occurs because design choices in parsing, matching, and thresholding interact with organizational factors such as decision rights, monitoring routines, and stakeholder behavior, causing small technical limitations to scale into systematic exclusion outcomes when governance controls are insufficient. Therefore, the literature relevant to this case must be organized around the screening pipeline and its governance, rather than treated as separate technical or organizational domains, to explain how automated hiring systems can simultaneously improve efficiency while degrading selection quality.

Automated Applicant Screening as a Pipeline

Automated applicant screening should be understood as a multi-stage decision pipeline in which candidate information is sequentially transformed, evaluated, and filtered, rather than as a single matching function, because errors introduced at early stages propagate and intensify through downstream decisions. In automated screening systems, information extraction and algorithmic decision-making show that résumé parsing converts unstructured text into structured representations, which are then evaluated through keyword matching or semantic scoring models before being subjected to threshold-based filtering that determines candidate progression (Yu et al., 2005; Retyk et al., 2023; Jiang & Zhai, 2007). Therefore, each stage in the pipeline depends on the outputs of the previous stage, so inaccuracies in representation reduce the fidelity of matching, and reduced match scores are subsequently converted into categorical inclusion or exclusion decisions through thresholding logic. As a result, small deviations in how candidate qualifications are captured or interpreted can produce disproportionately large exclusion effects when they accumulate across stages, particularly in high-volume hiring environments where automated systems operate without continuous validation.

This pipeline structure is critical because it reveals that false-negative outcomes are not random errors but predictable consequences of interacting design choices that remain unmonitored within the system. Research on algorithmic pipelines demonstrates that when sequential decision stages lack validation and feedback mechanisms, upstream distortions are not corrected but instead become embedded in downstream outputs, reinforcing systematic bias or exclusion patterns over time (Corbett-Davies et al., 2023; Pleiss et al., 2017). The mechanism is that the absence of validation at each stage prevents detection of representation and scoring errors, while the absence of monitoring across stages obscures how those errors accumulate and affect overall system performance. Therefore, understanding automated screening as a pipeline shifts the focus from individual technical components to the governance of the entire decision process, where validation, calibration, and oversight must be applied at each stage to ensure that efficiency gains do not come at the expense of selection quality.

Résumé Parsing as a Root Cause of Screening Error

Résumé parsing serves as a foundational stage in automated applicant screening, and errors introduced at this stage are root causes of downstream evaluation failures because they distort how candidate qualifications are represented before any matching or scoring occurs. Parsing systems must identify and structure entities such as skills, roles, and experience from highly variable résumé formats, which often include inconsistent layouts, abbreviations, and non-standard phrasing (Yu et al., 2005; Retyk et al., 2023). When parsing fails to extract or normalize this information correctly, candidate attributes are either omitted or inaccurately represented, reducing the completeness and fidelity of the structured profile used as input for subsequent evaluation stages. As a result, candidates may appear less qualified within the system than they are in reality, creating latent distortions that are not visible to users but directly influence downstream scoring and filtering decisions.

These representation errors become consequential because they systematically reduce candidate match scores regardless of actual capability, increasing the likelihood of exclusion when threshold-based filtering is applied. Research on text processing and information retrieval demonstrates that incomplete or fragmented representations weaken the signals used in matching algorithms, leading to lower similarity scores between candidate profiles and job requirements (Jiang & Zhai, 2007). As a result, missing or misclassified skills reduce the overlap between candidate and job representations, thereby lowering the computed relevance score used for ranking or filtering. Consequently, candidates affected by parsing errors are disproportionately likely to fall below decision thresholds, converting representation inaccuracies into categorical rejection outcomes. This process illustrates how parsing errors function not as isolated technical issues, but as structural inputs that shape the entire evaluation pipeline.

From a governance perspective, parsing errors persist because organizations often lack validation mechanisms to assess the accuracy of résumé data extraction and representation within the system. Digital transformation research emphasizes that system performance depends on the alignment between technical design and organizational oversight, yet parsing outputs are rarely audited against original résumé content once systems are deployed (Sebastian et al., 2017; Vial, 2019). The emerging pattern is that without validation, representation errors remain undetected, and without feedback loops, these errors are neither corrected nor accounted for in downstream decision-making. Therefore, résumé parsing should be understood as a governed input process that requires systematic validation and monitoring to ensure that candidate information is accurately captured before it is used to drive automated hiring decisions.

Matching and Interpretation in Automated Screening: Lexical and Semantic Approaches

Matching and interpretation processes in automated applicant screening determine how candidate qualifications are evaluated against job requirements, and both lexical and semantic approaches introduce distinct sources of error that affect scoring outcomes. Lexical matching relies on tokenization and normalization to compare candidate and job representations based on exact or near-exact term overlap, making evaluation highly sensitive to surface-level variation such as formatting, spacing, and phrasing differences (Jiang & Zhai, 2007). The mechanism is that variations such as “cross-functional” versus “cross functional” can yield different token representations, which in turn affect whether candidate skills are recognized and matched, leading to inconsistent scoring despite equivalent qualifications. As a result, lexical approaches introduce brittleness into the evaluation process, where minor linguistic differences can disproportionately influence candidate rankings.

Semantic matching approaches, including embedding-based transformer models, address this brittleness by capturing contextual meaning rather than relying on exact term overlap, but they introduce new challenges related to interpretability and transparency. Research on neural language models shows that embeddings represent candidate and job information as high-dimensional vectors, enabling systems to identify relationships between semantically similar concepts even when expressed differently (Vaswani et al., 2017; Vanetik, 2023). As a result, semantic models infer similarity from learned contextual patterns, improving recall by identifying relevant candidates that would otherwise be excluded by strict keyword matching. However, because these similarity scores are derived from complex representations, users cannot easily trace how specific attributes influence outcomes, reducing transparency and making it difficult to assess the validity of system decisions (Kozak & Fel, 2024; Leichtmann et al., 2023). Consequently, while semantic approaches reduce sensitivity to linguistic variation, they shift evaluation risk from under-recognition of qualifications to limited explainability of scoring outcomes.

