

# Rejected by Code: Algorithmic Bias in AI-Driven Recruitment and the Case for HR-IT Governance Frameworks

**Sowmik Sengupta**

College of Business, California State University, Long Beach, CA, USA

Email: [sowmik.sengupta@gmail.com](mailto:sowmik.sengupta@gmail.com)

**How to cite this paper:** Sengupta, S. (2026) Rejected by Code: Algorithmic Bias in AI-Driven Recruitment and the Case for HR-IT Governance Frameworks. *World Journal of Engineering and Technology*, **14**, 869-888.  
<https://doi.org/10.4236/wjet.2026.144053>

**Received:** August 24, 2026

**Accepted:** October 16, 2026

**Published:** October 19, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Organizations increasingly delegate the first evaluation of job candidates to artificial intelligence, adopted on a dual promise of efficiency and freedom from human prejudice. Scholarship has shown the second promise to be contested: models trained on historical hiring data reproduce past discrimination through facially neutral proxy variables rather than through explicit reference to protected characteristics. Legal scholarship has clarified who may be held liable, and technical scholarship has proposed how such systems might be audited, but the organizational seam between the human resources function that procures a hiring tool and the information technology function that deploys it remains less developed. This paper argues that *Mobley v. Workday, Inc.* exposes that seam. Drawing on a close reading of the litigation alongside scholarship on algorithmic bias, employment law, and AI governance, it advances three propositions: that AI recruitment tools encode and amplify historical bias through proxy variables; that the *Mobley* rulings mark a legal inflection point extending potential liability from the employer to the technology vendor; and that managing this exposure requires a joint HR-IT governance framework in which neither function can discharge its responsibility by deferring to the other, to the vendor, or to the algorithm. The paper proposes such a framework, assigning accountable ownership and explicit decision rights across procurement, configuration, monitoring, and review.

## Keywords

Algorithmic Bias, AI-Driven Recruitment, Applicant Tracking Systems, Disparate Impact, HR-IT Governance, Vendor Liability

## 1. Introduction

### 1.1. Background and Research Question

Derek Mobley applied for jobs at a large number of employers that used Workday, Inc.'s AI-based applicant recommendation system to screen candidates. Mobley holds a graduate degree in economics, is a military veteran, is over the age of forty, is Black, and is a person with a disability. Time and again he was rejected without an interview, often through automated responses issued with no apparent human review. What began as personal frustration resolved, in his account, into a pattern: the common thread across the rejections was not any single employer but the automated system standing between him and each of them. In 2023, he brought suit against Workday itself, not the employers who used it, alleging that its algorithmic screening discriminated against applicants on the basis of race, age, and disability.

The Mobley litigation has become a focal point in a larger transformation of hiring. Organizations increasingly delegate the first evaluation of candidates to artificial intelligence systems that promise speed, scale, and, paradoxically, freedom from human prejudice [1]. Yet a substantial body of scholarship shows that such systems can reproduce and even intensify historical discrimination, because models trained on past hiring data learn the patterns embedded in that data, including patterns that track protected characteristics through seemingly neutral proxy variables [2] [3]. The result is a governance problem that neither technologists nor human resource professionals can resolve alone.

This paper asks how algorithmic bias manifests in AI-powered recruitment tools and application tracking systems (ATS), and what joint HR-IT governance framework organizations must adopt to ensure legal compliance, ethical hiring, and equitable outcomes. Existing research has approached the question from two largely separate directions. Legal scholars have examined how antidiscrimination statutes apply to automated decisions [4], while computer scientists have proposed technical methods for auditing and mitigating bias [5]. Less attention has been paid to the organizational seam between these domains: the point at which an employer's HR function procures a technology it does not build, and its IT function deploys a system whose legal exposure it does not evaluate. Recent work has begun to address AI hiring governance at the level of the organization and its vendors, but the internal division of responsibility between the HR and IT functions, under the US statutory framework, remains less developed.

The central argument of this paper is that *Mobley v. Workday* represents not merely a legal dispute but a failure of organizational governance, specifically a shared accountability gap between the HR and IT functions in the procurement and oversight of AI hiring tools. Three propositions follow. First, AI recruitment tools encode and amplify historical bias through proxy variables rather than through explicit reference to protected classes. Second, the rulings in Mobley mark a legal inflection point that extends potential liability from the employer to the technology vendor itself. Third, managing this risk requires a joint HR-IT governance framework for AI procurement and oversight, one in which neither func-

tion can discharge its responsibility by deferring to the algorithm. The third proposition is the paper's original contribution to the literature.

The significance of this contribution lies in its vantage point. Much of the existing conversation stops at legal analysis of who may be sued. This paper extends the discussion into the operational decisions that precede any lawsuit: how vendors are selected, what contractual guardrails are written, and who monitors outcomes after deployment. Viewing the problem through technology procurement and governance connects the legal exposure documented in the case to the everyday organizational choices that create or prevent it.

## 1.2. Research Approach

This paper is a qualitative, analytical review organized around a single case. Rather than collecting new data, it examines an ongoing lawsuit, *Mobley v. Workday, Inc.*, alongside published scholarship and draws the two together to build its argument. This design suits a question that is interpretive and normative rather than statistical: it asks how bias appears in AI hiring tools and what governance response organizations should adopt.

The case material consists of three orders of the U.S. District Court for the Northern District of California in *Mobley v. Workday, Inc.*, No. 23-cv-00770-RFL: the order on Workday's motion to dismiss [6], the order granting preliminary collective certification of May 16, 2025 [7], and the order of June 22, 2026 on the amended complaint [8]. Each was read to identify what it decided, on what reasoning, and at what procedural stage. The scholarly material consists of twelve scholarly and regulatory sources on algorithmic bias, employment law, and AI governance, selected for their relevance to AI-driven recruitment, and included if they addressed algorithmic hiring, its legal exposure, or AI governance. Sources were located through Google Scholar and general web searches, with full texts retrieved from publisher platforms, SSRN, and the issuing agencies; searches were run using terms such as algorithmic bias, AI recruitment, disparate impact, and HR-IT governance in AI hiring, most recently in September 2026. The literature was assembled by citation chaining from seed sources rather than by systematic review, so coverage is not exhaustive, a limitation noted in the conclusion. The paper synthesizes these sources in the literature review, applies them to the case in the discussion, and uses that analysis to propose a joint HR-IT governance framework. Because the framework is drawn from existing evidence rather than tested empirically, its contribution is conceptual, a limitation noted in the conclusion.

