
Research Article

The Impact of Artificial Intelligence on Recruitment Decision-Making: An Integrative Critical Review of Opportunities, Risks, And Human Oversight

Ben Bander Abudawood

Abertay University

Abstract:

Artificial intelligence (AI) is moving recruitment from a largely human-administered process toward a socio-technical decision system in which algorithms source, rank, screen, assess, and recommend candidates. This integrative critical review examines how AI alters the validity, fairness, legitimacy, and accountability of recruitment decision-making and what forms of human oversight are required when employment opportunities are at stake. The paper synthesizes peer-reviewed empirical studies, established systematic reviews and meta-analytic evidence, foundational research in personnel selection and human–automation interaction, and selected authoritative regulatory sources. It uses a transparent structured literature-identification strategy and critical thematic synthesis across human resource management, organizational psychology, information systems, business ethics, law, and AI governance. The evidence reveals a persistent duality. AI can increase processing capacity, consistency, documentation, reach, and disciplined use of job-relevant information, and under validated conditions may reduce some forms of idiosyncratic human bias. These benefits are conditional rather than inherent. Historical labels, proxy variables, selective outcome data, construct misspecification, opaque vendor systems, automation bias, privacy intrusion, accessibility barriers, and weak contestability can reproduce or amplify disadvantage at scale. Applicant-reaction evidence further indicates that strongly algorithmic procedures can weaken perceived justice, trust, organizational attractiveness, and willingness to pursue employment. The review rejects both technological solutionism and a return to unaided human judgment. It advances a risk-calibrated model of constrained human–AI augmentation in which automation is strongest for administrative and information-processing tasks, whereas consequential exclusion decisions require validated criteria, documented human review, genuine override authority, subgroup monitoring, explanation, accommodation, and reconsideration. Responsible AI recruitment therefore depends less on whether a human is nominally 'in the loop' than on whether the complete decision architecture preserves validity, fairness, accountability, and genuine human agency.

Keywords: artificial intelligence; recruitment; personnel selection; algorithmic hiring; human oversight; critical literature review

1. Introduction

Recruitment is an unusually revealing setting in which to study artificial intelligence because it combines high-volume information processing with decisions that are both economically consequential and socially sensitive. Hiring decisions distribute access to income, career mobility, organizational membership, and professional identity. They are also made under uncertainty: employers rarely observe future job performance directly and must infer it from imperfect indicators such as education, work history, tests, interviews, work samples, and references. Long before contemporary AI, personnel-selection research therefore treated prediction, validity, standardization, and fairness as central design problems. Meta-analytic work showed that structured, job-relevant selection methods can provide substantial predictive utility, but also that the choice and combination of methods matter (Schmidt & Hunter, 1998). AI does not remove this classical selection problem. It changes its scale, speed, data sources, and institutional visibility. AI-enabled recruitment now encompasses a broad family of technologies rather than a single intervention. Systems may recommend vacancies to prospective applicants, identify passive candidates, parse curricula vitae, rank applicants, administer chatbots, score tests or games, transcribe and analyse interviews, infer person–job fit, or generate shortlists for recruiters. Black and van Esch (2020) describe AI-enabled recruiting across outreach, screening, assessment, and coordination, illustrating that automation can enter almost every stage of the recruitment funnel. This breadth is important because the ethical and evidential stakes are not uniform. Automating interview scheduling is not equivalent to automatically rejecting an applicant; summarising a résumé is not equivalent to predicting “culture fit”; and presenting a recruiter with a ranked list is not equivalent to allowing a model to make the final selection.

The attraction of AI is understandable. Digital recruitment produces application volumes that can overwhelm human recruiters, while machine-learning systems can process large amounts of information consistently and rapidly. Research and practitioner literature therefore associates AI with lower administrative burden, faster screening, wider sourcing, improved matching, and the

possibility of making decisions on a more explicit evidential basis (Black & van Esch, 2020; Ore & Sposato, 2022).

At the same time, the claim that algorithms are objective because they are mathematical is difficult to sustain. Data-driven models learn from choices about what to measure, which cases to include, how outcomes are labelled, which features are permitted, and what objective is optimised. Historical employment data can encode unequal opportunities and managerial preferences; apparently neutral variables can operate as proxies for protected characteristics; and a statistically accurate model can still distribute errors or opportunities unevenly. Barocas and Selbst (2016) established the broader legal and technical problem of disparate impact in data-driven decision systems, while Raghavan et al. (2020) showed that algorithmic hiring vendors make consequential design choices about data, prediction targets, validation, and bias mitigation that are often difficult for clients and candidates to inspect.