These matching approaches become particularly consequential when combined with upstream representation errors and downstream threshold-based filtering within the screening pipeline. This occurs when parsing inaccuracies reduce the completeness of candidate representations, lexical or semantic matching interprets those incomplete inputs, and thresholding converts resulting score differences into categorical inclusion or exclusion decision Consequently, both lexical brittleness and semantic opacity contribute to systematic distortions in candidate evaluation, albeit through different pathways, reinforcing the need to analyze matching as part of an interconnected pipeline rather than as an isolated function. From a governance perspective, neither lexical nor semantic matching resolves the underlying challenges of automated screening; instead, each requires calibration, validation, and monitoring to ensure reliable and consistent outcomes. Digital transformation and AI governance research emphasize that system performance depends on aligning technical configuration with oversight mechanisms, particularly as model complexity increases (Bharadwaj et al., 2013; Vial, 2019; NIST, 2023). The mechanism is that without calibration, linguistic variation or model inference errors persist, and without monitoring, organizations cannot detect how these errors influence candidate selection outcomes over time. Therefore, matching and interpretation should be governed as dynamic processes within the screening pipeline, requiring continuous validation, transparency controls, and performance monitoring to balance flexibility, accuracy, and accountability in automated hiring decisions.

Thresholding and the False-Negative Problem: Why Calibration is Not Neutral

Thresholding functions as the decisive stage in automated applicant screening, where continuous match scores are converted into categorical inclusion or exclusion decisions, making calibration choices inherently consequential rather than neutral. Threshold-based systems determine which candidate progresses by applying predefined cutoffs to similarity or relevance scores, thereby operationalizing tradeoffs between different types of error (Hardt et al., 2016; Pleiss et al., 2017). Even small reductions in candidate scores, caused by upstream representation errors or matching limitations, can result in categorical exclusion when thresholds are rigidly applied, particularly in high-volume hiring environments where automated filtering replaces human review. As a result, thresholding does not simply reflect candidate quality but actively shapes selection outcomes by determining how scoring uncertainty is resolved.

These threshold decisions are not neutral because they encode organizational priorities regarding efficiency, risk tolerance, and resource allocation, often privileging speed and volume over completeness of candidate evaluation. Research on algorithmic fairness demonstrates that different threshold settings systematically redistribute error, where stricter cutoffs increase false-negative rates by excluding candidates whose scores fall just below the decision boundary (Corbett-Davies et al., 2023). Therefore, threshold calibration determines how borderline cases are treated, and, when combined with imperfect upstream inputs, it amplifies exclusion effects by disproportionately filtering out candidates affected by representation or matching distortions. Consequently, false-negative outcomes emerge not as random errors but as predictable consequences of how organizations configure decision thresholds within the screening pipeline.

From a governance perspective, the false-negative problem persists because threshold calibration is often treated as a one-time technical configuration rather than an ongoing decision process requiring validation and oversight. Research on AI adoption and organizational decision-making emphasizes that complex systems require continuous alignment between technical settings and organizational objectives, particularly when decision outcomes carry operational and reputational consequences. Without monitoring pass-through rates, analyzing near-threshold candidates, and recalibrating decision criteria over time, organizations cannot detect or correct how threshold settings shape exclusion patterns. Therefore, thresholding should be treated as a dynamic control point within the screening pipeline, requiring continuous calibration, performance monitoring, and alignment with objectives to ensure that efficiency gains do not systematically suppress the flow of qualified candidates.

AI-generated job description keywords/job analysis validity

AI-generated job description keywords function as the reference standard against which candidate qualifications are evaluated, and inaccuracies at this stage introduce validity errors that distort the entire screening pipeline. Research on job analysis emphasizes that selection systems depend on the accurate specification of role-relevant knowledge, skills, and abilities (KSAs), yet AI-assisted tools increasingly generate or refine job requirements using historical data, templates, or inferred patterns that may not fully reflect current role needs (Köchling & Wehner, 2020). This occurs when job descriptions omit critical competencies, overemphasize easily measurable attributes, or encode legacy biases; the resulting keyword sets define an incomplete or skewed representation of job requirements. As a result, candidate evaluation becomes anchored to a flawed reference point, meaning that even accurately represented candidate profiles may be misaligned with system-defined criteria.

These validity issues are consequential because matching algorithms optimize on the provided job representation rather than the role's true requirements. Research on algorithmic decision systems shows that models learn and operationalize the objective functions and input specifications they are given, regardless of whether those inputs accurately reflect real-world conditions (Adensamer et al., 2021). Lexical or semantic matching processes maximize similarity between candidate profiles and the defined keyword set, so any distortion in job requirements directly shapes scoring outcomes by rewarding alignment with incomplete or biased criteria. Consequently, false-negative exclusions can occur not because candidates lack relevant qualifications, but because those qualifications are not captured or prioritized within the system's definition of the job, extending the source of error beyond candidate representation to the construction of the evaluation standard itself.

From a governance perspective, job description generation and validation are critical control points that are often overlooked in automated screening systems, despite their central role in shaping downstream decisions. Digital transformation and AI governance research emphasize that system outputs are only as valid as the inputs and assumptions embedded in their design, yet organizations rarely implement formal validation processes to ensure that AI-generated or standardized job requirements accurately reflect evolving role demands (Vial, 2019; NIST, 2023). Without periodic review, stakeholder input, and alignment with actual job performance criteria, errors in job analysis persist and propagate through matching and thresholding stages, reinforcing misalignment between hiring decisions and organizational needs. Therefore, job description generation should be treated as a foundational input process that requires validation, continuous updating, and alignment with real-world role requirements to ensure that automated screening systems evaluate candidates against accurate and relevant criteria.

