Algorithmic Human Resource Management: Bias Mitigation, Recruitment Analytics, and Performance Prediction

Mehmethan Erduran[1] & Kadir Demir[2]

[1] Res. Assist., İzmir Demokrasi University, Department of Management Information Systems, İzmir, Türkiye
erduranmehmethan@gmail.com · ORCID: 0000-0003-1620-2523

[2] Assoc. Prof. Dr., İzmir Demokrasi University, Department of Management Information Systems, İzmir, Türkiye
kadir.demir@idu.edu.tr · ORCID: 0000-0001-9568-9450

Please cite the published version:

Erduran, M., & Demir, K. (2026). Algorithmic human resource management: Bias mitigation, recruitment analytics, and performance prediction. In O. Kaynar & A. G. Yüksek (Eds.), *Artificial intelligence in business: Strategic transformation, functional applications, and future paradigms* (pp. 93–121). Livre de Lyon. https://doi.org/10.5281/zenodo.21717248

1. Introduction

Human resource management is increasingly shaped by data-driven technologies that promise to make organizational decisions faster, more consistent, and more predictive (Marler & Boudreau, 2017; Tursunbayeva et al., 2018). Applicant tracking systems, resume parsers, automated ranking tools, online assessments, performance dashboards, workforce planning models, and attrition prediction systems are now part of the ordinary vocabulary of digital HRM. This transformation goes beyond replacing paper forms with software. It changes what is visible to organizations, what is counted as evidence, who is considered employable or promotable, and how managerial judgment is supported or constrained. Following Tambe et al. (2019), algorithmic HRM therefore sits at the intersection of prediction, bias, decision-making, and governance.

The digital transformation of HRM began with operational systems that stored employee records, standardized workflows, and reduced administrative burdens. Over time, the growth of people analytics expanded the role of HR data from record keeping to inference. Organizations now use large volumes of structured and unstructured data to identify patterns in hiring, turnover, engagement, productivity, learning, collaboration, and performance. Scholars have described HR analytics and people analytics as emerging fields that combine human resource data, statistical reasoning, and organizational decision support (Marler & Boudreau, 2017; Tursunbayeva et al., 2018). In this sense, HR has moved from a largely administrative function toward a strategic decision domain in which data and models are expected to inform talent allocation, workforce planning, and organizational design.

This shift has been accelerated by remote work, platform-based labor, digital communication systems, and the routine production of digital traces (Kellogg et al., 2020; Tambe et al., 2019). Employees and applicants now generate data through applications, assessments, learning platforms, productivity tools, calendars, collaboration software, and online profiles. These traces may appear to measure behavior, effort, and potential, yet they are shaped by technologies, job design, organizational norms, and power relations (Schwartz et al., 2022). A digitally visible worker or candidate is therefore not necessarily the most capable one.

Algorithmic HRM also represents a movement from human judgment toward algorithmic decision support (Tambe et al., 2019). In recruitment, algorithms can target job advertisements, parse resumes, rank applicants, score assessments, and recommend shortlists (Fabris et al., 2025; Raghavan et al., 2020). In performance management, they can estimate future performance, attrition risk, development needs, promotion potential, or compensation-relevant indicators. Even when humans make the formal decision, the system structures what they see, which cases receive attention, and which risk categories shape judgment.

The central problem addressed in this chapter is that algorithmic HRM often appears objective while its outputs remain shaped by historical and organizational assumptions (Mehrabi et al., 2021; Schwartz et al., 2022). If past hiring favored certain universities, career paths, regions, or demographic groups, models may learn those patterns as signals of merit (Raghavan et al., 2020). If performance ratings reflect gender, age, disability, visibility, or unequal access to assignments, performance models may reproduce these biases as statistical regularities. The central concern is that algorithmic systems can make bias scalable, less visible, and harder to contest.

This chapter examines algorithmic HRM through three interconnected domains: recruitment analytics, bias mitigation, and performance prediction. It further proposes a Responsible

Algorithmic HRM Governance Model to conceptualize how fairness, explainability, accountability, and human oversight can be embedded across the algorithmic HRM decision cycle. The argument is that the success of algorithmic HRM cannot be evaluated by predictive accuracy alone (National Institute of Standards and Technology, 2023; Tambe et al., 2019). It depends on data validity, construct validity, fairness across groups, transparency to decision makers and affected individuals, meaningful human oversight, continuous auditing, and mechanisms through which candidates and employees can challenge or correct decisions. Following Selbst et al. (2019), algorithmic HRM should therefore be treated as a socio-technical system, not as a purely technical tool.

This chapter is a conceptual and narrative review and synthesis rather than an empirical study or a systematic literature review. The sources were selected from algorithmic HRM, HR analytics, personnel selection, algorithmic fairness, explainable AI, and AI governance literatures, with priority given to peer-reviewed studies, official legal and regulatory texts, and widely used technical governance frameworks. The purpose is to integrate these literatures into a governance-oriented framework for understanding how prediction, bias, human judgment, and accountability interact in algorithmic HRM, rather than to produce new empirical findings.

2. Conceptual Foundations of Algorithmic HRM

Algorithmic HRM refers to the use of data-driven and algorithmic systems to support, structure, or automate human resource decisions such as recruitment, selection, performance evaluation, promotion, training, workforce planning, compensation, and retention (Leicht-Deobald et al., 2019; Tambe et al., 2019). The term includes simple scoring rules, statistical models, machine learning systems, natural language processing tools, recommender systems, and increasingly generative AI applications. Drawing on the HR analytics logic described by Marler and Boudreau (2017), their common feature is the conversion of HR data into decision-relevant outputs such as predictions, rankings, classifications, alerts, or recommendations.

A useful distinction can be made between decision support and automated decision-making. Decision-support systems provide information to human decision makers; automated systems execute or strongly determine decisions through rules or model outputs (European Parliament and Council, 2024; National Institute of Standards and Technology, 2023). In practice, the boundary is blurred. If a recruiter sees only top-ranked candidates, or a manager sees employees through risk categories, the algorithm has already shaped the decision space.

Algorithmic HRM differs from traditional HRM because it introduces a predictive and computational layer into decisions that were historically based on interviews, supervisor judgment, performance appraisals, job analysis, assessment centers, and professional HR expertise (Schmidt & Hunter, 1998). Traditional HRM certainly has its own biases and inconsistencies. Human interviewers can be influenced by stereotypes, similarity attraction, halo effects, fatigue, and organizational politics (Neumann et al., 2022). Algorithmic systems may reduce some forms of inconsistency by standardizing evaluation criteria and applying the same scoring logic across candidates. Yet standardization does not guarantee fairness. Raghavan et al. (2020) caution that if the standardized criteria are invalid or historically biased, the system can reproduce inequality with greater efficiency.