These concerns have moved from academic debate into regulation. The European Union Artificial Intelligence Act classifies AI systems used for recruitment and selection as high-risk because they can materially affect careers, livelihoods, and workers' rights. New York City's rules for automated employment decision tools require a recent bias audit, public information about the audit, and candidate or employee notice before covered tools are used. In the United States, the Equal Employment Opportunity Commission (EEOC) has clarified that established anti-discrimination principles under Title VII and the Americans with Disabilities Act apply when employers use software, algorithms, and AI in employment selection. In the United Kingdom, the Information Commissioner's Office (ICO) reported in 2026 that many employers using automated recruitment may be making solely automated decisions and emphasized transparency, fairness monitoring, and genuinely meaningful human involvement. These developments make human oversight a practical governance requirement rather than a rhetorical preference.

The literature, however, remains fragmented. Computer-science research tends to emphasize model performance and formal fairness; personnel psychology focuses on validity and applicant reactions; human resource management considers adoption and organizational value; legal scholarship examines discrimination, data protection, and accountability; and human-computer interaction studies trust, explanation, and reliance. Reviews have addressed algorithmic discrimination and fairness in recruitment (Köchling & Wehner, 2020), ethics in AI-enabled recruiting and selection, and validity and applicant reactions to digital selection procedures (Woods et al., 2020). Yet the decision-making problem itself deserves a more integrated treatment. The central question is not simply whether AI is "good" or "bad" for recruitment, but how it redistributes epistemic authority between models, recruiters, vendors, managers, and candidates.

This review therefore examines four linked questions: (1) What opportunities does AI create for recruitment decision quality and organizational efficiency? (2) What technical, ethical, legal, and human risks can undermine those benefits? (3) What constitutes meaningful human oversight across different levels of recruitment automation? and (4) What governance architecture is required when predictive validity, fairness, candidate legitimacy, and accountability are treated as joint requirements? The review advances a socio-technical argument: AI recruitment should be evaluated as a decision architecture rather than as a standalone model. A system may be technically sophisticated yet organizationally unsafe if recruiters cannot interrogate its output, if vendors withhold validation evidence, if candidates cannot seek accommodation, or if no actor has clear responsibility for adverse outcomes. Conversely, human involvement is not automatically protective; an untrained recruiter who rubber-stamps algorithmic rankings can add the appearance of accountability without its substance.

2. Conceptual Background and Theoretical Lens

2.1 Recruitment decision-making as prediction and judgment

Recruitment and selection involve prediction under uncertainty, but they also involve normative judgments about which evidence should count. Classical personnel selection distinguishes reliability, criterion-related validity, incremental validity, and utility. AI can strengthen some of these dimensions by applying rules consistently and detecting multivariate patterns, yet predictive performance alone does not settle whether a criterion is legitimate. A model trained to predict supervisor ratings, for example, inherits whatever the ratings measure—including possible differences in opportunity, managerial preference, or workplace bias. The distinction between predicting an observed label and measuring the underlying construct is therefore fundamental. Tambe, Cappelli, and Yakubovich (2019) warn that HR phenomena are complex, organizational datasets are often small or selective, and fairness and accountability constraints limit simple transplantation of data-science methods into HR.

2.2 Human–AI complementarity

Human–AI decision-making research provides a useful alternative to the replacement narrative. Jarrahi (2018) argues that AI contributes computational capacity and analytical consistency, whereas humans remain comparatively strong in interpreting uncertainty, equivocality, contextual meaning, and value conflicts. Parasuraman, Sheridan, and Wickens (2000) similarly distinguish automation of information acquisition, information analysis, decision selection, and action implementation. Applied to recruitment, these frameworks imply that automation should be decomposed by function. A system may appropriately automate résumé parsing or scheduling while leaving interpretation of unusual career histories, accommodation requests, and final exclusion decisions to accountable human decision-makers.

2.3 Organizational justice and applicant reactions

Recruitment is not only a prediction system; it is also a social encounter in which applicants infer whether an organization is respectful, trustworthy, and fair. Acikgoz et al. (2020) found that AI-based interviewing was generally perceived as less procedurally and interactionally just than traditional human interviewing, with two-way communication playing an important role. Lavanchy et al. (2023) likewise found across four studies that applicants perceived algorithm-driven hiring as less fair than human-only or algorithm-assisted human processes, partly because they doubted an algorithm could recognize their uniqueness. These findings caution against evaluating AI solely through recruiter-side efficiency. A process that saves recruiter time but reduces perceived voice, dignity, or opportunity to explain may damage organizational attractiveness and candidate trust.