Impact on Key Constituencies

Applicants (External Impact: Access and Exclusion)

False-negative outcomes in automated screening disproportionately affect applicants by restricting access to employment opportunities despite relevant qualifications. Research on algorithmic decision systems shows that when upstream representation errors and matching distortions interact with threshold-based filtering, qualified candidates can be systematically excluded before human evaluation occurs (Köchling & Wehner, 2020; Corbett-Davies et al., 2023). The mechanism is that incomplete or misinterpreted candidate information reduces match scores, and threshold cutoffs convert those reductions into categorical rejection decisions, eliminating candidates from consideration without visibility into the decision process. As a result, applicants experience exclusion that is not readily explainable or contestable, raising concerns about fairness and transparency while limiting access to opportunities and exposing organizations to legitimacy and compliance risks in technology-mediated hiring.

Recruiters and Hiring Managers (Decision Quality and Productivity)

Automated screening systems reshape recruiter and hiring manager decision-making by altering both the quality of candidate pools and the nature of evaluation work. Prior research indicates that algorithmic tools are often used to manage high application volumes and improve efficiency, but their outputs depend on upstream data quality and model configuration (Kokina & Davenport, 2017). Therefore, false-negative filtering reduces the number and diversity of qualified candidates reaching human review, while opaque scoring processes limit recruiters' ability to understand or challenge system recommendations. Consequently, recruiters may experience reduced productivity not because of workload volume, but because of diminished candidate quality and increased effort required to verify or override system outputs, reducing decision confidence and weakening the effectiveness of system-supported hiring processes.

Organization (Performance, Risk, and Culture)

At the organizational level, false-negative screening outcomes translate into measurable performance and governance risks by degrading hiring effectiveness and introducing misalignment between efficiency objectives and workforce quality. When automated decision systems are not properly governed, operational efficiencies can mask underlying performance degradation (Bharadwaj et al., 2013; Vial, 2019). Repeated exclusion of qualified candidates reduces interview-slate quality, delays staffing, and increases the likelihood of suboptimal hiring decisions, while the absence of monitoring obscures these effects from organizational visibility. These dynamics also shape organizational culture, as reliance on automated screening without transparency or validation can normalize passive acceptance of system outputs, reducing critical evaluation and reinforcing over-reliance on technology. As a result, organizations may experience interview-slate collapse, delayed staffing, and diminished workforce quality despite apparent efficiency gains, while reinforcing cultural patterns that weaken accountability and decision quality over time.

Leadership and Governance (Accountability and Strategic Risk)

Leadership bears responsibility for ensuring that automated screening systems operate as governed decision processes rather than as opaque technical tools, as these systems introduce strategic and accountability risks that extend beyond operational hiring outcomes. Research on AI adoption and organizational governance emphasizes that complex decision-making systems require clearly defined decision rights, monitoring mechanisms, and alignment with organizational objectives to ensure responsible use (Vial, 2019). Without governance structures, leaders lack visibility into how system configurations affect outcomes, limiting their ability to detect patterns of systemic exclusion or intervene when performance misalignments occur. Consequently, accountability for hiring outcomes remains with the organization despite limited control over system behavior, creating strategic risk related to workforce quality, regulatory scrutiny, and stakeholder trust. Therefore, organizations face sustained strategic exposure when leadership cannot fully observe, explain, or govern automated decision processes that directly influence hiring outcomes.

Strategy/Planning Model: NIST AI RMF as the Governance Backbone

The National Institute of Standards and Technology Artificial Intelligence Risk Management Framework (AI RMF) provides a structured governance model that directly addresses the systemic failures observed in automated applicant screening by treating AI-enabled decision systems as lifecycle processes requiring continuous oversight. Unlike static technical fixes, the AI RMF emphasizes four core functions, map, measure, manage, and govern, that collectively guide how organizations identify risks, evaluate system performance, and implement controls across the full AI lifecycle (NIST, 2023). This leads to the explicit linking of system design, deployment, and monitoring to organizational objectives, enabling organizations to detect how errors in parsing, matching, thresholding, and job definition propagate through the screening pipeline and influence decision outcomes. As a result, the AI RMF reframes automated screening from a configuration problem to a governance problem, in which performance must be continuously assessed and aligned with hiring objectives.

This governance approach is particularly relevant because it aligns with the multi-stage nature of automated screening systems, where risks emerge not from individual components but from their interaction across the pipeline. Research on digital transformation emphasizes that system effectiveness depends on the integration of technology, processes, and organizational controls rather than on isolated technical improvements (Bharadwaj et al., 2013; Vial, 2019). The AI RMF mechanism is that governance interventions enable organizations to map each stage of the screening pipeline to specific risks, measure performance through indicators such as pass-through rates and decision consistency, and manage those risks through validation, calibration, and oversight practices. Consequently, organizations can move from reactive problem-solving to proactive governance by identifying where and how exclusion patterns emerge within the system, rather than addressing symptoms after they occur.

From an implementation perspective, the AI RMF serves as a planning model that supports the phased recommendations outlined in this paper by providing a structured approach to sequencing governance interventions across the screening pipeline. Immediate actions, such as validation of parsing outputs and human review of near-threshold candidates, align with the "measure" and "manage" functions, while longer-term actions, including threshold calibration, monitoring systems, and decision-rights frameworks, align with "map" and "govern." This alignment ensures that governance is not treated as an abstract principle but as an operational process embedded within system design and use. Therefore, the AI RMF serves as the governance backbone, enabling organizations to systematically align automated screening systems with organizational objectives, reducing false-negative outcomes while improving transparency, accountability, and decision quality in technology-mediated hiring.