Table 1. Conceptual distinctions among e-HRM, HR analytics, people analytics, and algorithmic HRM

Concept	Primary Focus	Role in Decision Process	Illustrative HR Use
e-HRM	Digitalization of HR transactions and records	Operational support	Online applications, employee self-service, digital leave requests
HR Analytics	Analysis of HR data and metrics	Reporting and decision support	Turnover dashboards, time-to-fill metrics, engagement analysis
People Analytics	Modeling employee behavior and organizational outcomes	Strategic analysis	Workforce planning, network analysis, learning and retention models
Algorithmic HRM	Algorithmic structuring or automation of HR decisions	High decision impact	Candidate ranking, job-fit scores, performance prediction, attrition risk scoring

The relationship among e-HRM, HR analytics, people analytics, and algorithmic HRM is important. e-HRM primarily concerns the digitalization of HR processes, such as online applications, electronic employee records, self-service portals, and workflow automation. HR analytics involves the analysis of HR data to support reporting, diagnosis, and decision-making (Marler & Boudreau, 2017). People analytics broadens the analytical lens to employee behavior, organizational outcomes, and strategic workforce questions (Tursunbayeva et al., 2018). Algorithmic HRM goes further by embedding algorithmic outputs directly into high-impact HR decisions (Tambe et al., 2019). In this sense, algorithmic HRM is not simply analytics about people; it is analytics acting on people through organizational decision processes.

2.1. Algorithmic Management and Organizational Control

Algorithmic HRM should also be understood in relation to algorithmic management. Algorithmic management refers to organizational control mechanisms that use algorithms to direct, evaluate, discipline, reward, or replace worker judgment. Kellogg et al. (2020) describe algorithmic control through mechanisms such as recording, rating, recommending, restricting, rewarding, and replacing. These mechanisms are relevant beyond platform work. In conventional organizations, HR algorithms may record employee behavior, rate performance, recommend actions, restrict access to opportunities, reward predicted high performers, or identify workers who are considered retention or performance risks.

This control dimension matters because HR algorithms help constitute organizational reality (Kellogg et al., 2020). A performance score can affect trust, assignments, or promotion opportunities; an attrition score can change how committed an employee is perceived to be; and a job-fit score can prevent a candidate from being reviewed. Leicht-Deobald et al. (2019) show how algorithmic outputs can affect personal integrity, while Schwartz et al. (2022) help explain why datafication can narrow contextual judgment and intensify surveillance.

Algorithmic HRM is best conceptualized as a socio-technical system. A socio-technical view rejects the idea that fairness can be solved only inside the model. The model is one part of a larger system that includes data collection, problem definition, construct measurement, organizational incentives, legal obligations, human oversight, vendor relationships, employee voice, and feedback loops. Selbst et al. (2019) warn that fairness efforts can fail when technical systems are abstracted

away from the social contexts in which they operate. In HRM, this warning is especially important because the constructs being modeled, such as talent, fit, potential, performance, culture, and leadership, are not simple natural properties. They are organizational judgments shaped by power, norms, job design, and opportunity structures.

Tambe et al. (2019) argue that AI in HRM faces distinctive challenges because HR phenomena are complex, data sets may be small, legal and ethical accountability is high, and employees may react negatively to algorithmic management decisions. These constraints mean that algorithmic HRM cannot simply import methods from consumer recommendation or advertising. Employment decisions affect livelihoods, career trajectories, dignity, and access to organizational resources. For this reason, algorithmic HRM requires stronger standards of validity, fairness, explanation, accountability, and governance than many lower-stakes uses of predictive analytics.

The literature on AI-enabled recruitment also shows why algorithmic HRM should not be framed as a simple contest between biased humans and neutral machines. Systematic and multidisciplinary reviews emphasize that AI recruitment tools create ethical opportunities, risks, and ambiguities across the hiring funnel, including questions of fairness, accountability, privacy, autonomy, and candidate experience (Fabris et al., 2025; Hunkenschroer & Luetge, 2022; Köchling & Wehner, 2020; Rigotti & Fosch-Villaronga, 2024). Raghavan et al. (2020) therefore caution that vendor claims about debiasing should be evaluated against evidence about validation, target construction, data provenance, and legal compatibility rather than accepted as a default benefit of automation.

3. Recruitment Analytics: Data-Driven Hiring and Selection

Recruitment analytics refers to the use of data and analytical techniques to improve, structure, or automate recruitment and selection processes (Fabris et al., 2025; Rigotti & Fosch-Villaronga, 2024). It can include analyzing candidate pools, optimizing job advertisements, predicting candidate-job fit, screening resumes, ranking applicants, monitoring candidate experience, forecasting offer acceptance, and evaluating hiring outcomes (Raghavan et al., 2020). Recruitment analytics is attractive because hiring often involves large applicant pools, costly screening, time pressure, and uncertainty about future job performance. Algorithms appear to offer a way to manage scale while reducing inconsistency.

Data sources in recruitment analytics are diverse. Common sources include resumes, application forms, education history, work experience, skills tests, cognitive assessments, personality inventories, online work simulations, structured interview responses, video interviews, professional network data, referral information, and prior employee performance records (Fabris et al., 2025; Raghavan et al., 2020). Some tools use NLP to extract skills from resumes; others use tests, games, or historical employee data to infer candidate-job fit. These sources may support matching, but their richness does not automatically make decisions fairer. Hunkenschroer and Luetge (2022) and Raghavan et al. (2020) show that more data can also create more opportunities for proxy discrimination, irrelevant correlations, privacy intrusion, and construct confusion.

Algorithmic recruitment can enter the hiring process at multiple stages (Fabris et al., 2025; Köchling & Wehner, 2020). Job advertisement targeting determines who sees the vacancy in the first place. Candidate sourcing tools identify potential applicants from databases or professional networks. Resume screening systems parse and filter applications. Automated ranking tools order

candidates according to predicted fit or quality. Assessment systems score tests or simulations. Interview analytics systems may analyze text, speech, or video responses. Shortlisting systems recommend candidates for recruiter review. Final decision tools may integrate assessment scores, interview ratings, and predicted performance (Raghavan et al., 2020). Fabris et al. (2025) similarly show that bias can enter at each stage, and small biases can accumulate across the recruitment funnel.

The potential benefits of recruitment analytics should be taken seriously. Automated tools can reduce administrative burden, shorten time-to-hire, enable recruiters to manage large applicant pools, standardize initial screening, and identify candidates who might otherwise be overlooked. Structured assessments and consistent scoring can reduce some forms of human subjectivity. When validated carefully, predictive models may improve the connection between selection criteria and job-relevant outcomes. The personnel selection literature has long shown that structured, evidence-based methods can outperform unstructured judgment in predicting job performance (Schmidt & Hunter, 1998). Algorithmic approaches can be understood as an extension of this broader movement toward evidence-based selection, provided that the evidence is valid and the system is governed responsibly.