2.4 Fairness as a socio-technical property

Formal fairness metrics are necessary but insufficient. Different fairness criteria can conflict, and statistical parity does not by itself establish substantive justice. Selbst et al. (2019) show that fairness failures often arise when designers abstract a technical system away from the social institutions in which it operates. Recruitment illustrates this problem clearly: the relevant unit is not merely a classifier but a pipeline involving job analysis, sourcing, data collection, model development, thresholds, recruiter interpretation, interviews, accommodations, and final decisions. Fairness can therefore fail upstream through exclusionary job requirements, within the model through biased data, or downstream through selective human overrides.

2.5 Trust, algorithm aversion, and automation bias

Human oversight is complicated by inconsistent reliance on algorithms. Dietvorst, Simmons, and Massey (2015) show that people can become excessively averse to algorithms after observing errors, even when the algorithm remains more accurate than a human forecaster. Logg, Minson, and Moore (2019), however, demonstrate algorithm appreciation in other settings, where people weight algorithmic advice more heavily than human advice. Recruitment systems must therefore be designed for calibrated reliance rather than maximum trust. Both under-reliance and over-reliance can degrade decisions: recruiters may ignore useful evidence because it is algorithmic, or accept a ranking uncritically because it appears quantitative and authoritative.

Table 1. Review framework and evidence-selection criteria

Element	Operationalization
Review purpose	Critical integration of evidence on AI-enabled recruitment decision-making, opportunities, risks, and human oversight
Context	Job applicants, recruiters, hiring managers, employers, and recruitment-technology providers
Phenomenon	AI-, ML-, algorithm-, or automation-enabled recruitment and selection
Core outcomes	Efficiency, validity, decision quality, fairness, discrimination, transparency, privacy, applicant reactions, accountability, and human oversight
Evidence prioritized	Peer-reviewed empirical studies, meta-analyses, systematic reviews, substantive conceptual studies, and authoritative regulatory sources
Exclusions	Unrelated HR functions; promotional vendor claims; unsupported commentary; peripheral sources
Synthesis	Critical integrative thematic synthesis using socio-technical and human–AI decision-making lenses

3. Methodology

3.1 Review design

This article is a structured integrative critical literature review based exclusively on secondary sources. It does not report primary data and does not claim to be a completed PRISMA systematic review. This positioning is deliberate. A conventional systematic review requires a fully documented search history, database exports, deduplication records, screening decisions, exclusion reasons, and reproducible study-flow counts. Those records were not generated as part of the present manuscript; consequently, reconstructing PRISMA counts retrospectively would create a false impression of methodological precision. The review instead uses explicit search concepts, transparent source-selection principles, evidential traceability, and critical thematic synthesis to integrate a multidisciplinary body of evidence.

An integrative design is appropriate because AI-enabled recruitment is not a single standardized intervention. It includes sourcing, résumé parsing, ranking, chatbot interaction, assessment scoring, interview analysis, matching, and generative-AI assistance. Outcomes are similarly heterogeneous: efficiency, predictive validity, subgroup error, discrimination, applicant reactions, privacy, accessibility, transparency, accountability, and human oversight. These questions are distributed across organizational psychology,

HRM, information systems, computer science, business ethics, law, and public policy. The purpose is therefore explanatory synthesis: to identify where findings converge, where they conflict, and which socio-technical mechanisms explain those differences.

3.2 Analytical questions

Four analytical questions guide the review. First, where in the recruitment process does AI alter decision authority, and what forms of organizational value are supported by evidence? Second, through which mechanisms can AI introduce or amplify error, discrimination, opacity, privacy intrusion, accessibility barriers, or adverse applicant reactions? Third, when does human involvement improve decisions, and what distinguishes meaningful oversight from nominal human presence or rubber-stamping? Fourth, what governance architecture follows when predictive validity, fairness, candidate legitimacy, and accountability are treated as joint requirements rather than independent objectives?