## 2. Literature Review

### 2.1. AI in Recruitment: Evolution and Adoption

The use of technology to screen job applicants is not new, but its character has changed. For decades, applicant tracking systems performed clerical work, storing resumes and matching keywords, while human recruiters retained judgment over who advanced. Over the past decade that division has shifted. Employers now de-

ploy machine learning systems that rank, score, and filter candidates, functions that once belonged to people [1]. This reflects a broader movement in which employers use algorithmic systems to take over direction, evaluation, and screening functions once performed by managers and recruiters, relocating human judgment into automated tools [9]. The appeal is straightforward. Hiring at scale is expensive and slow, and organizations receive far more applications than any recruiter can read, so tools that promise to surface the strongest candidates automatically address a genuine operational pressure. The reach of these systems is now substantial: in one large employer survey, more than 90% of firms used automated systems to initially filter or rank candidates, including 94% for middle-skill roles and 92% for high-skill roles [10]. Screening by algorithm is no longer an experiment at the margins of hiring; for many applicants it is the default first gate.

Two motivations drive adoption, and the second is where the tension begins. The first is efficiency: automated screening reduces time-to-hire and cost-per-hire. The second is a claim about fairness. AI hiring tools have also been promoted on the rationale that removing humans from the initial screen removes human prejudice with them; by standardizing criteria and stripping out information irrelevant to performance, such as a candidate's race or sex, algorithms are said to reduce the bias a human screener might introduce [1]. This framing positioned AI not merely as faster but as a corrective to discrimination.

That promise rests on a questionable assumption, which later sections examine in detail: that a system trained on an organization's past hiring decisions will be neutral. If those past decisions carried bias, the model learns the bias as a pattern to reproduce [2]. The very feature that makes these tools attractive, their ability to learn what a "successful" hire looks like from historical data, is also what allows them to inherit historical discrimination. Recognizing this tension is essential to understanding the Mobley dispute, because Workday's defense rests substantially on the objectivity claim, while the plaintiffs' case rests on the argument that the tools are only as fair as the data and design behind them.

In short, AI recruitment tools have moved from supporting human decisions to making them, and they have been adopted on a dual promise of efficiency and fairness. The efficiency is largely real. The fairness is contested, and it is that contested claim that the remainder of this review, and the case itself, interrogates.

## **2.2. Algorithmic Bias: Definition, Sources, and Proxy Variables**

Algorithmic bias refers to the systematic production of outcomes that disadvantage a group defined by a protected characteristic such as race, age, or disability. Mehrabi *et al.* [11] catalog numerous forms this can take, but for hiring the important point is that bias rarely enters as explicit prejudice coded into a rule. It enters through the data and design choices that shape what the system learns.

The most examined source is the training data. A machine learning model built to identify promising candidates is typically trained on an organization's records

of past hires and past performance. If the organization historically favored certain groups, the model learns those historical patterns as the definition of a good candidate and reproduces them at scale [2]. The bias is not introduced by the algorithm so much as inherited from the human decisions the algorithm was asked to imitate. Barocas and Selbst [2] further note that even the choice of what to predict, the “target variable,” can embed bias, because a proxy for success such as tenure or manager ratings may itself reflect discriminatory conditions.

This leads to the mechanism most central to the Mobley dispute: proxy variables. A model does not need access to an applicant’s race, age, or disability status to discriminate on those grounds. Facially neutral data points can stand in for protected characteristics because they correlate with them. Residential zip code can proxy for race, dates such as graduation year can proxy for age, and gaps in employment history can proxy for disability or caregiving status. The last of these is particularly consequential: in one survey, 48% of employers filtered middle-skill candidates on the basis of employment gaps longer than six months [10], a filter that effectively screens out applicants whose histories reflect medical leave or rehabilitation, tracking disability without ever naming it. Evidence also shows how directly models can reconstruct protected traits from ordinary application data. Testing large language models in a resume-screening task, Wilson and Caliskan [12] found that resumes with White-associated names were significantly preferred in 85.1% of tests, while resumes with Black-associated names were preferred in only 8.6%. Removing the protected attribute from the model therefore does not remove the bias; the model reconstructs the attribute from its correlates [2] [3]. This is why the common defense that a system is “blind” to protected characteristics is technically true and practically insufficient.

The deeper implication, developed by Selbst *et al.* [13], is that bias is not merely a technical defect to be patched but a property of the sociotechnical system in which the tool is built, sold, configured, and used. Purely technical de-biasing efforts, they argue, tend to fail because they abstract the model away from the organizational context that produces the harm. Raghavan *et al.* [3] reinforce this empirically, finding that vendors’ claims to have mitigated bias vary widely in rigor and transparency, with little external verification. Taken together, this literature establishes two claims the remainder of the paper relies on: that algorithmic bias in hiring is real and largely traceable to data and design rather than malice, and that fixing it cannot be left to the technology alone. Both claims point toward governance.

### 2.3. Legal Framework: Title VII, the ADA, and the ADEA

Three federal laws govern the discrimination claims at the center of AI-driven hiring. Title VII of the Civil Rights Act of 1964 [14] prohibits discrimination based on race, color, religion, sex, and national origin. The Age Discrimination in Employment Act (ADEA) [15] protects workers aged forty and older. The Americans with Disabilities Act (ADA) [16] protects qualified individuals with disabilities

and, notably, limits an employer's ability to make medical inquiries before a job offer. Together they map onto the Mobley plaintiffs, whose claims span race, age, and disability.

The concept that ties these laws to algorithmic hiring is disparate impact. Discrimination law recognizes two paths to liability: disparate treatment, meaning intentional differential treatment, and disparate impact, meaning a neutral practice that nonetheless screens out a protected group at a higher rate without adequate business justification. Disparate impact is the better fit for algorithmic hiring, because AI screening rarely reflects intent to exclude; it produces skewed outcomes through the neutral-seeming mechanisms discussed above.

The three statutes are not interchangeable on this point, however, and the difference matters for any governance framework built on them. Under Title VII, once a plaintiff establishes that a selection practice produces a disparate impact, the employer must demonstrate that the practice is job-related for the position in question and consistent with business necessity; the plaintiff may then prevail by identifying a less discriminatory alternative that the employer refused to adopt [14] [17].

The ADEA framework is materially more permissive. In *Smith v. City of Jackson* [18], the Supreme Court held that disparate-impact claims are cognizable under the ADEA but that the employer's defense is that the practice rests on reasonable factors other than age, a standard the Court expressly distinguished from, and described as narrower in reach than, Title VII's business-necessity test. The less-discriminatory-alternative inquiry that Title VII permits does not transfer to the ADEA in the same form.