However, recruitment analytics also introduces significant risks. A model trained on past hiring decisions may learn the preferences of previous recruiters rather than the requirements of the job (Raghavan et al., 2020). A model trained on current high performers may reproduce historical opportunity patterns that allowed certain groups to become visible as successful. Variables such as university name, postal code, employment gaps, extracurricular activities, social network size, language style, commute distance, or patterns of online availability may act as proxies for socioeconomic status, gender, disability, caregiving responsibilities, race, ethnicity, or class advantage (Rigotti & Fosch-Villaronga, 2024). Raghavan et al. (2020) and the U.S. Equal Employment Opportunity Commission (2023) therefore show why removing legally protected attributes from the data is insufficient when many non-sensitive variables remain correlated with protected or disadvantaged group membership.

Video interview analytics and emotion recognition raise particularly serious concerns. These tools may interpret facial expressions, voice features, word choice, or eye contact as evidence of personality, confidence, honesty, or cultural fit, even though such signals can be shaped by disability, neurodiversity, accent, language background, cultural norms, internet quality, interview anxiety, lighting, and camera access. The EU AI Act classifies many employment-related AI systems as high-risk (European Parliament and Council, 2024). It also treats AI systems that infer emotions in workplaces and educational institutions as prohibited practices, except for narrowly specified medical or safety reasons (European Parliament and Council, 2024); as of May 26, 2026, the European Commission's implementation materials list workplace and education emotion recognition among the AI Act prohibitions (European Commission, n.d.).

Recruitment analytics must therefore be evaluated through a double lens. The first lens is predictive validity: Does the system actually predict job-relevant performance, learning, retention, or other legitimate outcomes? Selection procedures should be job-related and valid for their intended employment use (Schmidt & Hunter, 1998; U.S. Equal Employment Opportunity Commission, 2007). The second lens is fairness and governance: Does the system distribute opportunities equitably, explain its outputs, allow meaningful human review, and remain accountable to legal and ethical standards (U.S. Equal Employment Opportunity Commission,

2023)? A system that predicts historical hiring success but reproduces exclusion is not responsible recruitment analytics. Conversely, a system that satisfies a narrow fairness metric but predicts an invalid outcome is also not responsible. The challenge is to align predictive usefulness with organizational justice.

4. Bias in Algorithmic HRM

Bias in algorithmic HRM can arise at multiple points in the machine learning life cycle and the organizational decision cycle. It is tempting to think of bias as a property of a model after it has been trained. In practice, bias can be introduced through problem definition, data collection, labeling, measurement, representation, feature selection, model optimization, deployment, user interpretation, and feedback effects. Suresh and Guttag (2021) provide a useful framework for understanding how sources of harm can emerge across the machine learning life cycle, while Mehrabi et al. (2021) synthesize many technical categories of bias and fairness in machine learning. In HRM, these technical categories must be connected to organizational histories and employment law.

Historical bias occurs when data reflect past inequalities (Mehrabi et al., 2021; Suresh & Guttag, 2021). If prior promotions favored men, leadership models may treat male-coded career patterns as potential. If recruitment favored elite universities, models may treat institutional prestige as capability. If older, disabled, or caregiving candidates were previously screened out, models may learn to discount similar profiles. Schwartz et al. (2022) help explain why such data can be descriptively accurate while normatively unjust.

Measurement bias arises when variables do not validly measure the construct of interest (Suresh & Guttag, 2021). Performance is a central example because it is multidimensional rather than reducible to one score (Campbell & Wiernik, 2015). Supervisor ratings, KPIs, sales records, absenteeism, productivity metrics, training completion, or digital activity each capture only part of performance and may reflect visibility, territory quality, managerial perception, or platform use. Schwartz et al. (2022) therefore help show why these measures can become distorted targets when treated as objective labels.

Representation bias occurs when some groups are underrepresented or differently represented in the data (Mehrabi et al., 2021; Suresh & Guttag, 2021). If a model is trained mostly on majority-group employees, it may perform worse for minority groups because it has learned fewer patterns relevant to their career paths. Underrepresentation can also affect intersectional groups, such as older women, disabled women, migrant workers, or candidates from less common educational backgrounds. Even when average model performance appears acceptable, subgroup performance may vary. A responsible HR analytics process therefore requires group-based validation, not only aggregate accuracy (Barocas et al., 2023). A model with high overall accuracy can still produce systematic errors for smaller or marginalized groups.

Proxy discrimination is one of the most important issues in algorithmic HRM (Barocas et al., 2023; Mehrabi et al., 2021). Sensitive variables may be removed, yet university name, postal code, career gaps, social network size, language style, commute distance, or constant online availability can still encode socioeconomic status, gender, disability, caregiving, race, ethnicity, or class advantage. Raghavan et al. (2020) show why removing protected attributes does not remove discrimination when correlated variables remain.

Automation bias is a decision-layer risk rather than only a data-layer risk. It occurs when human decision makers over-rely on algorithmic outputs because they appear objective, technical, or authoritative (Parasuraman & Manzey, 2010). A recruiter who would have considered a candidate may defer to a low job-fit score. A manager may accept a performance risk flag without asking whether the input data are valid. A promotion committee may treat a model ranking as a neutral ordering of talent. Automation bias is especially problematic when the model output is presented without uncertainty, explanation, limitations, or alternative interpretations (Neumann et al., 2022). Leicht-Deobald et al. (2019) similarly show that algorithmic systems may not formally make decisions, but their scores can strongly guide human attention and judgment.

Bias can also be amplified through feedback loops (Barocas et al., 2023; Suresh & Guttag, 2021). If an algorithm ranks certain candidates lower, those candidates receive fewer interviews and fewer opportunities to generate positive outcome data. If employees flagged as high potential receive better assignments, mentoring, and training, they may later perform better because the organization invested in them. The model may then appear to have predicted potential, when it partially created the conditions for the predicted outcome. Similarly, employees labeled as attrition risks may receive less investment or trust, increasing disengagement. Selbst et al. (2019) help explain why such feedback loops require algorithmic HRM to be governed as an ongoing system rather than a one-time model deployment.

5. Bias Mitigation Strategies

Bias mitigation in algorithmic HRM includes technical, organizational, legal, and ethical strategies (Barocas et al., 2023; Bellamy et al., 2019). Technical mitigation matters, but bias also exists outside the model. It can be embedded in the definition of success, the choice of target variable, the composition of the data set, the interpretation of model outputs, and the organizational consequences attached to those outputs (Schwartz et al., 2022). Responsible mitigation must therefore begin before model training and continue after deployment.

Pre-processing strategies intervene before model training. They include data cleaning, missing-data analysis, representation analysis, reweighting, feature review, detection of sensitive and proxy variables, and documentation of data provenance (Bellamy et al., 2019; Mehrabi et al., 2021). In HRM, pre-processing should also include construct validation: does a resume keyword measure skill, a rating measure contribution, or a response-speed measure motivation? Following Gebru et al. (2021), such questions determine whether the model has a legitimate foundation.