3.3 Literature-identification strategy

Relevant literature was identified through combinations of technology, recruitment, and governance concepts. Technology terms included "artificial intelligence", "machine learning", algorithm*, automat*, "predictive analytics", "large language model*", and generative AI. Recruitment terms included recruit*, hiring, "personnel selection", "employee selection", "talent acquisition", "resume screening", interview*, candidate matching, and applicant*. Decision and governance terms included validity, fairness, bias, discrimination, transparency, explainab*, privacy, accessibility, "applicant reaction*", trust, "human oversight", "human in the loop", accountability, and audit*. Searches and citation chaining were used to identify peer-reviewed studies and established reviews through multidisciplinary scholarly discovery services and publisher databases. Foundational literature was retained selectively where necessary to interpret contemporary AI recruitment, particularly research on personnel-selection validity, disparate impact, organizational justice, algorithmic fairness, and human–automation reliance.

Recent, clearly identifiable evidence syntheses were used as triangulation points rather than as substitutes for a screening log of the present paper. Dadaboyev et al. (2025) systematically reviewed 49 peer-reviewed articles on AI in employee recruitment published between 2018 and 2025. Ogunoye et al. (2026) reviewed 79 peer-reviewed articles published between 2002 and 2024 and distinguished short-term recruitment efficiency from broader effectiveness. Moritz et al. (2026) conducted a meta-analysis of reactions to algorithmic decision-making in HRM, synthesizing 365 effect sizes from 73 samples in 53 studies (N = 24,578). These numerical counts describe those published reviews only; they are not presented as identification, screening, or inclusion counts for the current article.

3.4 Source-selection principles

Sources were retained when they made a direct and substantive contribution to AI-enabled recruitment or to a theoretical mechanism necessary for evaluating recruitment decisions. Priority was given to peer-reviewed empirical studies, meta-analyses, systematic reviews, well-established conceptual articles, and primary regulatory or legal materials. Foundational non-AI studies were included selectively where they establish the comparison standard, such as evidence on selection validity or discrimination in conventional hiring. Purely promotional vendor material, unsupported commentary, and sources with only peripheral relevance to recruitment decision-making were excluded from the evidential core.

The selection logic emphasized evidential traceability. Precise numerical claims were retained only when the underlying source could be clearly identified and verified. Where publication metadata were ambiguous or future-dated, the source was either described according to the publisher's formal bibliographic status or omitted if it was not essential to the argument. This rule reduces citation uncertainty in a rapidly changing field.

3.5 Critical appraisal

The literature was appraised through six questions rather than a single quality score. First, construct validity: does the technology measure a job-relevant attribute or merely predict a convenient historical label? Second, criterion quality: is the outcome used to train or validate the system itself a defensible measure of performance? Third, external validity: do hypothetical scenarios, single occupations, or proprietary demonstrations generalize to consequential hiring? Fourth, comparator quality: is AI compared with realistic human practice or with an idealized unbiased recruiter? Fifth, fairness completeness: are subgroup errors, accessibility, and intersectional effects considered alongside average accuracy? Sixth, transparency and independence: are data provenance, model purpose, validation procedures, vendor involvement, and limitations sufficiently visible for scrutiny?

3.6 Evidence synthesis

Evidence was first mapped to the recruitment pipeline—sourcing, screening, assessment, interviewing, ranking, and final selection—and then compared across recurrent mechanisms. The cross-study synthesis focused on standardization versus construct misspecification; predictive scale versus historical and proxy bias; efficiency versus contextual information loss; opacity versus auditability; automation versus candidate voice; and algorithmic consistency versus meaningful human judgment.

This structure allows apparently contradictory findings to coexist when they arise from different technologies, autonomy levels, samples, or outcome definitions.

Greater interpretive weight was given to meta-analytic and systematic-review evidence for broad patterns, controlled experiments for applicant-reaction findings, validated personnel-selection research for predictive-quality claims, technical and legal scholarship for mechanisms of algorithmic discrimination, and primary regulatory sources for governance obligations. The paper does not pool heterogeneous effect sizes or claim exhaustive coverage. Its contribution is analytical integration: explaining why benefits and harms emerge under different recruitment architectures and deriving governance implications from those mechanisms.

3.7 Methodological limitations

The integrative design has limitations. It is more vulnerable than a fully protocolled systematic review to selection effects in literature identification, and it cannot legitimately provide a PRISMA flow diagram or claim exhaustive coverage of every eligible study. The evidence base is heterogeneous, proprietary systems are often difficult to validate independently, and many applicant-reaction studies use experimental scenarios rather than actual employment decisions. Rapid technological change, particularly generative AI, can also outpace peer-reviewed evaluation. These limitations are acknowledged explicitly rather than masked by reconstructed screening counts. The conclusions should therefore be interpreted as a rigorous critical synthesis of identifiable secondary evidence, not as an exhaustive meta-analytic estimate of a single treatment effect.