The ADA differs again: a qualification standard or selection criterion that screens out individuals with disabilities must be job-related and consistent with business necessity, but the employer must additionally show that the required performance cannot be achieved with reasonable accommodation [16]. Because the Mobley claims span all three statutes, and because the age claim is the one on which preliminary collective certification was granted, a framework that applies a single uniform test across them would misstate the governing standard for the paper's most developed claim. Mobley's age claim proceeds on this basis.

Regulators have started to respond. The Equal Employment Opportunity Commission has issued technical assistance confirming that software, algorithms, and AI tools fall under existing obligations of the ADA and Title VII. That guidance is influential but does not carry the force of law, and it shares a built-in assumption with the statutes themselves: it presumes the employer is the party responsible. Whether that responsibility also reaches the technology vendor that builds and runs the screening system is the question the next section takes up.

## **2.4. Vendor Liability and the Agent Question**

The literature reviewed so far treats the employer as the party responsible for a biased hiring decision. Automated recruitment complicates that assumption, because the tool doing the screening is often built, trained, and operated by a third

party the employer merely licenses. This raises the question at the center of the Mobley dispute: when a vendor's algorithm screens out applicants, who is legally answerable, the employer, the vendor, or both?

The traditional answer places responsibility on the employer. Antidiscrimination law generally holds that an employer cannot escape liability by outsourcing a discriminatory practice; if a selection procedure produces a disparate impact, the employer remains responsible even when an outside vendor designed it [2]. The EEOC has affirmed this directly, stating in its technical assistance that an employer may be liable under Title VII or the ADA for a vendor-built tool, and further that employers "may be held responsible for the actions of their agents, which may include entities such as software vendors, if the employer has given them authority to act on the employer's behalf" [19] [20]. Notably, the EEOC framed this in the language of agency well before the Mobley rulings, signaling that the vendor's role was a live legal question at the regulatory level, not merely a novel theory advanced by one plaintiff.

The harder and more consequential question is whether the vendor itself can be held directly liable. This is where agency becomes central. Under an agency theory, a party that performs a function on another's behalf, and with authority delegated by that other party, can bear legal responsibility for how it performs that function. Applied to hiring technology, the argument is that a vendor whose system does not merely store or transmit applications, but actively scores, ranks, and effectively rejects candidates, is performing a screening function traditionally carried out by the employer, and is therefore acting as the employer's agent rather than as a passive tool [21]. Ajunwa [21] argues that automation should not dissolve accountability, contending that "the creators of runaway algorithms should not be permitted to disavow their creations" but must instead "bear the ultimate liability for any harm wrought by their creations" ([21]: p. 1707).

The distinction that determines the outcome is the one between a tool and a decision-maker. A vendor that provides neutral infrastructure, executing criteria the employer alone specifies, looks like a tool and attracts little independent liability. A vendor whose model independently determines which applicants advance looks like a participant in the employment decision, and it is that characterization that opens the door to direct liability. This tool-versus-participant line, which the next section shows the Mobley court drawing explicitly, is the pivot on which vendor liability turns, and it is precisely the seam where the governance failure this paper identifies takes root: an employer that treats the system as a mere tool while the vendor treats itself the same way leaves the screening function accountable to no one.

## 2.5. Existing Governance Models for AI in Hiring

If algorithmic bias cannot be solved by the technology alone, the question becomes how organizations should govern the tools they adopt. A growing body of work has proposed answers, and they fall into three broad groups, each useful and each incomplete for the problem this paper identifies.

The first group is technical governance: methods for auditing and testing algorithmic systems for bias. Raji *et al.* [5] propose an end-to-end internal auditing framework that traces an AI system's behavior across its development lifecycle, with the aim of catching harmful outcomes before deployment. Such audits are valuable, but they are principally engineering practices, situated inside the team that builds or configures the model. They tell an organization how to inspect a system; they say less about who inside the organization owns that obligation, or how it survives the moment a tool is purchased from an outside vendor rather than built in-house.

The second group is sociotechnical governance, which argues that bias cannot be understood or fixed within the model in isolation. Selbst *et al.* [13] show that purely technical de-biasing efforts fail when they abstract the system away from the social and organizational context in which it operates. This literature correctly locates the problem at the level of the whole system rather than the code, but it is largely diagnostic. It establishes that governance must be organizational without specifying which organizational functions must act, or how they should divide responsibility.

The third group is standardized risk governance, exemplified by the National Institute of Standards and Technology's AI Risk Management Framework (2023), which offers organizations a structured, voluntary process for identifying and managing AI-related risks across a system's lifecycle. The framework is comprehensive and increasingly influential, and it provides useful vocabulary for mapping and measuring risk. Its limitation for present purposes is that it is general by design, written to apply to any AI system in any sector, and so it does not address the specific division of labor in hiring, where a human resources function selects and procures a tool that an information technology function deploys and maintains.

Across all three groups, a common assumption is visible. Each treats governance as something an organization does to a system, and each is largely silent on the internal boundary that this paper argues is decisive: the line between the HR function, which owns the hiring process and its legal exposure, and the IT function, which owns the technology and its operation. Existing frameworks tell organizations to audit, to think in sociotechnical terms, and to manage risk, but they do not resolve who is accountable when a biased hiring tool is procured by one function and run by another, each reasonably assuming the other has addressed the problem. That unresolved boundary, surfaced starkly by the Mobley case, is the gap the framework proposed in this paper is designed to close.

### 3. Analysis & Framework

#### 3.1. Mobley v. Workday: Case Analysis

Because this analysis draws on a matter in active litigation, the discussion that follows distinguishes plaintiffs' allegations and the court's preliminary procedural rulings from established fact. Surviving a motion to dismiss establishes only that

a claim is plausibly pleaded; no merits determination has been made, and the question of vendor liability under an agency theory remains unresolved.

The dispute began with one applicant's persistence. Since 2017, Derek Mobley applied to more than one hundred positions at companies that screened candidates through Workday's platform and was, he alleges, rejected from every one [7]. In 2023 he sued Workday itself, and as the case developed, four additional plaintiffs over the age of forty joined, each alleging they had applied for hundreds of jobs through Workday and been rejected almost every time, often through automated responses. The claims spanned race, age, and disability, brought under Title VII, the ADA, the ADEA, and California law, and they proceeded on a disparate impact theory: not that Workday intended to discriminate, but that its screening produces discriminatory outcomes.

Two features of the platform sit at the center of the case. The first, Candidate Skills Match, scores applicants against a job posting and reports results to the employer on a scale from "strong" down to "low." The second, the Workday Assessment Connector, is alleged to use machine learning in a way that can, in the plaintiffs' description, build a "homogenous workforce" by learning which candidates an employer's existing patterns favor. On the plaintiffs' account, the significance of these tools is that they do not merely store applications; they evaluate them, performing the screening judgment a human recruiter once exercised.