Data documentation is a critical pre-processing practice (Gebru et al., 2021). Dataset documentation should specify where the data came from, why each feature is used, what population is represented, what groups may be underrepresented, what labels are used, how missing values are treated, and what known limitations exist. In HRM, documentation should also record whether data were collected for the purpose for which they are now being used (National Institute of Standards and Technology, 2023). Data collected for payroll, scheduling, or platform access may not be valid for performance prediction. Repurposing such data without validation can create hidden measurement problems and privacy concerns.

In-processing strategies intervene during model training. These methods include fairness-aware learning, regularization, adversarial debiasing, group-sensitive optimization, and modeling approaches that include fairness constraints (Bellamy et al., 2019; Mehrabi et al., 2021).

The goal is to optimize not only predictive accuracy but also fairness-related criteria. However, fairness-aware learning requires choices about which fairness definition matters. Equal selection rates, equal true positive rates, equal false positive rates, equal calibration, and equal predictive value are not always simultaneously achievable (Barocas et al., 2023). The appropriate choice depends on the organizational context, the harm being addressed, the decision threshold, and the legal environment. Technical fairness must therefore be connected to normative reasoning and stakeholder deliberation.

Post-processing strategies intervene after the model produces scores or classifications. These include threshold adjustment, score calibration, group-based error analysis, review triggers, confidence intervals, and escalation to human review (Barocas et al., 2023; Bellamy et al., 2019). In recruitment, a system may route borderline cases or cases with missing data to human assessment rather than automatic rejection. In performance prediction, a model may flag uncertainty and require contextual review before any employment consequence follows. Post-processing is useful because organizations often deploy vendor systems that cannot be fully modified internally. Yet post-processing should not become a way to hide deeper problems in data or model design.

Fairness metrics are central tools for detecting and evaluating bias (Barocas et al., 2023). Demographic parity asks whether selection rates are similar; equal opportunity asks whether qualified individuals are selected at similar rates; equalized odds compare false positive and false negative rates; predictive parity asks whether a score has similar meaning across groups; and calibration asks whether predicted probabilities match actual outcomes. These metrics make fairness observable, but Barocas et al. (2023) show that formal criteria must still be interpreted within legal, institutional, and social understandings of discrimination.

Algorithmic fairness tools can support these analyses. IBM AI Fairness 360, for example, provides metrics and mitigation algorithms for assessing and reducing bias in machine learning workflows (Bellamy et al., 2019). Such tools can be valuable for HR data scientists and auditors, but they should not be treated as compliance guarantees. A fairness toolkit can measure group disparities in a data set, but it cannot decide whether the target variable is legitimate, whether the job analysis is valid, whether candidates received appropriate notice, whether disabled applicants were accommodated, whether workers can contest a decision, or whether the organizational use of the model is ethically justified.

Beyond technical debiasing, organizations must examine the full decision process. They must clarify the purpose of the model, validate the relationship between predictors and job requirements, document the reasons for feature inclusion, test subgroup performance, provide explanations to decision makers, train users to question model outputs, create appeal mechanisms, and monitor outcomes over time. The EEOC emphasizes that employment tests and selection procedures must be job-related and consistent with business necessity, and that employers remain responsible for selection tools even when vendors provide them (U.S. Equal Employment Opportunity Commission, 2007, 2023). Bias mitigation is therefore a governance practice, not simply a model adjustment practice. Operationally, responsibility should be assigned rather than shared vaguely: HR leaders define acceptable uses, data scientists validate models and subgroup performance, legal teams review compliance, procurement units set vendor requirements, vendors document limitations, and line managers are trained to use and challenge outputs.

6. Performance Prediction in HRM

Performance prediction is one of the most attractive and challenging domains of algorithmic HRM (Marler & Boudreau, 2017; Tambe et al., 2019). Organizations want to know which applicants will perform well, which employees have promotion potential, who may need development, which teams are at risk, and who may leave the organization. Predictive analytics appears to offer a way to move from retrospective evaluation to proactive talent management. Yet employee performance is not a simple property that can be captured by a single score. Campbell and Wiernik (2015) therefore support treating performance as contextual, relational, dynamic, and partly constructed by the organization itself.

Employee performance can include task performance, contextual performance, adaptive performance, citizenship behavior, teamwork, leadership potential, learning agility, creativity, customer impact, safety behavior, ethical behavior, and contribution to organizational culture (Campbell & Wiernik, 2015). Some roles have clear quantitative outputs, while others depend on coordination, mentoring, problem solving, judgment, or invisible relational labor. The same employee may perform differently depending on team resources, managerial support, workload, autonomy, technology, organizational climate, and market conditions. A model that treats performance as a stable individual trait may miss these contextual factors.

Data sources for performance prediction include performance appraisal scores, key performance indicators, sales records, attendance data, training completion, engagement surveys, 360-degree feedback, project management data, customer ratings, supervisor notes, communication patterns, and digital work traces (Campbell & Wiernik, 2015; Marler & Boudreau, 2017). These sources vary in validity and sensitivity: some are direct but narrow, some are broad but subjective, and some are behaviorally rich but intrusive. Leicht-Deobald et al. (2019) and Schwartz et al. (2022) help explain why learning analytics and collaboration traces may support development while also creating surveillance risks.

Predictive models in performance analytics range from interpretable statistical models to complex machine learning systems (Marler & Boudreau, 2017; Tambe et al., 2019). Logistic regression, decision trees, random forests, gradient boosting, support vector machines, neural networks, survival analysis, natural language processing, and network-based models can all be used to classify employees, estimate probabilities, or detect patterns. The technical diversity of models should not distract from the central HR question: what is being predicted, why, with what data, and for what decision?

The personnel selection literature provides a useful caution. Schmidt and Hunter (1998) showed that some selection methods have stronger predictive validity than others, and that validity matters for utility. Yet more complex models are not automatically better. Harman and Scheuerman (2023) found that a simple rule-based model in a personnel selection competition outperformed most submitted machine learning models while retaining transparency. This finding does not imply that machine learning is never useful. Rather, it challenges the assumption that complexity should be preferred by default. In high-impact HR decisions, simpler and more explainable models may be preferable when they perform comparably and are easier to audit, explain, and govern.

Explainable AI is especially important in performance prediction. Feature importance, SHAP values, LIME explanations, partial dependence plots, counterfactual explanations, and model cards can help decision makers understand a score or classification (Lundberg & Lee, 2017; Mitchell et al., 2019; Ribeiro et al., 2016). In HRM, explanation is tied to legitimacy, trust, contestability, and dignity: managers should understand recommendations before acting, employees should be able to

identify inaccurate data, and compliance teams should audit whether the model is used for its intended purpose.

The risks of performance prediction are substantial. Employees may be reduced to narrow metrics, monitored in ways that reduce trust and autonomy (Kellogg et al., 2020; Leicht-Deobald et al., 2019), labeled in ways that become self-fulfilling, or denied opportunities when models favor profiles resembling past high performers. Predictive systems may also obscure team-based contributions and weaken managerial responsibility. Schwartz et al. (2022) further show why employees may perceive algorithmic evaluation as unfair if they cannot understand or challenge it.