4. Critical Literature Review

4.1 From electronic recruitment to algorithmic decision systems

The literature shows an important conceptual shift from digitizing recruitment to delegating judgment. Early e-recruitment primarily changed communication and application handling; contemporary AI increasingly transforms how evidence is generated and weighted. Black and van Esch (2020) map AI across outreach, screening, assessment, and coordination, while more recent reviews show that machine learning still dominates implemented automated-hiring research even as LLMs and hybrid systems expand.

This distinction matters because administrative digitization mostly changes transaction costs, whereas algorithmic scoring changes epistemic authority: the system does not merely carry information to the recruiter but participates in defining which candidates appear promising.

The literature is therefore best understood as a continuum of decision delegation. At one end, AI performs low-discretion tasks such as scheduling. In the middle, it structures information through parsing, matching, summaries, and ranked recommendations. At the high-discretion end, it evaluates tests, interviews, or behavioral signals and can determine who proceeds. The evidential burden should rise across this continuum because the cost of measurement error and the difficulty of detecting wrongful exclusion increase as automation moves closer to the final employment decision.

4.2 Efficiency is well supported; effectiveness is more conditional

Efficiency is the most robust positive theme. Automated systems can process large applicant pools, standardize repetitive screening, and reduce recruiter workload. Yet efficiency is not equivalent to effectiveness. Ogunoye et al. (2026), in a systematic review of 79 peer-reviewed articles published between 2002 and 2024, identify a 'resourcing paradox': AI can improve operational speed and data processing while creating candidate-experience, opacity, and governance costs that weaken broader recruitment effectiveness. The strongest interpretation of the evidence is consequently not that AI improves hiring, but that AI improves particular information-processing operations under particular conditions. Whether this becomes an effective selection system depends on job analysis, predictor validity, data quality, threshold choice, human use of the output, and post-deployment monitoring. This returns AI recruitment to a central principle of personnel psychology: sophisticated prediction cannot compensate for a weak criterion or an invalid measure.

4.3 Validity: prediction is not the same as measuring merit

A recurring weakness in the AI-recruitment discourse is the tendency to treat predictive accuracy as self-validating. Schmidt and Hunter (1998) established the importance of criterion-related validity in personnel selection, but AI complicates the criterion itself. Historical hiring decisions, supervisor ratings, tenure, or sales outcomes may be convenient labels without being unbiased measures of capability. Tambe et al. (2019) emphasize that HR datasets are generated inside organizations with complex social processes; they are not neutral scientific observations.

Thus, a model can predict a historical outcome accurately while learning managerial preferences, unequal opportunity, or organizational structures that should not be reproduced.

This creates a target-validity problem. Before asking whether a model predicts well, researchers and employers must ask whether the target represents the construct that legitimate recruitment should optimize. The problem is especially acute for 'fit', personality, engagement, or inferred soft skills, where behavioral proxies can be distant from established constructs. Numerical precision can conceal conceptual weakness. Deep analysis therefore requires separating model discrimination/calibration from psychometric construct validity and from the normative legitimacy of the outcome being predicted.

4.4 Bias and fairness: AI can interrupt and reproduce discrimination

The fairness literature resists a one-directional conclusion. Conventional human recruitment is demonstrably vulnerable to discrimination; Bertrand and Mullainathan (2004), for example, showed unequal callbacks for otherwise equivalent résumés associated with racially coded names. Standardized algorithmic procedures can therefore remove some idiosyncratic discretion. At the same time, Barocas and Selbst (2016), Köchling and Wehner (2020), and Raghavan et al. (2020) show why data-driven systems can reproduce disadvantage through historical labels, proxy variables, sampling, feature engineering, and optimization choices. The correct comparison is not biased algorithms versus unbiased humans, but alternative decision systems with different error-generating mechanisms.

Fairness must also be distinguished across levels. Measurement fairness asks whether a tool measures the same construct comparably across groups. Error fairness asks how false positives and false negatives are distributed. Allocation fairness asks who receives interviews and jobs. Procedural fairness asks whether applicants have voice, consistency, explanation, and correction. Legal fairness asks whether protected groups experience prohibited adverse impact or inaccessible processes. These dimensions can diverge, so declaring a system 'fair' on one statistical metric is analytically incomplete.

4.5 Applicant reactions: the missing relational dimension

Applicant-reaction studies expose a dimension that technical validation can miss. Acikgoz et al. (2020) found lower procedural and interactional justice for AI interviewing than for traditional human interviewing, and Lavanchy et al. (2023) found that algorithm-driven hiring can be perceived as less fair than human or algorithm-assisted human processes because applicants doubt that algorithms recognize their uniqueness.