Change Management Model: ADKAR as the Implementation Pathway

The Prosci ADKAR model provides an implementation pathway to operationalize governance controls in automated screening systems by addressing behavioral and organizational barriers to adoption. While governance frameworks such as the NIST AI RMF define what controls are required, research on change management shows that successful implementation depends on individual and organizational alignment with new processes, roles, and accountability structures (Hiatt, 2006). This occurs because governance interventions, such as validation protocols, threshold calibration, and monitoring practices, require changes in how recruiters, hiring managers, and leaders interpret and act on system outputs. As a result, without structured change management, governance controls may be inconsistently applied or bypassed in practice, limiting their effectiveness in addressing false-negative outcomes.

The ADKAR model is particularly relevant because it aligns each stage of organizational change with the specific governance challenges identified in the screening pipeline. Change management research demonstrates that successful adoption of new processes depends on progressing individuals through stages of awareness, capability development, and sustained reinforcement, rather than relying solely on structural or technical changes (Hiatt, 2006; Armenakis & Harris, 2009). Therefore, Awareness and Desire shift organizational understanding from viewing automated screening as a neutral efficiency tool to recognizing it as a governed decision system with measurable risks, while Knowledge and Ability enable users to interpret system outputs, apply validation and calibration practices, and exercise judgment in near-threshold or ambiguous cases. Reinforcement ensures that governance behaviors, such as monitoring system performance and questioning outputs, are sustained over time rather than treated as one-time interventions. Consequently, each stage of ADKAR addresses a specific failure point in adoption, ensuring that governance practices are not only designed but consistently executed across the organization.

From a strategic perspective, integrating ADKAR with governance frameworks ensures that automated screening systems evolve from technical implementations into organizational capabilities that support reliable and accountable decision-making. Digital transformation research emphasizes that system effectiveness depends not only on technical design but on how users engage with and trust system outputs within organizational contexts (Vial, 2019; Sebastian et al., 2017). In parallel, change management research demonstrates that sustained adoption of new practices requires reinforcement mechanisms that embed new behaviors into routine decision-making (Hiatt, 2006; Armenakis & Harris, 2009). The mechanism is that when change management aligns user behavior with governance objectives, organizations can create feedback loops that improve system performance, detect emerging risks, and adjust decision processes over time. Therefore, ADKAR functions as the implementation pathway that enables governance frameworks to translate into sustained behavioral change, ensuring that automated screening systems improve hiring outcomes rather than perpetuate exclusion risks through inconsistent or unexamined use.

Digital Transformation Theory

Digital transformation theory explains failures in automated applicant screening as governance breakdowns in technology-enabled decision systems rather than as isolated technical deficiencies. Prior research demonstrates that digital transformation outcomes depend on the alignment of technology, processes, and governance structures, not simply the deployment of digital tools (Bharadwaj et al., 2013; Vial, 2019). When organizations implement AI-enabled screening without clearly defined decision rights, monitoring processes, and accountability structures, technical components such as parsing, matching, and thresholding operate independently of organizational objectives, allowing small errors to propagate unchecked across the system. Therefore, false-negative exclusions reflect a failure to align system behavior with organizational intent, positioning breakdowns in automated screening as governance failures in digital transformation rather than as limitations of the underlying technology.

Automated screening pipelines further illustrate how digital transformation failures arise from misalignment among interconnected system components rather than from isolated errors in individual stages. Research on digital operating models emphasizes that effective digital systems require coordination between technical architecture, data flows, and human decision-making processes to ensure consistent and reliable outcomes (Sebastian et al., 2017). Therefore, parsing, matching, and thresholding generate outputs that must be interpreted and validated within organizational workflows, and when these processes are not governed holistically, localized distortions accumulate into systemic patterns of exclusion. Consequently, organizations may achieve efficiency gains through automation while simultaneously degrading decision quality, demonstrating how digital transformation efforts can produce unintended negative outcomes when governance mechanisms are insufficiently integrated across the system.

Sustained value from digital transformation depends on the organization's ability to continuously govern, monitor, and adapt technology-enabled decision processes over time. Organizations that realize value from digital initiatives establish feedback loops, performance monitoring systems, and governance structures that maintain alignment between system outputs and business objectives (Vial, 2019; Weill & Ross, 2004). Continuous monitoring enables organizations to detect deviations in system performance, recalibrate decision thresholds, and integrate human judgment into automated processes, preventing errors from becoming institutionalized. Therefore, the persistent false-negative outcomes observed in automated screening systems reflect a failure to implement adaptive governance within digital transformation, reinforcing the need for continuous oversight to ensure that efficiency gains do not come at the expense of hiring effectiveness.

Critical Evaluation

While AI-enabled applicant tracking systems provide operational benefits and are influenced by contextual factors, these considerations do not fully explain the observed pattern of false-negative exclusions, which is more convincingly attributed to governance failure within the screening pipeline. Automated screening systems improve efficiency by enabling organizations to process large applicant volumes, reduce recruiter workload, and standardize early-stage decision-making, particularly in high-volume hiring environments (Köchling & Wehner, 2020). Additionally, organizations may intentionally configure screening thresholds to prioritize efficiency or reduce downstream evaluation costs, while contextual factors such as labor market conditions, organizational risk tolerance, and regulatory expectations can influence how these systems are designed and used. Although these factors shape system adoption and configuration, they do not inherently result in systematic exclusion unless combined with unvalidated parsing, misaligned matching logic, and unmonitored thresholding within the screening pipeline. Therefore, the persistence of false-negative outcomes suggests that the primary issue lies not in the existence of automation itself, but in the absence of governance mechanisms required to validate, calibrate, and monitor system performance in alignment with organizational hiring objectives.