Performance prediction should distinguish developmental use from high-stakes evaluative use. Developmental uses include identifying training needs, suggesting coaching, or flagging workload risks, where the appropriate response is support rather than sanction. High-stakes evaluative uses include promotion eligibility, compensation, performance ranking, disciplinary action, or termination. Because employment procedures must be job-related and consistent with business necessity (U.S. Equal Employment Opportunity Commission, 2007), evaluative uses require stronger evidence of criterion validity, subgroup validity, explainability, and human review than developmental uses. Campbell and Wiernik (2015) support this distinction by treating performance as multidimensional and context bound.

Validation in performance prediction should be treated as a continuing requirement rather than a one-time model test. Criterion validity asks whether the predicted outcome is a defensible measure of performance; subgroup validation asks whether errors differ across gender, age, disability, race, ethnicity, or other relevant groups; and drift monitoring asks whether the model remains valid as jobs, teams, technologies, and work arrangements change. False positives and false negatives also carry different employment harms. A false positive in promotion potential may allocate scarce development resources to the wrong employee, while a false negative may deny a qualified employee advancement. A false positive performance-risk flag may trigger unnecessary monitoring, while a false negative may delay needed support.

Three scenarios illustrate the distinction. A training-need model may predict which employees would benefit from reskilling and route them to optional support. A promotion-potential model may influence access to leadership tracks, so it requires stronger justification, explanation, and human review. An attrition or performance-risk score may support retention conversations, but it should not become a hidden label that reduces trust, opportunity, or investment. In each case, the same predictive technique can be developmental or punitive depending on the organizational action attached to the score.

Research on personnel selection also complicates the human-versus-algorithm narrative. Mechanical or algorithmic combination of information can improve prediction compared with holistic judgment, but decision makers may resist algorithms or use them selectively because of stakeholder perceptions and concerns about autonomy (Neumann et al., 2022). At the same time, algorithm-based HR decision-making can challenge personal integrity and organizational sense-making if employees and managers are expected to defer to outputs that they cannot interpret or contest (Leicht-Deobald et al., 2019). These findings reinforce the need for governance structures that combine predictive discipline with human accountability.

7. Governance and Responsible Algorithmic HRM

Governance is necessary because algorithmic HRM operates in high-impact domains. Hiring, promotion, evaluation, compensation, development, and termination decisions affect access to income, identity, status, opportunity, and dignity. Employment-related AI systems are therefore treated as high-risk in emerging legal frameworks, including the EU AI Act (European Parliament and Council, 2024). Model accuracy alone is insufficient because a system can be statistically accurate on average and still unfair, invalid, opaque, or harmful for particular groups (National Institute of Standards and Technology, 2023). Conversely, a system can satisfy a narrow fairness metric while being based on an invalid definition of performance or fit. The U.S. Equal Employment Opportunity Commission (2023) reinforces this point by framing governance as the organizational structure through which selection-related risks are identified, assigned, monitored, and corrected.

The NIST AI Risk Management Framework provides a structure for mapping, measuring, managing, and governing AI risks (National Institute of Standards and Technology, 2023), while NIST SP 1270 highlights systemic, human, and computational sources of bias across the AI lifecycle (Schwartz et al., 2022). In HRM, bias may come from historical employment patterns, rating practices, organizational incentives, vendor design, data limitations, interfaces, or managerial interpretation. Erduran (2025) similarly frames bias, opacity, privacy, autonomy, and accountability in AI-driven MIS as governance concerns that require proactive design rather than passive compliance. This is directly relevant because HR systems are management information systems that collect organizational data and shape high-stakes managerial action.

Human oversight is a core governance requirement, but not all human oversight is meaningful (National Institute of Standards and Technology, 2023). A nominal human-in-the-loop process may exist only to rubber-stamp algorithmic outputs. Meaningful human oversight requires that HR professionals and managers understand the system's purpose, limitations, uncertainty, and appropriate use. They must have authority to challenge or override the output, access to relevant explanations, and responsibility to document their reasoning. Oversight also requires training. Leicht-Deobald et al. (2019) and Parasuraman and Manzey (2010) help explain why a recruiter cannot meaningfully evaluate a model ranking if the organization treats the score as unquestionable or if the user interface hides the factors behind it.

Algorithmic delegation should not become accountability delegation. Even when a vendor builds the tool, the employer remains responsible for how it is used in employment decisions. The EEOC's guidance on employment tests and AI-related selection procedures underscores that employers should ensure selection tools are validated, job-related, and monitored for adverse impact (U.S. Equal Employment Opportunity Commission, 2007, 2023). The Americans with Disabilities Act (ADA) also raises concerns when algorithmic tools screen out applicants or employees with disabilities or fail to provide reasonable accommodation (U.S. Equal Employment Opportunity Commission, 2022). Accountability therefore requires role clarity: data scientists, HR leaders, legal teams, procurement specialists, managers, vendors, and audit committees must know who is responsible for validation, monitoring, explanation, documentation, and corrective action.

Transparency, explainability, and contestability are related but distinct. Transparency concerns whether the organization discloses that an algorithmic system is being used, what purpose it serves, and what data it relies on (European Parliament and Council, 2024; National Institute of Standards and Technology, 2023). Explainability concerns whether affected individuals and decision makers can understand why a particular output was produced. Contestability concerns whether candidates or employees can challenge the decision, correct inaccurate data, request

accommodation, or obtain human review (U.S. Equal Employment Opportunity Commission, 2022). A system may be transparent in the general sense that its use is disclosed but still not explainable or contestable in practice. Responsible algorithmic HRM requires all three.

Legal developments increasingly reflect these governance concerns. New York City's Local Law 144 requires certain automated employment decision tools used in hiring or promotion to undergo bias audits and requires public disclosure and notice to candidates or employees (City of New York Department of Consumer and Worker Protection, n.d.). The EU AI Act classifies many employment, worker management, and self-employment access systems as high-risk, including systems used to filter applications, evaluate candidates, allocate tasks, monitor behavior, or evaluate performance (European Parliament and Council, 2024). As of May 26, 2026, the European Commission's implementation materials indicated that high-risk rules for areas such as employment would apply from December 2, 2027, and the Commission was also consulting on draft guidelines for transparency obligations and high-risk AI classification (European Commission, n.d., 2026a, 2026b). These details should be checked again before publication because AI Act guidance and compliance timelines continue to evolve.

Responsible algorithmic HRM also requires lifecycle monitoring (National Institute of Standards and Technology, 2023). Models can drift as jobs, labor markets, data pipelines, policies, and applicant behavior change. Because bias can arise across the AI lifecycle, Schwartz et al. (2022) and Tambe et al. (2019) support periodic validation, subgroup performance analysis, fairness audits, user feedback, incident reporting, and sunset criteria for models that no longer meet validity or fairness standards.