Workday's defense rested on a single characterization: that it is a neutral tool, not a decision-maker. It argued that employers choose whether to enable its AI features, that it does not itself reject applicants, and that any discriminatory outcome traces to the employer's own preferences rather than to Workday's software. If accepted, this framing would place all legal responsibility on the employer and none on the vendor.

The court declined to accept it at the pleading stage, and the reasoning is the pivot of the entire case. In denying Workday's motion to dismiss, the court found that Mobley had plausibly alleged that Workday's tools were "participating in the decision-making process by recommending some candidates to move forward and rejecting others," rather than "simply implementing in a rote way the criteria that employers set forth" [6]. The court drew an explicit line: if Workday's system merely executed criteria the employer specified, it would be, in the court's own analogy, "no different from an excel spreadsheet operated by the employer," and no vendor liability would attach. But because the system was alleged to exercise its own screening judgment, it could plausibly be treated as the employer's agent, engaged in "conduct that is at the heart of equal access to employment opportunities." That distinction, between a passive tool and an active participant, is what allowed the claims against the vendor to survive.

In May 2025, the case advanced further when the court granted preliminary certification of a nationwide collective action on the age discrimination claim, covering applicants aged forty and over who were denied recommendations through Workday's platform since September 2020 [7]. The scale is striking. In opposing notice to

the collective, Workday stated that “1.1 billion applications were rejected using Workday” during the time period at issue—a rough estimate, as the court characterized it, covering all rejections across Workday’s entire client base and carrying no implication about the rate at which any protected group was screened out. Whatever the ultimate outcome, the certification transformed the matter from one applicant’s grievance into a test case with industry-wide implications, and it did so on the strength of the same theory, that a vendor’s screening tool can operate as a common policy producing disparate impact across employers.

The case continued to expand through 2026. Following amended complaints that added plaintiffs and new claims, the court issued a significant ruling on June 22, 2026, that broadened the litigation in two ways relevant to this analysis [8].

First, the court declined to dismiss claims under the California Fair Employment and Housing Act, reasoning that because Workday designs, maintains, and operates its screening algorithms from its California headquarters, a sufficient territorial connection exists to apply California law even to applicants residing outside the state.

Second, and most relevant here, the court permitted a disability claim to proceed under a proxy discrimination theory, accepting that a machine-learning system need not record disability status directly to discriminate on that basis; it may infer and screen out candidates through correlated signals such as lengthy employment gaps or patterns consistent with medical leave and recovery. This is the same mechanism examined earlier in this review, now recognized by a federal court as a plausible basis for liability rather than a purely theoretical concern.

### **3.2. Procurement without Oversight—The HR Gap**

If Mobley reveals a governance failure rather than merely a legal one, the first place to locate it is in how organizations acquire hiring technology. Human resources departments are the functional owners of the hiring process; they select the tools, define the criteria, and answer for the outcomes. Yet the evidence in Mobley suggests that the human resources function too often procures algorithmic screening tools the way it would procure any other software service, as a purchased efficiency, without the oversight that its own legal exposure demands.

The core failure is treating an AI hiring tool as a neutral utility rather than as a delegation of a core HR responsibility. When an employer enables an automated screen, it is not outsourcing a clerical task; it is handing the judgment of who advances, historically an HR judgment, to a system it did not build and often does not understand. The EEOC has been explicit that this delegation does not transfer the liability: an employer “may be held responsible for the actions of their agents, which may include entities such as software vendors, if the employer has given them authority to act on the employer’s behalf” [19] [20]. In other words, the law already treats the procured tool as an extension of the employer, even when the HR function treats it as a detached product.

Two oversight gaps follow from this mis-framing. The first is at the point of

selection. The EEOC's guidance lists concrete questions an employer should ask a vendor before purchase: whether the tool was tested for disparate impact, whether it makes disability-related inquiries, whether its measured traits correlate with protected characteristics [19] [20]. These are procurement questions, and they are HR's to ask. When they go unasked, bias enters the organization not through a hiring manager's prejudice but through the procurement decision itself.

The second gap is the absence of any ongoing duty of care after purchase. Vendor claims of a "bias-free" or "validated" tool are frequently accepted at face value, yet the scholarship shows such claims vary widely in rigor and are rarely subject to external verification [3]. An HR function that relies on the vendor's marketing rather than on independent evidence has not exercised oversight; it has substituted the vendor's judgment for its own.

The Mobley facts illustrate the cost of this posture. Employers using Workday's platform relied on scores and recommendations, "strong," "good," "fair," "low," generated by a system whose internal workings they did not control and, in many cases, could not inspect. On the plaintiffs' account, applicants were screened out by a purchasing decision made upstream, with no human in the process positioned to catch a discriminatory pattern. This is the HR failure at the heart of the case: not that the technology was adopted, but that it was adopted without the governance that the delegation of a hiring decision requires. Adoption is a legitimate business choice; abdication of oversight is not.

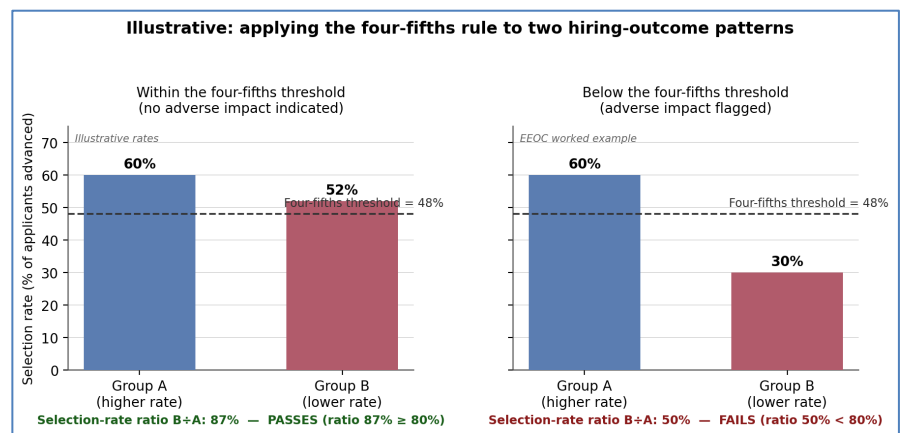
### 3.3. Deployment without Guardrails—The IT Gap

If the human resources function fails at the point of selection, the information technology function fails at the point of deployment. Once a hiring tool is procured, it must be integrated, configured, connected to applicant data, and maintained, and these are IT responsibilities. Yet the same mis-framing that lets HR treat the tool as a neutral purchase lets IT treat it as neutral infrastructure: a system to be stood up and kept running, rather than one whose outputs carry legal and ethical weight. The result is deployment without guardrails.