Methods

Research Design

This study employs a qualitative case-based analytical design to examine false-negative outcomes in automated applicant screening systems as a governance failure within a digital decision pipeline. Case study methodology is appropriate for investigating complex, context-dependent phenomena in which technical, organizational, and behavioral factors interact to produce observable outcomes (Yin, 2018). The focus of this study is on understanding how automated screening systems process candidate information through sequential stages, including parsing, matching, and thresholding, and how these stages contribute to systematic exclusion patterns when governance mechanisms are insufficient.

This design enables an in-depth examination of the interaction between system processes and organizational decision-making.

Data Sources

The data for this study consist of secondary sources, including existing literature on applicant tracking systems, algorithmic decision-making, natural language processing, and digital transformation, as well as documented examples of automated screening processes. Prior research provides the empirical and theoretical foundation for analyzing how candidate information is extracted, evaluated, and filtered within automated systems (Yu et al., 2005; Jiang & Zhai, 2007; Corbett-Davies et al., 2023). These sources collectively support the examination of how system design and configuration influence hiring outcomes.

Analytical Approach

The analysis is conducted using a structured, process-oriented approach that traces how candidate information moves through the screening pipeline and how errors introduced at each stage propagate through subsequent stages. This approach focuses on identifying causal relationships among parsing accuracy, matching logic, and threshold-based decision rules, and on their combined effect on candidate exclusion. A conceptual framework integrating digital transformation theory, AI governance principles, and change management models is used to interpret system behavior and organizational response (Bharadwaj et al., 2013; Vial, 2019; NIST, 2023; Hiatt, 2006). This framework enables the analysis of automated screening systems as socio-technical processes rather than isolated technical components.

Rigor and Limitations

To ensure analytical rigor, the study applies a consistent framework across all stages of analysis and draws on multiple streams of literature to support interpretation, enabling triangulation of concepts and findings. The structured examination of the screening pipeline reduces the risk of attributing outcomes to isolated factors and supports a comprehensive understanding of system behavior. However, the study is limited by its reliance on secondary data and conceptual analysis rather than primary empirical data. Despite this limitation, the approach is appropriate for identifying systemic patterns, causal mechanisms, and governance implications in technology-mediated decision systems.

Literature Search and Selection Strategy

Search Sources

The literature search targets scholarly work on automated hiring and résumé screening systems, with an emphasis on text extraction, matching logic, threshold calibration, governance, and stakeholder impacts, as these components most directly explain the observed failure modes. These databases provide complementary coverage: Semantic Scholar supports efficient retrieval of recent and highly cited research across technical domains, Google Scholar broadens access to interdisciplinary and computer science literature, and ProQuest enables targeted retrieval of peer-reviewed management and social science research relevant to applicant reactions, procedural justice, and organizational change. Using multiple sources increases coverage of technical and organizational studies needed to interpret the case.

Keywords

Keywords were selected to align directly with the mechanism stack underlying automated applicant screening, including résumé ingestion and extraction, lexical matching and normalization, confidence scoring and thresholding, and downstream fairness and stakeholder impacts. Core keywords included the following terms and variants: applicant tracking system, ATS, automated screening, algorithmic hiring, résumé screening, resume parsing, CV parsing, information extraction, keyword matching, lexical matching, tokenization, normalization, semantic matching, embeddings, learning to rank, confidence threshold, false negatives, selection validity, procedural justice, applicant reactions, audit, governance, and AI risk management. The inclusion of both technical and organizational terms ensures that the search captures research addressing (a) system-level brittleness at the text-processing and decision stages and (b) the translation of these technical limitations into trust, legitimacy, and staffing outcomes within organizational contexts. This keyword strategy supports retrieving studies that explain both screening-accuracy mechanisms and organizational adoption risks.

Boolean Search Terms

Boolean strings are constructed to (1) retrieve empirical and review literature on automated hiring, (2) isolate research on parsing and matching mechanisms, and (3) capture governance, fairness, and stakeholder impact studies. The following Boolean strings guide searches (with minor adjustments by database syntax):

1. ("applicant tracking system" OR ATS) AND ("resume parsing" OR "CV parsing" OR "information extraction")
2. ("algorithmic hiring" OR "automated screening" OR "AI hiring") AND ("resume screening" OR "résumé screening")
3. ("keyword matching" OR "lexical matching" OR tokenization OR normalization) AND (resume OR résumé OR CV) AND (screening OR hiring OR recruitment)
4. ("semantic matching" OR embeddings OR transformer OR "learning to rank") AND (resume OR résumé OR CV) AND (job description OR vacancy)
5. ("automated screening" OR "algorithmic decision-making") AND ("procedural justice" OR fairness OR "applicant reactions")
6. ("AI governance" OR "AI risk management" OR audit OR "model documentation") AND (hiring OR recruitment OR selection)

These queries are designed to retrieve studies that can be mapped onto the literature review structure: (a) core screening mechanisms, (b) stakeholder impacts, and (c) governance and change requirements. Using multiple strings reduces the risk that the review is overly weighted toward one discipline (e.g., only computer science) and misses relevant organizational research.