8. A Responsible Algorithmic HRM Governance Model

This chapter proposes a Responsible Algorithmic HRM Governance Model as a conceptual synthesis of the preceding discussion. The model is built on a simple claim: bias mitigation in algorithmic HRM is a continuous decision cycle that spans data collection, model development, human decision-making, governance mechanisms, and organizational outcomes (National Institute of Standards and Technology, 2023; Schwartz et al., 2022). Selbst et al. (2019) emphasize that fairness problems cannot be solved through abstraction from social context, and Tambe et al. (2019) show why HRM is a distinctive domain for AI. Building on these insights, the model organizes responsible algorithmic HRM into five interrelated layers: Data, Algorithmic Model, Decision, Governance, and Outcome.

The model contributes in four ways. First, it translates general AI risk governance into HR-specific decision points: candidate and employee data, predictive targets, human review, appeal, and outcomes. Second, it functions as both conceptual synthesis and practical analysis tool: researchers can locate where bias enters, while practitioners can use it as a governance checklist. Third, it makes the decision layer explicit by showing how scores become managerial action through recruiters, supervisors, HR managers, and committees. Fourth, it applies most directly to recruitment, performance management, promotion, retention, and training allocation; it can be extended to pay and dismissal decisions when organizations add domain-specific legal review, validation standards, and safeguards for high-stakes employment consequences.

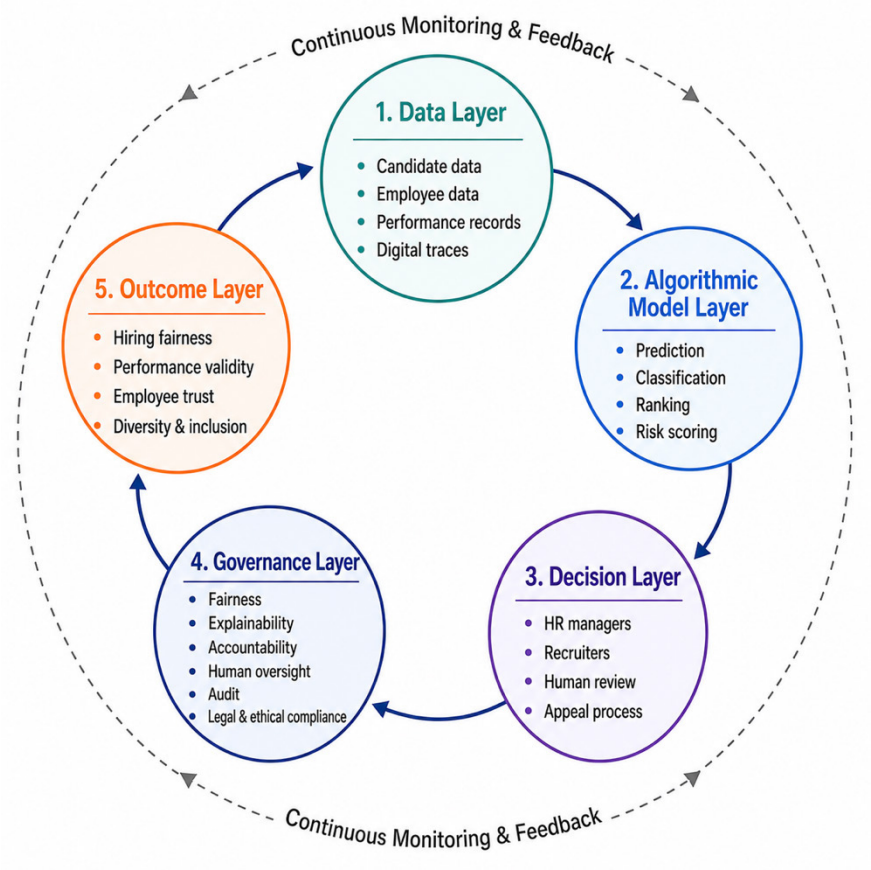

Figure 1. Responsible Algorithmic HRM Governance Model

The Data Layer asks which human attribute is represented by which data. HR data are traces of prior decisions, measurement systems, job designs, and opportunity structures (Schwartz et al., 2022; Suresh & Guttag, 2021). Candidate data may reflect unequal access to education, networks, credentials, or resume coaching, while employee data may reflect unequal access to visibility, mentoring, assignments, or supportive managers. Performance records may mix contribution with managerial perception, and digital traces may measure platform visibility more than meaningful work. The main risks are historical bias, missing data, representation bias, proxy discrimination, and weak construct validity. Following Gebru et al. (2021), responsible practices include data audits, representation analysis, data-source documentation, variable justification, proxy review, and validation of data use for the intended HR decision.

The Algorithmic Model Layer asks what the model optimizes and how its errors affect different groups. It covers target definition, feature engineering, model selection, validation, explanation, and fairness evaluation. Key risks include invalid targets, overfitting, opacity, subgroup error differences, and overreliance on aggregate accuracy (Barocas et al., 2023; Bellamy et al., 2019). Responsible practices include model comparison, fairness metrics, subgroup validation, explainable AI, model cards, and explicit statements of model limitations, as recommended in model reporting work by Mitchell et al. (2019). The layer matters because a technically strong model can still optimize a poorly chosen HR target.

The Decision Layer asks how an algorithmic score becomes organizational action. Scores may operate as advice, rankings, thresholds, review triggers, or near-final decisions. Neumann et al. (2022) and Parasuraman and Manzey (2010) help explain the risks of automation bias, unclear

responsibility, and uncritical reliance. Responsible practices include meaningful human oversight, documented reasoning, alternative assessment channels, accommodation procedures, decision logs, and contestability mechanisms. The U.S. Equal Employment Opportunity Commission (2023) similarly emphasizes that AI-based employment selection procedures require monitoring and accountability.

The Governance Layer asks how the system is audited, documented, and held accountable over time. It includes policies, legal review, procurement standards, vendor accountability, audit trails, risk assessment, incident response, communication, and monitoring (National Institute of Standards and Technology, 2023). Responsible practices include algorithmic audits, model and dataset documentation, fairness reports, legal and ethical review, and role-based accountability. The New York AEDT rules and EU AI Act show that such practices are becoming formal employment-governance expectations (City of New York Department of Consumer and Worker Protection, n.d.; European Parliament and Council, 2024), especially for systems used in hiring, promotion, monitoring, and worker management.

The Outcome Layer asks what effects the system produces. Outcomes include hiring fairness, performance validity, diversity and inclusion, employee trust, perceived justice, privacy perceptions, legitimacy, dignity, and contestation. Organizations should ask whether disadvantaged groups are screened out, whether development resources concentrate among already advantaged employees, and whether monitoring reduces trust (Kellogg et al., 2020; Leicht-Deobald et al., 2019). This layer also requires attention to second-order consequences: a system that changes hiring, promotion, or evaluation patterns may gradually reshape the organization's diversity profile, career norms, and performance culture. Rigotti and Fosch-Villaronga (2024) support treating outcomes as feedback for revising data, models, decisions, and governance.