The central failure is the assumption that technical soundness equals fitness for use. An IT function evaluates a system on availability, security, integration, and performance, and an AI hiring tool can pass every one of those tests while still producing discriminatory outcomes. A model that runs flawlessly and returns results quickly is, from a conventional IT standpoint, a success. But the design choices that determine whether the tool discriminates, what data it was trained on, which variables it weighs, whether it reconstructs protected characteristics from proxies, sit outside the traditional IT checklist. Selbst *et al.* [13] make precisely this point: treating fairness as a property that can be engineered into a model in isolation, abstracted from the social context of its use, is why technical interventions so often fail. Bias is a property of the whole sociotechnical system, and deploying a tool without accounting for that context is a governance gap, not merely a technical oversight.

The most consequential and least examined of these gaps is configuration governance. Enterprise hiring platforms are not deployed in a single fixed state; they expose configurable settings that materially change who is screened out. Employers can enable or disable AI features, and scoring tools report results on graded scales, such as the “strong,” “good,” “fair,” and “low” bands generated by the system in Mobley, that require someone to decide where an acceptable score ends and a rejection begins. That threshold is not a technical default; it is a policy decision with direct disparate-impact consequences, because moving the cutoff changes which protected groups clear the bar. In practice, such settings are frequently configured once during implementation, by technical staff optimizing for fit and throughput, and then left unexamined, with no defined owner accountable for their downstream effect on protected classes. The organization that would never let a single administrator quietly set a hiring policy nonetheless allows a configuration screen to set one by default. Deciding who owns these thresholds and requiring that changes to them be reviewed for adverse impact rather than treated as routine system configuration, is a guardrail that ad hoc deployment omits.

The legal test for whether a resulting pattern is lawful is concrete. Under the four-fifths rule described in EEOC guidance, a selection rate for one group that is less than 80% of the rate for the most-selected group may indicate adverse impact [20]. **Figure 1** illustrates this test applied to two outcome patterns, one within the threshold and one, drawn from the EEOC’s own worked example, that falls below it and would be flagged for review. The four-fifths rule is a rule of thumb rather than a bright line; the same guidance notes it can be inappropriate for very large samples and that courts sometimes rely on tests of statistical significance instead [20]. Even as a rough screen, however, it gives an organization a concrete trigger for review that ungoverned deployment simply never applies.



Note. The right panel uses the worked example from EEOC (2023): 80 White and 40 Black applicants, with selection rates of 60% and 30%, yielding a ratio of 50%, which falls below the four-fifths (80%) threshold and indicates adverse impact. The left panel uses illustrative rates to show a pattern that clears the threshold. Values are illustrative of the legal standard and do not represent measured outcomes from any specific hiring tool.

**Figure 1.** Applying the four-fifths rule to two hiring-outcome patterns.

Two further guardrails are typically absent. The first is monitoring after deployment. A model's behavior can shift as applicant pools change, so a system acceptable at launch can develop a disparate impact over time; without ongoing outcome monitoring, no one detects the change. The second is auditability. When an applicant is rejected, the organization often cannot explain why, because the decision lives inside a vendor model it cannot inspect. Raji *et al.* [5] argue that closing this accountability gap requires deliberate internal auditing across a system's lifecycle, precisely the end-to-end scrutiny that unstructured deployment lacks.

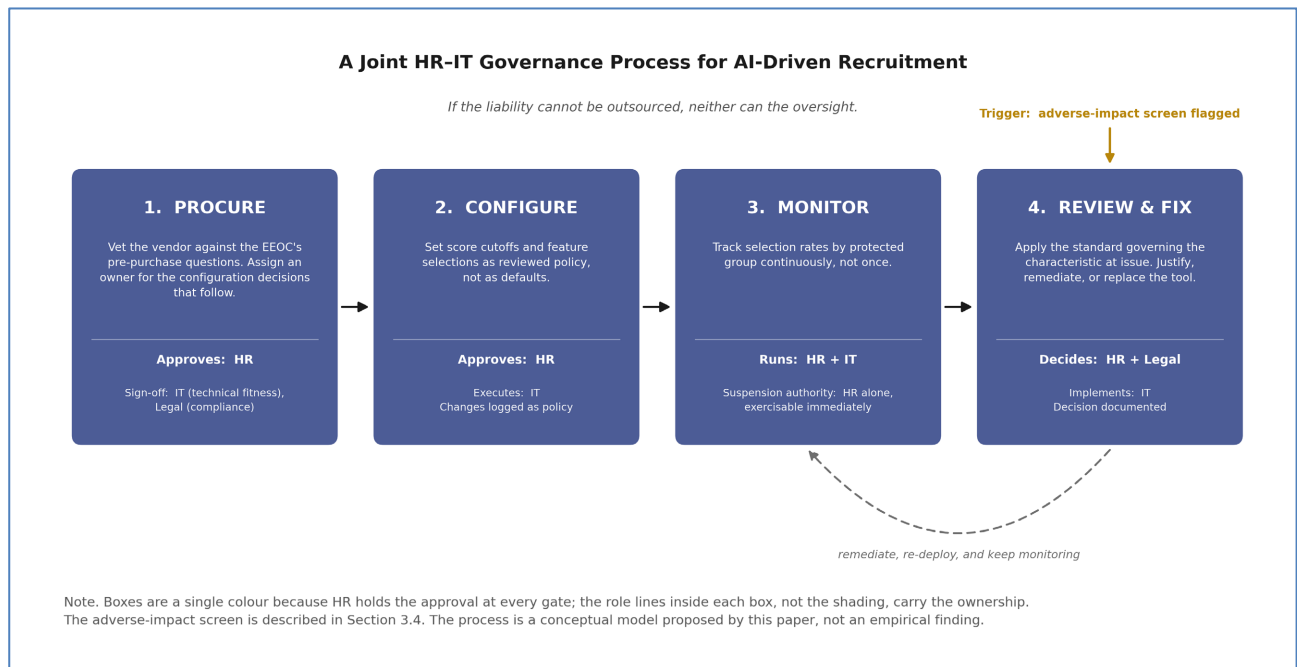
The allegations in Mobley describe what deployment without guardrails can produce. A system was integrated as the first gate every applicant passed through; it scored and sorted candidates at scale; and it did so without configuration review, outcome monitoring, or any mechanism to explain a rejection. From an operational standpoint the deployment worked. From a governance standpoint it was ungoverned, and that distinction is the point: the failure was not that the system malfunctioned, but that its normal, functioning output was never treated as something requiring oversight. As with the human resources failure, the problem is not deployment itself but deployment stripped of the guardrails that a consequential automated decision demands.