Inclusion Strategy

Inclusion criteria were established to prioritize studies directly relevant to the case mechanism and to ensure the use of peer-reviewed evidence suitable for explanatory analysis and actionable recommendations. Studies were included if they met all of the following criteria:

- Peer-reviewed journal articles or peer-reviewed conference proceedings, with limited inclusion of authoritative standards and framework documents used explicitly as analytical models (e.g., NIST AI Risk Management Framework).
- Substantive relevance to at least one of the following domains: automated hiring and screening, résumé parsing and information extraction, keyword and lexical matching, semantic matching, scoring and thresholding with associated error tradeoffs, fairness and discrimination risks in hiring automation, applicant, recruiter, or manager reactions, governance and auditing practices, or organizational change in digital decision systems.
- Publication timeframe primarily between 2013 and 2026 to reflect contemporary developments in digital transformation, change management, and AI-enabled hiring,

with selective inclusion of foundational works (e.g., procedural justice and selection validity) to support theoretical grounding (Gilliland, 1993; Schmidt & Hunter, 1998). These inclusion criteria ensured that selected studies could explain the observed failure mechanism, specifically false-negative exclusions driven by parsing, matching, and thresholding processes, while also providing the organizational context necessary to interpret stakeholder impacts and implementation constraints. This approach supports a literature base that is both causally specific and organizationally actionable, linking technical system behavior to governance implications.

Exclusion Strategy

Exclusion criteria were applied to prevent the inclusion of sources that do not support rigorous, evidence-based analysis of screening mechanisms or organizational impacts. Sources were excluded if they met any of the following conditions:

- Non-peer-reviewed opinion pieces, blog posts, vendor marketing materials, or non-methodological commentary, unless used solely for contextual background and clearly identified as such.
- Studies focused on AI decision-making in domains not transferable to employment selection, where underlying mechanisms differ materially (e.g., medical diagnosis systems without analogous text extraction and matching processes).
- Publications lacking sufficient methodological detail to assess credibility or without clear relevance to applicant screening, selection processes, or stakeholder outcomes.

The application of these exclusion criteria supports the defensibility of the analysis by ensuring that conclusions are grounded in credible, relevant, and methodologically sound sources. This approach is consistent with APA expectations for evidence-based synthesis in case-based research.

Screening, Synthesis, and Article Count

A staged screening process was used to ensure that included studies were both relevant to the case and collectively sufficient to support the literature review and associated recommendations. Titles and abstracts were screened initially to remove clearly irrelevant sources, followed by deeper screening to confirm alignment with the review structure, including core mechanisms, key constituencies, and the selected analytical models (NIST AI Risk Management Framework, ADKAR, and digital transformation theory). Key concepts, findings, and methodological characteristics were extracted into structured evidence organized by literature review subsection.

This structured approach supports analytic synthesis by grouping studies according to the mechanisms or organizational consequences they explain, rather than summarizing sources in isolation. For example, studies were categorized based on their contributions to understanding parsing errors as upstream drivers, lexical brittleness as a matching-stage limitation, thresholding as an error-amplification mechanism, and justice perceptions as stakeholder-level impacts. This process enables the integration of technical and organizational insights within a unified analytical framework. The final review includes 42 articles.

Recommendation

The findings of this analysis indicate that the organization’s automated screening challenges cannot be resolved through isolated technical adjustments but require a structured, phased approach that addresses the screening pipeline as a governed socio-technical system. While individual issues such as parsing errors, threshold miscalibration, and limited oversight contribute to false-negative outcomes, their combined effect reflects a broader misalignment between efficiency objectives and selection quality. Representation errors, scoring distortions, and rigid filtering logic interact within the pipeline to amplify exclusion outcomes at scale, particularly when governance controls are absent. Therefore, the following recommendations are organized into phased actions, beginning with immediate stabilization of candidate flow, followed by short-term calibration of decision logic, and culminating in long-term governance redesign and sustained organizational adoption, to ensure that corrective actions are both effective and strategically sequenced.

Phase 1: Immediate Actions (Stabilize the Screening Pipeline)

Organizations should immediately implement targeted validation checks and human oversight controls to prevent ongoing false-negative exclusions while longer-term fixes are developed. The current screening pipeline relies on automated parsing, keyword matching, and threshold filtering without sufficient verification of how candidate information is represented or evaluated, creating a high risk of excluding qualified applicants at scale (Yu et al., 2005). The mechanism is that representation errors reduce candidate scores, and threshold-based filtering converts those distortions into immediate rejection decisions before human review occurs. Therefore, organizations should implement rapid interventions, including manual review of near-threshold candidates, spot-checking parsed résumé outputs against original documents, and temporarily lowering or adjusting screening thresholds for critical roles, to stabilize candidate flow and prevent further degradation of hiring outcomes.

Phase 2: Short-Term Actions (30–60 Days – Calibrate and Control Decision Logic)

Within 30 to 60 days, organizations should systematically recalibrate matching algorithms and threshold criteria to align screening decisions with actual hiring outcomes and reduce error amplification in the pipeline. Threshold-based decision systems inherently allocate error tradeoffs, where rigid cutoffs can significantly increase false-negative rates when upstream signals are imperfect (Hardt et al., 2016; Pleiss et al., 2017). Small variations in candidate representation or keyword matching produce disproportionate exclusion effects when thresholds are not tuned to real-world data. Consequently, organizations should analyze historical hiring data to test alternative threshold levels, implement role-specific scoring criteria, and monitor pass-through rates to ensure that screening decisions reflect actual candidate quality rather than artifacts of system configuration.

Phase 3: Long-Term Actions (Governance and Operating Model Redesign)

Over the longer term, organizations must redesign the hiring operating model to treat automated screening as a governed socio-technical system with continuous oversight, clear decision rights, and embedded accountability. The National Institute of Standards and Technology Artificial Intelligence Risk Management Framework emphasizes that AI systems require lifecycle governance through ongoing measurement, monitoring, and risk management rather than static implementation (NIST, 2023). Without governance structures, errors in parsing, matching, and thresholding persist undetected and scale across hiring decisions, reinforcing misalignment between efficiency objectives and selection quality. Therefore, organizations should establish governance frameworks that include performance-monitoring dashboards, audit trails for screening decisions, defined ownership of system performance, and structured override policies to ensure human judgment remains integrated at critical decision points.