Table 2. Components of the Responsible Algorithmic HRM Governance Model

Model Layer	Core Question	Key Risk	Responsible Practice
Data Layer	Which human attribute is represented by which data?	Historical bias, missing representation, proxy discrimination, poor construct validity	Data audit, representation analysis, variable justification, proxy review
Algorithmic Model Layer	What does the model optimize?	Invalid target, subgroup error differences, black-box decisions, overfitting	Fairness metrics, explainability, subgroup validation, model documentation
Decision Layer	How does the score become a human decision?	Automation bias, unclear responsibility, lack of appeal	Meaningful human oversight, decision reasoning, contestability, review logs
Governance Layer	How is the system monitored and accountable?	Opaque use, legal and ethical risk, model drift, weak vendor oversight	Algorithmic audit, model cards, legal review, role-based accountability
Outcome Layer	What effects does the system produce?	Trust loss, reduced diversity, invalid performance decisions, dignity harms	Continuous monitoring, employee feedback, impact assessment, corrective action

The model's central contribution is to connect predictive analytics with governance. It integrates recruitment analytics, performance prediction, and bias mitigation by showing that fairness must be designed, measured, interpreted, governed, and revised across the decision cycle

(Barocas et al., 2023; Selbst et al., 2019). This framing moves the chapter away from a narrow toolkit view of debiasing and toward an institutional view of responsible decision infrastructure. Consistent with the NIST AI Risk Management Framework (National Institute of Standards and Technology, 2023), responsible algorithmic HRM should be understood as continuous monitoring rather than a one-time certification process.

In practical terms, the model asks organizations to examine what data represent, what models optimize, how errors are distributed, who can question outputs, who is accountable, and what outcomes are produced. Schwartz et al. (2022) help frame these questions as a way to shift algorithmic HRM from a narrow efficiency project to a responsible decision infrastructure.

A brief application example illustrates how the model can be used. For a candidate-ranking tool, the Data Layer would examine whether CV data, assessment tests, and historical hiring records validly represent job-relevant capability and whether some applicant groups are underrepresented. The Model Layer would review the target variable, compare model alternatives, and test subgroup error rates. The Decision Layer would define how ranking thresholds shape shortlisting, when recruiters must conduct human review, and how accommodation or appeal requests are handled. The Governance Layer would assign audit duties, vendor accountability, documentation, and corrective-action triggers. The Outcome Layer would monitor selection rates, candidate trust, diversity effects, and whether the tool improves hiring quality without excluding qualified applicants.

9. Future Research Directions

Future research on algorithmic HRM should address changing technologies, governance demands, and work contexts. One direction is generative AI in HRM: job-ad drafting, resume summarization, candidate messaging, interview questions, and performance-review narratives. These uses may improve efficiency but also introduce hallucination, privacy, hidden bias, and overreliance risks. They may also standardize HR language in ways that appear neutral while concealing weak evidence or reproducing generic managerial assumptions. Autio et al. (2024) link generative AI risks to lifecycle evaluation and monitoring in the NIST Generative AI Profile, a point that is directly relevant to HR communication and decision support.

A second direction concerns multimodal recruitment systems that combine text, video, audio, behavioral data, assessment responses, and digital interaction patterns (Fabris et al., 2025). Such systems may measure irrelevant or discriminatory attributes such as voice, accent, facial expression, camera quality, disability, neurodiversity, and cultural communication norms (Hunkenschroer & Luetge, 2022; Rigotti & Fosch-Villaronga, 2024). Future research should test incremental validity, necessity, and accommodation pathways. The European Parliament and Council (2024) make this evidentiary burden especially important by treating many employment AI systems as high-risk under the EU AI Act.

A third direction is algorithmic fairness beyond metrics. Fairness metrics are necessary but incomplete (Barocas et al., 2023). They can reveal disparities, but they cannot decide which disparities matter, what fairness means in context, or how trade-offs should be resolved. Building on socio-technical critiques of algorithmic fairness (Selbst et al., 2019) and legal-ethical analyses of AI recruitment (Rigotti & Fosch-Villaronga, 2024), future research should connect statistical fairness with organizational justice, labor law, disability accommodation, employee voice, and critical HRM.

It should also examine how candidates and employees perceive explanations, appeals, and algorithmic assessments over time.

A fourth direction is worker voice and participatory governance. Many algorithmic HRM systems are designed, purchased, and deployed without meaningful input from affected people. Participatory governance would involve employees, candidates, worker representatives, diversity officers, legal experts, and ethics committees in defining acceptable uses, identifying harms, reviewing audits, and designing appeals (National Institute of Standards and Technology, 2023). Kellogg et al. (2020) help explain why worker voice matters for understanding how work is actually performed, and Leicht-Deobald et al. (2019) show why personal integrity concerns should be part of participatory review.

A fifth direction concerns cross-cultural and Global South perspectives. Much of the algorithmic HRM literature and regulation is shaped by North American and European contexts (Fabris et al., 2025; Rigotti & Fosch-Villaronga, 2024). Because labor markets, legal frameworks, data infrastructures, and organizational cultures differ across regions, future research should examine how algorithmic HRM is adopted, resisted, regulated, and localized across institutional environments. This includes settings where informal hiring networks, uneven data quality, and limited audit capacity shape whether fairness models can be implemented in practice. This follows Tambe et al.'s (2019) broader call for context-sensitive HR AI research.

Finally, future research should evaluate algorithmic HRM through longitudinal and causal designs. Vendor claims and cross-sectional evidence are not enough. Randomized trials, quasi-experiments, audit studies, qualitative fieldwork, and mixed-method research can test whether algorithmic systems improve hiring quality, reduce bias, support development, or increase trust over time. Tambe et al. (2019) emphasize causal reasoning because HR interventions often have unintended effects.

10. Conclusion

This chapter examined algorithmic HRM as a high-impact socio-technical system. It situated algorithmic HRM within digital HRM, people analytics, and algorithmic decision support; then analyzed recruitment analytics, bias sources, mitigation strategies, performance prediction, and governance. It also proposed a Responsible Algorithmic HRM Governance Model that links the data, model, decision, governance, and outcome layers of algorithmic HRM. Overall, the chapter argued that bias is a property of the relationship among data, organizational history, measurement, model design, decision practice, and governance.

The discussion of bias mitigation showed that pre-processing, in-processing, and post-processing strategies are important but incomplete. AI Fairness 360 illustrates technical bias detection and mitigation (Bellamy et al., 2019), while broader reviews show that bias can arise beyond model training itself (Mehrabi et al., 2021). Fairness metrics are valuable diagnostic tools, but they embody normative assumptions and must be interpreted in context (Barocas et al., 2023). The chapter also emphasized that performance prediction can support development while still creating risks of surveillance, narrow measurement, and diminished trust.