Moritz et al. (2026) provide aggregate evidence: their meta-analysis synthesized 365 effect sizes from 73 samples in 53 studies ($N = 24,578$) and found negative associations between algorithmic decision-making and system-related reactions such as justice, fairness, and trust, as well as organization-related reactions such as organizational attractiveness and job-pursuit intention, with effects varying by interaction type and decision extent.

These results do not prove that candidates always prefer humans. Rather, they indicate that automation changes the social meaning of selection. Applicants care not only about consistent scoring but also about whether they can communicate, explain exceptional circumstances, receive recognition as individuals, and believe that the decision-maker can exercise empathy and contextual judgment. The deeper implication is that procedural legitimacy is itself an organizational outcome. A technically accurate process may still reduce applicant cooperation or organizational attractiveness if it is experienced as unresponsive or inscrutable.

4.6 Transparency, explainability, and contestability

Explainability is often presented as a technical property, but recruitment requires multiple forms of explanation. Developers need diagnostic explanations to identify failure modes. Recruiters need operational explanations that connect outputs to job-relevant evidence and communicate uncertainty. Candidates need process explanations sufficient to understand what information was used and how to seek correction, accommodation, or reconsideration. Barredo Arrieta et al. (2020) provide the broader XAI foundation, while recent recruitment research indicates that transparency can influence perceived fit and acceptance. A single feature-importance chart cannot satisfy all of these audiences.

Contestability is therefore a stronger governance concept than transparency alone. A candidate can be told that AI was used and still have no practical way to correct an erroneous résumé parse or challenge a misclassification. Meaningful accountability requires that explanation be connected to action: correction of data, accessible human contact, accommodation, review of borderline cases, and authority to reverse an automated recommendation.

4.7 Human oversight: why 'human in the loop' is not enough

The literature increasingly treats human involvement as necessary, but the phrase 'human in the loop' can obscure rather than clarify governance.

Parasuraman et al. (2000) show that automation can occur at different stages of information acquisition, analysis, decision selection, and implementation. Jarrahi (2018) and Puranam (2021) further suggest that human-AI complementarity depends on task allocation and organization design. In recruitment, a human who sees only a vendor score and has seconds to approve it is formally present but substantively weak.

Meaningful oversight requires competence, information, time, independence, and authority. It must also be monitored. Human override can reduce algorithmic error, but unconstrained override can reintroduce favoritism and inconsistency. The relevant design problem is therefore calibrated reliance: determine when humans should defer to validated evidence, when they should investigate disagreement, and when the system should be prevented from acting autonomously. Oversight is strongest when disagreement is treated as diagnostic information and when override patterns themselves are audited.

4.8 A synthesis gap: most studies evaluate components, not the whole hiring system

A major gap across the literature is unit of analysis. Studies often evaluate a model, an interview interface, a fairness metric, or

applicant perceptions in isolation. Real hiring is a pipeline. Bias can enter through sourcing before an AI screener operates; a fair ranking can be undermined by an inaccessible assessment; a valid model can be followed by an unstructured interview; and a transparent tool can still be used by recruiters who systematically override particular candidates. Recent socio-technical work therefore argues that AI recruitment should be analyzed through relationships among employers, recruitment firms, technology developers, recruiters, and applicants rather than as a self-contained algorithm.

This review builds on that gap by treating decision quality as an emergent property of the full architecture. The analytical focus shifts from 'Is the algorithm fair?' to 'Under what combination of data, measurement, automation, human judgment, institutional rules, and candidate rights does the recruitment system produce defensible decisions?' That reframing is the basis for the results and governance model that follow.

5. Thematic Findings and Deep Analysis

5.1 The changing location of decision authority

Across the literature, the most consequential effect of AI is not simple task automation but the relocation of decision authority. Traditional recruitment distributes judgment across recruiters, hiring managers, tests, interviews, and organizational rules. AI inserts additional actors—data scientists, model developers, vendors, platform owners and converts some tacit judgments into formal scores or rankings. This can improve consistency, but it can also make accountability diffuse. A recruiter may regard a vendor score as objective; the vendor may state that the employer controls the final decision; and the candidate may be unable to identify either the data or rationale that produced exclusion. The result is an accountability gap unless responsibilities are specified before deployment.