Phase 4: Sustained Adoption (Embed Behavioral and Organizational Change)

To ensure long-term effectiveness, organizations must embed governance practices into day-to-day hiring behavior through structured change management and capability development. Even well-designed systems with calibrated thresholds and governance controls can fail if users lack the knowledge, incentives, or authority to question and refine system outputs. The mechanism is that uncritical reliance on automated recommendations reinforces existing errors and limits feedback loops necessary for continuous improvement. As a result, organizations should implement training programs, establish clear usage guidelines, and align performance incentives with responsible use of the system to ensure that recruiters and hiring managers actively engage with, rather than passively accept, automated screening decisions.

To operationalize these recommendations, Table 1 outlines a phased implementation roadmap with associated validation metrics that enable the organization to monitor progress and ensure alignment between system performance and hiring objectives.

Table 1. Phased Implementation Roadmap

Phase	Objective	Key Actions	Key Metrics (Validation Signals)	Success Indicators
Phase 1: Immediate Stabilization	Prevent ongoing false-negative exclusions and restore candidate flow	Manual review of near-threshold candidates; Spot-check parsed outputs; Adjust thresholds for critical roles	% near-threshold reviewed; Parsing accuracy rate; Pass-through rate	More candidates reach interviews; Reduced rejection anomalies; Early error detection
Phase 2: Calibration (30–60 Days)	Align scoring and thresholds with hiring outcomes	Analyze historical data; Role-specific thresholds; Test scoring configs; Monitor pass-through	Pass-through rate by role; False-negative proxy; Threshold sensitivity	Stable pass-through; Better alignment with recruiter judgment; Reduced variance
Phase 3: Governance Redesign	Establish lifecycle oversight	Monitoring dashboards; Decision ownership; Audit trails; Override policies	Trend pass-through; Audit completeness; Override frequency; Time to fix errors	Visibility into performance; Clear accountability; Reduced systemic errors
Phase 4: Sustained Adoption	Embed responsible use and continuous improvement	Training; Usage guidelines; Incentives; Feedback loops	User engagement; Training completion; Feedback frequency; Recruiter satisfaction	Improved trust and engagement; Active oversight; Continuous improvement

Conclusion

This case demonstrates that false-negative outcomes in AI-enabled applicant screening are not isolated technical anomalies but reflect a broader socio-technical governance failure within digitally transformed hiring systems. The evidence shows that parsing limitations, lexical matching constraints, and threshold-based filtering do not operate independently; rather, they function as an interconnected decision pipeline in which small representation errors are amplified into large-scale exclusion outcomes. When these systems are implemented without sufficient validation, calibration, and oversight, organizations risk misinterpreting system outputs as accurate reflections of candidate quality, thereby suppressing the flow of qualified talent while believing they are improving efficiency.

This analysis suggests that the core issue is not the use of automation itself, but the absence of governance structures required to align system behavior with organizational hiring objectives. While AI-enabled screening systems provide clear operational benefits in managing scale and reducing recruiter workload, these benefits are undermined when organizations fail to establish decision rights, monitoring mechanisms, and feedback loops that ensure system outputs remain reliable and interpretable over time. As a result, efficiency gains can be offset by degraded selection quality, delayed staffing, and reduced organizational performance.

Therefore, organizations must treat AI-enabled applicant screening not as a standalone technology solution but as a governed decision system embedded within the broader digital transformation of hiring. This requires implementing lifecycle governance practices, including validation of input representation, calibration of matching and threshold logic, and continuous monitoring of downstream outcomes such as candidate pass-through rates and interview yield. Without these controls, automated screening systems will continue to function as high-volume exclusion mechanisms rather than effective decision-support tools, creating sustained risks to both organizational performance and stakeholder trust.

References

Adensamer, A., Gabl, M., & Neumann, G. (2021). Algorithmic decision-making in hiring: Assessing fairness, transparency, and accountability. *AI and Ethics, 1*(3), 365–381. https://doi.org/10.1007/s43681-021-00054-8

Armenakis, A. A., Harris, S. G., & Mossholder, K. W. (1993). Creating readiness for organizational change. *Human Relations, 46*(6), 681–703. https://doi.org/10.1177/001872679304600601

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104*(3), 671–732.

Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., & Venkatraman, N. (2013). Digital business strategy: Toward a next generation of insights. *MIS Quarterly, 37*(2), 471–482. https://doi.org/10.25300/MISQ/2013/37.2.3

Corbett-Davies, S., Gaebler, J., Nilforoshan, H., Shroff, R., & Goel, S. (2023). The measure and mismeasure of fairness. *Journal of Machine Learning Research, 24*(312), 1–117. https://jmlr.org/papers/v24/22-1511.html

Cropanzano, R., Bowen, D. E., & Gilliland, S. W. (2003). Procedural justice and organizational staffing: A tale of two paradigms. *Human Resource Management Review, 13*(1), 7–39. https://doi.org/10.1016/S1053-4822(02)00097-9

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019)* (pp. 4171–4186). https://doi.org/10.18653/v1/N19-1423

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114–126. https://doi.org/10.1037/xge0000033

Fabris, A., Baranowska, N., Dennis, M. J., Graus, D., Hacker, P., Saldivar, J., Zuiderveen Borgesius, F., & Biega, A. J. (2025). Fairness and bias in algorithmic hiring: A multidisciplinary survey. *ACM Computing Surveys*. https://doi.org/10.1145/3696457

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review, 18*(4), 694–734. https://doi.org/10.5465/amr.1993.9402210155

Guan, Z., Liu, S., Zhang, Y., et al. (2024). JobFormer: Skill-aware job recommendation with semantic-enhanced transformer. *ACM Transactions on Information Systems*. https://doi.org/10.1145/3701735

Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. *Advances in Neural Information Processing Systems*.