### 3.4. A Proposed HR-IT Governance Framework

The two preceding sections describe one gap seen from two sides. HR fails at procurement by treating a hiring tool as a neutral purchase; IT fails at deployment by treating it as neutral infrastructure. Neither failure is the whole problem, and neither function can close the gap alone, because the accountability falls precisely in the space between them. The central argument of this paper is that responsibility for algorithmic hiring is shared and cannot be delegated, neither to the other function nor to the vendor nor to the algorithm itself, and that organizations therefore need a governance framework that assigns joint HR-IT ownership across the full life of the tool. This framework is the paper's original contribution.

The organizing principle is that the law already treats the tool as an extension of the employer, so the organization must govern it as one. When a court in Mobley distinguished a system that merely executes an employer's criteria, "no different from an excel spreadsheet," from one that "participate[s] in the decision-making process," it drew the same line this framework depends on [6] [7]. A tool that exercises screening judgment carries the employer's legal exposure, and the EEOC has confirmed that this exposure is not transferred to the vendor by contract [19] [20]. If the liability cannot be outsourced, then the oversight cannot be either. The framework operationalizes that principle by naming who does what, and when.

**Figure 2** presents the framework as a lifecycle. It begins before any tool is deployed, at a procurement gate jointly owned by HR, IT, and legal counsel, where the EEOC's own vendor questions are asked and answered: has the tool been tested for disparate impact, does it make disability-related inquiries, do its

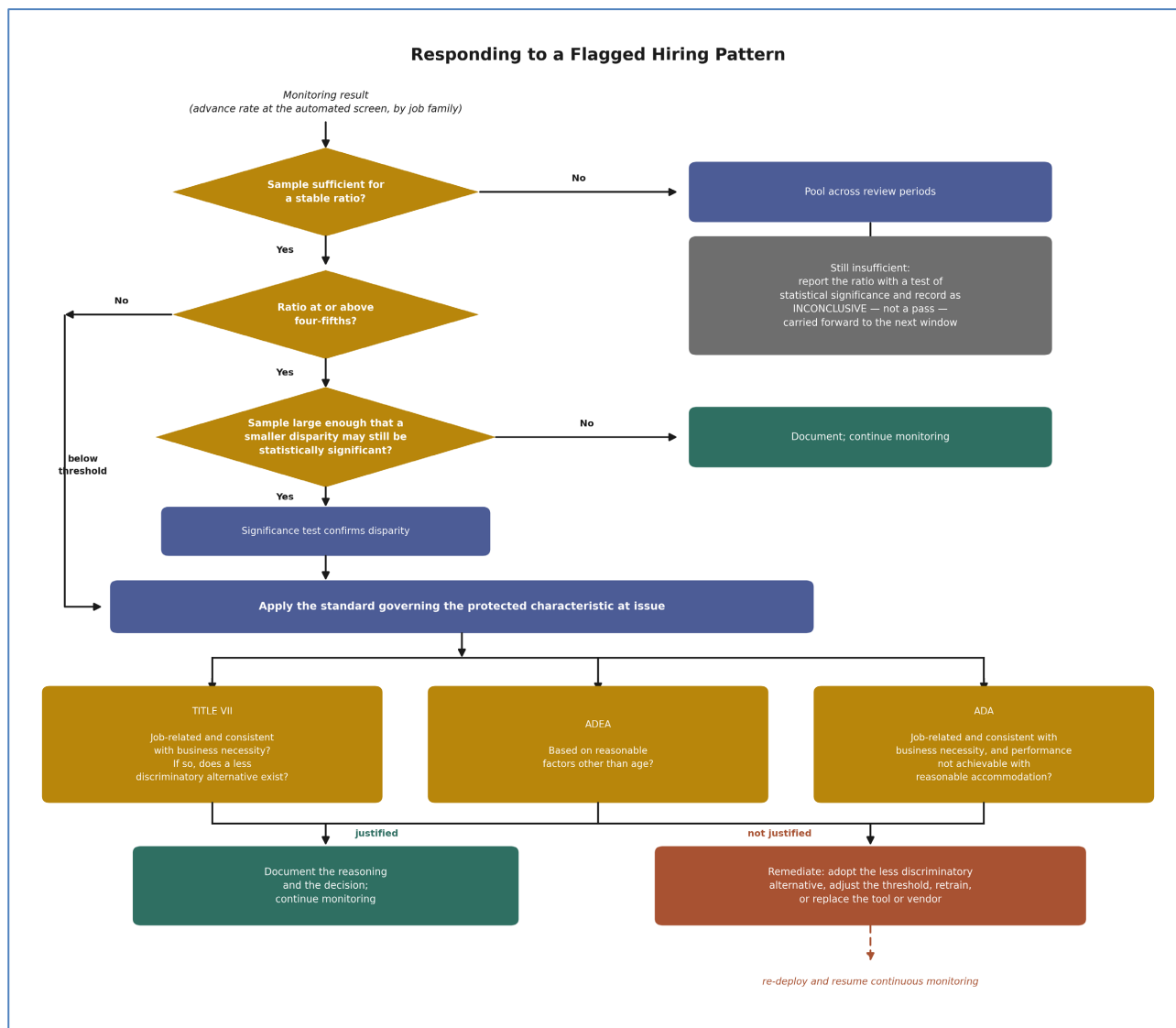


Note. Each stage names the function that holds the approval and the functions that execute or advise; approval rests with HR at every gate because HR carries the employer's legal exposure. The adverse-impact screen is described in Section 3.4. The process is a conceptual model proposed by this paper, not an empirical finding.

**Figure 2.** A Joint HR-IT governance process for AI-Driven recruitment.

measured traits correlate with protected characteristics [19] [20]. Critically, this gate also assigns an accountable owner for the configuration decisions that follow, closing the gap identified in Section 3.3. At deployment, IT configures the tool as reviewed policy rather than default settings, with score thresholds treated as consequential choices subject to review. Once live, HR and IT jointly monitor outcomes, measuring selection rates by protected class against the four-fifths screen described earlier. A pattern that falls below that threshold triggers a defined review by HR and legal, which applies the standard governing the protected characteristic at issue: for Title VII claims, whether the practice is job-related and consistent with business necessity, and whether a less discriminatory alternative exists; for ADEA claims, whether the practice rests on reasonable factors other than age; and for ADA claims, whether the screening criterion is job-related and consistent with business necessity and whether the required performance could be achieved with reasonable accommodation [16]-[18] [20]. **Figure 3** sets out this decision sequence, routing by the applicable statutory standard. If the practice cannot be justified, remediation follows: adjusting the threshold, retraining, or replacing the tool or vendor. Each decision, including a decision that no adverse impact exists, is documented, and monitoring continues. The loop is deliberate, because a model acceptable at launch can drift, and one-time approval is not governance.