The main argument is that algorithmic HRM should be judged by more than predictive accuracy. The deeper question is whether predictions are valid, fair, explainable, contestable, and accountable within a broader risk governance framework (National Institute of Standards and

Technology, 2023), while remaining compatible with employee dignity. The proposed Responsible Algorithmic HRM Governance Model contributes by connecting recruitment analytics, performance prediction, and bias mitigation within a single governance-oriented framework. In line with Selbst et al. (2019), it shifts the discussion from isolated technical debiasing to a socio-technical understanding of responsible algorithmic HRM. The model is nevertheless conceptual; it has not been empirically tested, and future research should examine its applicability across organizational, sectoral, and legal contexts.

For practitioners, algorithmic HRM should not be adopted as a black-box efficiency solution. Consistent with the NIST AI Risk Management Framework, organizations must ask what data represent, what models optimize, how scores influence decisions, who is accountable, and what outcomes emerge over time (National Institute of Standards and Technology, 2023). For policymakers, employment-related AI deserves heightened scrutiny because it affects opportunity, livelihood, and dignity, especially where regulation classifies employment AI as high-risk or scrutinizes AI-based selection procedures (European Parliament and Council, 2024; U.S. Equal Employment Opportunity Commission, 2023). Responsible algorithmic HRM requires treating algorithms as governed organizational systems, not neutral decision machines.

References

Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). *Artificial intelligence risk management framework: Generative artificial intelligence profile (NIST AI 600-1)*. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.600-1

Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities*. MIT Press. https://fairmlbook.org/

Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., Lohia, P., Martino, J., Mehta, S., Mojsilovic, A., Nagar, S., Ramamurthy, K. N., Richards, J., Saha, D., Sattigeri, P., Singh, M., Varshney, K. R., & Zhang, Y. (2019). AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development, 63*(4/5), 4:1-4:15. https://doi.org/10.1147/JRD.2019.2942287

Campbell, J. P., & Wiernik, B. M. (2015). The modeling and assessment of work performance. *Annual Review of Organizational Psychology and Organizational Behavior, 2*, 47-74. https://doi.org/10.1146/annurev-orgpsych-032414-111427

City of New York Department of Consumer and Worker Protection. (n.d.). *Automated employment decision tools (AEDT)*. Retrieved May 26, 2026, from https://www.nyc.gov/site/dca/about/automated-employment-decision-tools.page

Erduran, M. (2025). Ethical and legal implications of AI-driven MIS solutions. In A. G. Yüksek & O. Kaynar (Eds.), *Information technologies in managerial decision-making: Foundations and application* (pp. 283-306). Livre de Lyon.

European Commission. (n.d.). *AI Act*. Retrieved May 26, 2026, from https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

European Commission. (2026a, May 8). *Commission opens consultation on draft guidelines for AI transparency obligations*. Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/news/commission-opens-consultation-draft-guidelines-ai-transparency-obligations

European Commission. (2026b, May 19). *Commission seeks feedback on the draft guidelines for the classification of high-risk artificial intelligence systems*. Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/news/commission-seeks-feedback-draft-guidelines-classification-hi

gh-risk-artificial-intelligence-systems

European Parliament and Council. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

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 Transactions on Intelligent Systems and Technology, 16*(1), Article 16. https://doi.org/10.1145/3696457

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM, 64*(12), 86-92. https://doi.org/10.1145/3458723

Harman, J. L., & Scheuerman, J. (2023). Simple rules outperform machine learning for personnel selection: Insights from the 3rd annual SIOP machine learning competition. *Discover Artificial Intelligence, 3*, Article 2. https://doi.org/10.1007/s44163-022-00044-2

Hunkenschroer, A. L., & Luetge, C. (2022). Ethics of AI-enabled recruiting and selection: A review and research agenda. *Journal of Business Ethics, 178*, 977-1007. https://doi.org/10.1007/s10551-022-05049-6

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals, 14*(1), 366-410. https://doi.org/10.5465/annals.2018.0174

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*, 795-848. https://doi.org/10.1007/s40685-020-00134-w

Leicht-Deobald, U., Busch, T., Schank, C., Weibel, A., Schafheitle, S., Wildhaber, I., & Kasper, G. (2019). The challenges of algorithm-based HR decision-making for personal integrity. *Journal of Business Ethics, 160*(2), 377-392. https://doi.org/10.1007/s10551-019-04204-w

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems 30* (pp. 4765-4774). Curran Associates. https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions

Marler, J. H., & Boudreau, J. W. (2017). An evidence-based review of HR analytics. *The International Journal of Human Resource Management, 28*(1), 3-26. https://doi.org/10.1080/09585192.2016.1244699

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), Article 115. https://doi.org/10.1145/3457607

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220-229). Association for Computing Machinery. https://doi.org/10.1145/3287560.3287596

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). https://doi.org/10.6028/NIST.AI.100-1

Neumann, M., Niessen, S., Linde, M., Tendeiro, J., & Meijer, R. (2022). *"Adding an egg" in algorithmic decision making: Improving stakeholder and user perceptions and predictive validity.* PsyArXiv. https://doi.org/10.31234/osf.io/743hn

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors, 52*(3), 381-410. https://doi.org/10.1177/0018720810376055

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 469-481). Association for Computing Machinery. https://doi.org/10.1145/3351095.3372828

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135-1144). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939778

Rigotti, C., & Fosch-Villaronga, E. (2024). Fairness, AI & recruitment. *Computer Law & Security Review, 53*, Article 105966. https://doi.org/10.1016/j.clsr.2024.105966

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

Schwartz, R., Vassilev, A., Greene, K. K., Perine, L., Burt, A., & Hall, P. (2022). *Towards a standard for identifying and managing bias in artificial intelligence (NIST SP 1270)*. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.1270

Selbst, A. D., boyd, d., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 59-68). Association for Computing Machinery. https://doi.org/10.1145/3287560.3287598

Suresh, H., & Guttag, J. V. (2021). A framework for understanding sources of harm throughout the machine learning life cycle. In *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. Association for Computing Machinery. https://doi.org/10.1145/3465416.3483305

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review, 61*(4), 15-42. https://doi.org/10.1177/0008125619867910

Tursunbayeva, A., Di Lauro, S., & Pagliari, C. (2018). People analytics: A scoping review of conceptual boundaries and value propositions. *International Journal of Information Management, 43*, 224-247. https://doi.org/10.1016/j.ijinfomgt.2018.08.002

U.S. Equal Employment Opportunity Commission. (2007). *Employment tests and selection procedures*. https://www.eeoc.gov/laws/guidance/employment-tests-and-selection-procedures

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://www.eeoc.gov/eeoc-disability-related-resources/artificial-intelligence-and-ada

U.S. Equal Employment Opportunity Commission. (2023). *Select issues: 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://www.eeoc.gov/laws/guidance/select-issues-assessing-adverse-impact-software-algorithms-and-artificial-intelligence