Hiatt, J. (2006). *ADKAR: A model for change in business, government, and our community*. Prosci.

Jiang, J., & Zhai, C. (2007). An empirical study of tokenization strategies for biomedical information retrieval. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '07)* (pp. 657–664). https://doi.org/10.1145/1277741.1277852

Jobscan. (2025). *ATS statistics: The prevalence and impact of applicant tracking systems*. https://www.jobscan.co/blog/fortune-500-use-applicant-tracking-systems/

Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research, 13*(3), 795–848. https://doi.org/10.1007/s40685-020-00134-w

Kokina, J., & Davenport, T. H. (2017). The emergence of artificial intelligence: How automation is changing auditing. *Journal of Emerging Technologies in Accounting, 14*(1), 115–122. https://doi.org/10.2308/jeta-51730

Kozak, M., & Fel, M. (2024). Trust and transparency in AI-enabled decision-making systems: Implications for user adoption. *Journal of Business Research, 172*, 114421. https://doi.org/10.1016/j.jbusres.2023.114421

Lacroux, A., & Martin, N. (2022). Should I trust the artificial intelligence to recruit? Recruiters' perceptions and behavior when faced with algorithm-based recommendation systems during résumé screening. *Frontiers in Psychology, 13*, 895997. https://doi.org/10.3389/fpsyg.2022.895997

Leichtmann, B., Schlögl, S., & Stary, C. (2023). Explainable AI and user trust: A systematic review of empirical evidence. *Computers in Human Behavior, 139*, 107524. https://doi.org/10.1016/j.chb.2022.107524

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Martin, K. (2019). Designing ethical algorithms. *MIS Quarterly Executive, 18*(2), 129–142.

National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1). https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

Oreg, S., Vakola, M., & Armenakis, A. (2011). Change recipients' reactions to organizational change: A 60-year review of quantitative studies. *The Journal of Applied Behavioral Science, 47*(4), 461–524. https://doi.org/10.1177/0021886310396550

Pleiss, G., Raghavan, M., Wu, F., Kleinberg, J., & Weinberger, K. Q. (2017). On fairness and calibration. *Advances in Neural Information Processing Systems*. https://papers.neurips.cc/paper/7151-on-fairness-and-calibration.pdf

Ployhart, R. E., & Moliterno, T. P. (2011). Emergence of the human capital resource: A multilevel model. *Academy of Management Review, 36*(1), 127–150. https://doi.org/10.5465/amr.2009.0318

Retyk, F., Fabregat, H., Aizpuru, J., Taglio, M., & Zbib, R. (2023). *Résumé parsing as hierarchical sequence labeling: An empirical study*. In *Proceedings of the 3rd Workshop on Recommendation Systems for Human Resources (RecSys in HR '23)* (CEUR Workshop Proceedings, Vol. 3490). CEUR-WS. https://ceur-ws.org/Vol-3490/RecSysHR2023-paper_10.pdf

Rosenberger, J., et al. (2025). CareerBERT: Matching résumés to ESCO jobs in a shared embedding space. *Information Sciences*.

Sanchez, J. I., & Levine, E. L. (2009). What is (or should be) the difference between competency modeling and traditional job analysis? *Human Resource Management Review, 19*(2), 53–63. https://doi.org/10.1016/j.hrmr.2008.10.002

Sanchez, J. I., & Levine, E. L. (2012). The rise and fall of job analysis and the future of work analysis. *Annual Review of Psychology, 63*, 397–425. https://doi.org/10.1146/annurev-psych-120710-100401

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262–274. https://doi.org/10.1037/0033-2909.124.2.262

Sebastian, I. M., Ross, J. W., Beath, C., Mocker, M., Moloney, K. G., & Fonstad, N. O. (2017). How big old companies navigate digital transformation. *MIS Quarterly Executive, 16*(3), 197–213.

Senger, E., Zhang, M., van der Goot, R., & Plank, B. (2024). Deep learning-based computational job market analysis: A survey on skill extraction and classification from job postings. In *Proceedings of the First Workshop on Natural Language Processing for Human Resources (NLP4HR 2024)* (pp. 1–15). Association for Computational Linguistics. https://aclanthology.org/2024.nlp4hr-1.1.pdf

Stouten, J., Rousseau, D. M., & De Cremer, D. (2018). Successful organizational change: Integrating the management practice and scholarly literatures. *Academy of Management Annals, 12*(2), 752–788. https://doi.org/10.5465/annals.2016.0095

Vanetik, N. (2023). Job vacancy ranking with sentence embeddings, explainable matching, and retrieval baselines. *Information, 14*(8), 468. https://doi.org/10.3390/info14080468

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998–6008. https://arxiv.org/abs/1706.03762

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems, 28*(2), 118–144. https://doi.org/10.1016/j.jsis.2019.01.003

Warner, K. S. R., & Wäger, M. (2019). Building dynamic capabilities for digital transformation: An ongoing process of strategic renewal. *Long Range Planning, 52*(3), 326–349. https://doi.org/10.1016/j.lrp.2018.12.001

Weick, K. E. (1995). *Sensemaking in organizations*. Sage.

Weill, P., & Ross, J. W. (2004). *IT governance: How top performers manage IT decision rights for superior results*. Harvard Business School Press.

Yu, K., Guan, G., & Zhou, M. (2005). Resume information extraction with cascaded hybrid model. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL '05)* (pp. 499–506). https://doi.org/10.3115/1219840.1219902

Yu, X., et al. (2025). CONFIT V2: Improving resume–job matching using transformer-based encoders and data augmentation. *Findings of ACL 2025.*