Monitoring stated as an intention is not monitoring, so the framework specifies what is measured and when. The ratio is computed at the automated screen itself,



Note. Decisions are shown in gold, process steps in slate, compliant outcomes in green, and required action in red-brown. A ratio at or above four-fifths does not close the review: where the sample is large, a smaller disparity may still be statistically significant and is assessed before the pattern is cleared. Where the sample is too small for a stable ratio, results are pooled across review periods and otherwise recorded as inconclusive rather than compliant. The justification step applies the standard governing the protected characteristic at issue ([14] [16]-[18]). The model is a conceptual decision aid, not legal advice.

**Figure 3.** Responding to a flagged hiring pattern.

comparing the rate at which applicants from each group advance out of that stage, because that is the gate at which the harm occurs and the only point at which a disparity can be traced to the tool rather than to a later human judgment. End-to-end rates are tracked as a secondary check but cannot substitute for it. Rates are pooled by job family rather than per requisition, grouping postings that share a screening configuration, since individual openings are rarely large enough to yield a stable ratio. Protected-class data comes from voluntary applicant self-identification, stored apart from the record the tool reads. Disability requires a different approach: because the ADA limits pre-offer inquiry, the framework audits the

screening criteria for proxy effects rather than computing a rate against disability status, testing whether filters such as employment-gap thresholds track the correlated signals the *Mobley* court accepted as a plausible basis for liability.

IT maintains the data continuously, HR and IT review it formally each quarter, and a review is triggered outside that cycle by any threshold change, model retrain, or vendor version update. Where a job family is too small for a meaningful comparison, results are pooled across periods; where the sample remains insufficient, the ratio is reported with a test of statistical significance and recorded as inconclusive rather than compliant, carried forward to the next window. That distinction matters, because an inconclusive finding treated as a pass reproduces the vacuum this framework exists to close.

Shared responsibility is not shared decision-making, and a framework that stops at joint ownership risks reproducing the diffusion it sets out to correct. Each checkpoint therefore carries a named approver, on the rule that decision rights follow legal exposure: because HR remains accountable for who is hired and who is excluded, HR holds the decision at each gate, while IT executes and legal counsel exercises a defined veto. Deployment requires sign-off from IT on technical fitness and legal on compliance, but HR owns the approval. Threshold changes are executed by IT and approved by HR, logged as policy rather than configuration, so no cutoff is altered unilaterally by technical staff. Authority to suspend a tool sits with HR alone and is exercisable immediately, without awaiting a review, because a flagged pattern that keeps screening applicants while committees convene compounds the exposure the framework exists to limit. Remediation, including a decision to accept a disparity as justified, is decided jointly by HR and legal, implemented by IT, and documented with its reasoning. Naming these rights is what converts joint ownership into an allocation of accountability. The particular allocation proposed here is one defensible configuration rather than the only one; organizations may assign these rights differently to fit their structures. What the framework treats as non-negotiable is that the rights be assigned and documented, since an unnamed approver is the condition under which accountability disperses.

What distinguishes this framework from the existing models reviewed earlier is not that it invents new controls but that it assigns them across the organizational seam those models leave open. Technical auditing frameworks specify how to inspect a system; Raji *et al.* [5] are explicit that determining what to audit and why lies outside their scope ([5]: p. 38). That scoping leaves open the organizational question this paper foregrounds: which function owns the obligation once a hiring tool is purchased rather than built. Sociotechnical accounts establish that governance must be organizational without naming the functions responsible [13]. Standardized risk frameworks such as the NIST AI Risk Management Framework provide a general lifecycle process applicable to any AI system in any sector, but by design they do not address the specific division of labor in hiring, where HR procures a tool that IT deploys [22]. This framework's contribution is precisely

that specificity: it maps each governance task to an accountable function and defines the joint checkpoints, procurement, configuration, monitoring, and review, at which HR and IT must act together. It treats the HR-IT boundary not as a place where responsibility is handed off, but as a place where it is shared.

The framework is a conceptual proposal rather than a validated instrument. Its logic rests on the legal standards and the scholarship synthesized in this paper, not on empirical testing in operating organizations, and its effectiveness would depend on factors this analysis cannot measure, such as organizational willingness to fund ongoing monitoring or to act on an unfavorable finding. These limitations, and the further research they invite, are considered in the conclusion. What the framework offers is a direct response to the failure Mobley exposes: not a better algorithm, but a clear answer to the question the case leaves employers facing, which is who, inside the organization, is answerable when the code says no.

#### 4. Conclusions

This paper set out to answer how algorithmic bias manifests in AI-powered recruitment tools, and what joint governance framework organizations must adopt in response. The three propositions that opened the paper are borne out by the analysis. AI recruitment tools encode and amplify historical bias through proxy variables rather than explicit prejudice, a mechanism established in the literature and, in the disability context, now recognized by a federal court as a plausible basis for liability. The rulings in *Mobley v. Workday* mark a genuine inflection point, extending potential exposure from the employer to the vendor whose system exercises screening judgment. And managing that exposure requires a governance framework in which neither the human resources function nor the information technology function can discharge its responsibility by deferring to the other, to the vendor, or to the algorithm. Because the litigation remains unresolved, however, the extension of liability to vendors described here is an emerging rather than settled doctrine; the agency question, the merits of the disparate-impact claims, and the scope of the preliminarily certified collective all remain open.

The paper's contribution is to locate the problem where existing scholarship does not. Legal analysis has clarified who may be sued, and technical scholarship has proposed how systems might be audited, but the organizational seam between the function that procures a hiring tool and the function that deploys it has remained underexamined. Mobley exposes that seam. Workday's defense, that it was a mere tool, and the employers' implicit posture, that responsibility ran to the vendor, describe a gap in which an automated hiring decision would be accountable to no one. The framework proposed here treats that boundary not as a place where responsibility is handed off but as a place where it is jointly held, assigning specific governance tasks, procurement, configuration, monitoring, and review, to accountable functions across the life of the tool.

For human resource practice, the implication is direct. The adoption of AI screening is a legitimate response to real operational pressure, but it is not a dele-

gation that relieves HR of accountability for who is hired and who is excluded. HR professionals accustomed to owning the fairness of a hiring process must now own it through a technology they did not build, which means asking procurement questions they may not have asked, insisting on monitoring they may not have required, and treating a vendor's assurance of a "bias-free" tool as a claim to be verified rather than accepted. Equally, the technology function must recognize that a system passing every conventional test of performance can still produce unlawful outcomes, and that configuration and deployment choices are policy decisions, not merely technical ones.