5.2 Opportunities: where AI can improve recruitment decisions

Efficiency and scale

The strongest and most consistent opportunity is operational. AI can parse applications, identify minimum qualifications, schedule interactions, answer routine questions, and prioritize records for review. Black and van Esch (2020) emphasize efficiency across outreach, screening, assessment, and coordination. Ore and Sposato (2022) similarly report that recruiters value the delegation of routine work. These gains matter because high application volumes create a practical risk that human screening becomes cursory, inconsistent, or dependent on easily observed cues.

Consistency and standardization

Algorithms can apply the same scoring rule to every applicant, reducing day-to-day variation, fatigue effects, and some idiosyncratic preferences. Standardization is valuable when criteria are demonstrably job-related. The opportunity is therefore not “objectivity” in an absolute sense but procedural consistency around validated constructs. This mirrors the long-standing advantage of structured over unstructured selection methods.

Decision support and pattern recognition

Machine learning can integrate multiple job-relevant predictors and detect interactions that are difficult to process manually. In principle, this can improve prioritization and person–job matching, particularly when models are trained and validated on relevant outcomes. Recent experimental work on transparency also suggests that clearer information about how AI reaches person–job fit assessments can reduce discrepancies between AI and human judgments (Chen et al., 2025).

Bias interruption

AI can sometimes reduce human bias by removing irrelevant identity cues, structuring evaluation, or forcing decisions to rely on pre-specified evidence. This is an important counterweight to romanticized accounts of human judgment. Field experiments demonstrate that human hiring can exhibit discrimination, including the classic finding by Bertrand and Mullainathan (2004) that otherwise equivalent résumés with White-sounding names received substantially more callbacks than those with African-American-sounding names. The relevant comparison is therefore responsible AI versus realistic human practice, not AI versus an ideal unbiased recruiter.

Auditability and documentation

Digital systems can generate logs, version histories, score distributions, and subgroup outcomes that are difficult to reconstruct in informal human screening. When organizations preserve these records, they create the possibility of monitoring adverse impact, identifying model drift, reviewing overrides, and investigating complaints. Auditability is a major governance advantage, but only if the organization has access to the necessary data and does not treat proprietary vendor claims as a substitute for evidence.

Candidate access and responsiveness

Chatbots and automated coordination can provide 24-hour communication, status updates, and consistent answers. Accessibility benefits may arise when systems offer multiple channels and formats. However, this opportunity reverses if automation creates barriers for disabled candidates or if accommodation processes are difficult to reach.

5.3 Risks: how AI can degrade recruitment decisions

Historical bias and label bias

Machine learning learns from observed data, not from a neutral record of merit. If past hiring, performance ratings, promotion, or

retention reflect unequal opportunity, models can reproduce those patterns.

Barocas and Selbst (2016) explain how apparently neutral data mining can inherit structural disadvantage. In recruitment, the problem is especially acute when the target variable such as prior manager rating, tenure, sales, or historical hiring success is treated as a clean measure of worker quality.

Proxy discrimination and feature leakage

Removing protected characteristics does not remove all information correlated with them. Postcodes, schools, employment gaps, language patterns, names, online behavior, and occupational histories can act as proxies. Raghavan et al. (2020) show that bias mitigation therefore requires scrutiny of data construction and prediction targets, not simply deletion of explicit demographic variables.

Representation and intersectional error

Models trained on unrepresentative populations may perform differently across groups. The wider facial-analysis literature provides a cautionary example: Buolamwini and Gebru (2018) found large intersectional accuracy disparities in commercial gender-classification systems, with particularly high error for darker-skinned women. Recruitment tools that rely on face, voice, language, or behavioral signals can inherit analogous measurement problems. The lesson is not that every hiring model produces the same disparities, but that average accuracy can conceal severe subgroup error.

Construct validity and pseudoprecision

A technically accurate model can predict the wrong thing. Recruitment vendors may claim to infer personality, employability, engagement, or “fit” from digital traces, games, video, or language. Yet the more remote the feature is from a validated job construct, the greater the danger of pseudoprecision: a numerical score creates an appearance of measurement without adequate construct validity. Personnel-selection standards require evidence that predictors relate to job requirements and outcomes; AI should not receive a lower evidential threshold merely because the method is computational.

Opacity and explanation deficits

Complex models and proprietary systems can make it difficult to understand why an applicant received a score. Explainable AI research treats interpretability as central to responsible deployment (Barredo Arrieta et al., 2020).

In hiring, explanation has at least three audiences: developers need diagnostic explanations; recruiters need actionable reasons and limitations; candidates need intelligible information sufficient to understand the process, request accommodation, or challenge an error. One generic “AI was used” notice does not satisfy all three functions.