Several limitations bound these conclusions. The framework is a conceptual proposal grounded in legal standards and scholarship rather than a validated instrument tested in operating organizations; its effectiveness would depend on factors this analysis cannot measure, including whether organizations will fund ongoing monitoring and act on unfavorable findings. The analysis also rests on a single case still in active litigation, so its rulings are preliminary and may yet change on further proceedings or appeal, and the four-fifths screen it relies upon is a rule of thumb rather than a definitive test of discrimination. Because the literature was assembled by citation chaining rather than systematic review, closely adjacent work may not be represented. These limitations point to the research the field now needs. Empirical study of how organizations actually divide responsibility for AI hiring tools would test whether the accountability gap this paper describes is as common as the case suggests. Evaluation of the proposed framework in practice would show whether joint HR-IT governance measurably reduces disparate impact or merely adds process. And as litigation and regulation continue to develop, further work should track whether vendor liability, still an emerging doctrine, hardens into a settled expectation that reshapes how AI hiring tools are built, sold, and governed.

In the end, *Mobley v. Workday* is less a story about a flawed algorithm than about an unanswered question of responsibility. On the account the litigation presents, the technology performed as designed; what was missing was any person or function prepared to answer for its outcomes. As automated systems take over more of the hiring decision, that question, who inside the organization is accountable when the system says no, will only grow more pressing. The contribution of this paper is to insist that the answer cannot be the algorithm, and to offer a framework for ensuring it is not.

### Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

### References

- [1] Tambe, P., Cappelli, P. and Yakubovich, V. (2019) Artificial Intelligence in Human Resources Management: Challenges and a Path Forward. *California Management Review*, **61**, 15-42. <https://doi.org/10.1177/0008125619867910>
- [2] Barocas, S. and Selbst, A.D. (2016) Big Data's Disparate Impact. *California Law Re-*

- view, 104, 671-732. <https://doi.org/10.15779/Z38BG31>
- [3] Raghavan, M., Barocas, S., Kleinberg, J. and Levy, K. (2020) Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Barcelona, 27-30 January 2020, 469-481. <https://doi.org/10.1145/3351095.3372828>
  - [4] Kim, P.T. (2017) Data-Driven Discrimination at Work. *William & Mary Law Review*, **58**, 857-936.
  - [5] Raji, I.D., Smart, A., White, R.N., Mitchell, M., Gebru, T., Hutchinson, B., *et al.* (2020) Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Barcelona, 27-30 January 2020, 33-44. <https://doi.org/10.1145/3351095.3372873>
  - [6] Mobley v. Workday, Inc., 740 F. Supp. 3d 796 (N.D. Cal. 2024). <https://storage.courtlistener.com/recap/gov.uscourts.cand.408645/gov.uscourts.cand.408645.80.0.pdf>
  - [7] Mobley v. Workday, Inc., No. 3:23-cv-00770-RFL, 2025 WL 1424347 (N.D. Cal. May 16, 2025). <https://storage.courtlistener.com/recap/gov.uscourts.cand.408645/gov.uscourts.cand.408645.128.0.pdf>
  - [8] Mobley v. Workday, Inc., No. 3:23-cv-00770-RFL (N.D. Cal. June 22, 2026). <https://storage.courtlistener.com/recap/gov.uscourts.cand.408645/gov.uscourts.cand.408645.360.0.pdf>
  - [9] Kellogg, K.C., Valentine, M.A. and Christin, A. (2020) Algorithms at Work: The New Contested Terrain of Control. *Academy of Management Annals*, **14**, 366-410. <https://doi.org/10.5465/annals.2018.0174>
  - [10] Fuller, J.B., Raman, M., Sage-Gavin, E. and Hines, K. (2021) Hidden Workers: Untapped Talent. Harvard Business School Project on Managing the Future of Work and Accenture.
  - [11] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A. (2021) A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, **54**, 1-35. <https://doi.org/10.1145/3457607>
  - [12] Wilson, K. and Caliskan, A. (2024) Gender, Race, and Intersectional Bias in Resume Screening via Language Model Retrieval. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, **7**, 1578-1590. <https://doi.org/10.1609/aies.v7i1.31748>
  - [13] Selbst, A.D., Boyd, D., Friedler, S.A., Venkatasubramanian, S. and Vertesi, J. (2019) Fairness and Abstraction in Sociotechnical Systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, Atlanta, 29-31 January 2019, 59-68. <https://doi.org/10.1145/3287560.3287598>
  - [14] (1964) Civil Rights Act of 1964, Title VII, 42 U.S.C. §§ 2000e to 2000e-17. <https://www.law.cornell.edu/uscode/text/42/2000e>
  - [15] Age Discrimination in Employment Act of 1967, 29 U.S.C. §§ 621-634 (1967). <https://www.law.cornell.edu/uscode/text/29/chapter-14>
  - [16] Americans with Disabilities Act of 1990, 42 U.S.C. § 12101 et seq (1990). <https://www.law.cornell.edu/uscode/text/42/chapter-126>
  - [17] Griggs v. Duke Power Co., 401 U.S. 424 (1971). <https://supreme.justia.com/cases/federal/us/401/424/>
  - [18] Smith v. City of Jackson, 544 U.S. 228 (2005). <https://supreme.justia.com/cases/federal/us/544/228/>

- [19] U.S. Equal Employment Opportunity Commission (2022) The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees. [https://data.aclum.org/storage/2025/01/EOCC\\_www\\_eeoc.gov\\_laws\\_guidance\\_americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence.pdf](https://data.aclum.org/storage/2025/01/EOCC_www_eeoc.gov_laws_guidance_americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence.pdf)
- [20] U.S. Equal Employment Opportunity Commission (2023) Assessing Adverse Impact in Software, Algorithms, and Artificial Intelligence Used in Employment Selection Procedures under Title VII of the Civil Rights Act of 1964. [https://data.aclum.org/storage/2025/01/EOCC\\_www\\_eeoc.gov\\_laws\\_guidance\\_select-issues-assessing-adverse-impact-software-algorithms-and-artificial.pdf](https://data.aclum.org/storage/2025/01/EOCC_www_eeoc.gov_laws_guidance_select-issues-assessing-adverse-impact-software-algorithms-and-artificial.pdf)
- [21] Ajunwa, I. (2020) The Paradox of Automation as Anti-Bias Intervention. *Cardozo Law Review*, **41**, 1671-1742.
- [22] National Institute of Standards and Technology (2023) Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1. U.S. Department of Commerce. <https://doi.org/10.6028/NIST.AI.100-1>