Privacy and data minimization

AI recruitment can expand the amount and intimacy of data collected about candidates. Video, voice, social media, inferred traits, geolocation, behavioral telemetry, and third-party data may extend evaluation beyond information applicants reasonably expect to be job-relevant. The risk is not only unauthorized disclosure but function creep: data gathered for one purpose may be repurposed for prediction without clear necessity.

Applicant justice and organizational attractiveness

Efficiency gains can impose relational costs. Acikgoz et al. (2020) and Lavanchy et al. (2023) show that applicants can perceive algorithmic procedures as less fair, particularly when human interaction, two-way communication, or recognition of individual circumstances is reduced. These findings indicate that applicant reactions are not merely a public-relations issue; they can affect acceptance, withdrawal, employer reputation, and the diversity of the eventual applicant pool.

Automation bias and rubber-stamping

Human oversight can fail when a nominal reviewer defers to a system. Quantified outputs can acquire unwarranted authority, particularly when recruiters lack statistical literacy or cannot see the underlying model. A human who clicks “approve” after viewing an opaque score is not exercising meaningful judgment. The ICO’s 2026 recruitment work specifically emphasizes that meaningful human involvement must be real and consistently applied, not a formal step added to a predominantly automated decision.

Inconsistent override and reintroduction of bias

The opposite problem also occurs: humans may override algorithms selectively and without documentation. If overrides are more common for favored candidates or based on intuition, the organization can reintroduce the very inconsistency that automation was intended to reduce. Oversight therefore requires structured override criteria and audit trails rather than unlimited discretion.

Accessibility and disability discrimination

Automated assessments can disadvantage candidates whose disabilities affect interaction with timed tests, video interfaces, speech recognition, games, or other standardized inputs. The EEOC and U.S. Department of Justice have warned that employers must consider disability impacts and reasonable accommodations when using algorithmic tools. A tool can be statistically valid for an average population and still be unlawful or unfair when it screens out qualified disabled candidates because of interface or measurement design.

Vendor dependency and accountability diffusion

Recruitment AI is frequently purchased rather than built. Employers may not receive training data, source code, validation reports, or sufficient subgroup results. Yet outsourcing technical development does not outsource the employment decision’s consequences. Sánchez-Monedero, Dencik, and Edwards (2020) and Raghavan et al. (2020) both show why vendor claims about bias mitigation

require independent scrutiny. Procurement is therefore a core part of AI governance.

Table 2. Synthesis of representative evidence

Study/source	Design/focus	Principal contribution to this review
Schmidt & Hunter (1998)	Meta-analytic personnel-selection evidence	Establishes validity/utility as baseline criteria for selection methods.
Tambe et al. (2019)	Conceptual HRM/data-science analysis	Identifies HR complexity, small data, fairness/accountability, and employee reaction challenges.
Black & van Esch (2020)	AI-enabled recruiting framework	Maps AI across outreach, screening, assessment, and coordination; emphasizes efficiency.
Raghavan et al. (2020)	Technical/legal analysis of hiring vendors	Shows bias risks in data, targets, validation, and vendor practices.
Köchling & Wehner (2020)	Systematic review	Synthesizes discrimination and fairness risks in algorithmic HR decisions.
Acikgoz et al. (2020)	Two applicant-reaction studies	AI interviewing generally perceived as less procedurally/interactionally just.
Lavanchy et al. (2023)	Four experimental studies	Algorithm-driven hiring perceived as less fair than human or algorithm-assisted human processes.
Chen (2023)	Review of ethics/discrimination	Synthesizes quality/efficiency potential alongside discrimination risks and mitigation.
EU AI Act (2024)	Regulation	Classifies employment recruitment/selection AI as high-risk.
ICO (2026)	Regulatory findings from employer engagement	Emphasizes transparency, fairness monitoring, and meaningful human involvement.

5.4 Human oversight: from symbolic review to decision governance

Human oversight is frequently recommended but poorly specified. The phrase “human in the loop” can describe radically different arrangements: a recruiter may merely see an AI score; may be allowed to override it; may be required to independently review evidence; or may make the decision first and use AI only as a second opinion. These designs produce different risks. Parasuraman et al. (2000) provide a useful starting point because they treat automation as a continuum across information acquisition, analysis, decision selection, and implementation rather than a binary human-versus-machine choice.

For recruitment, oversight should be risk-calibrated. Low-stakes administrative tasks can be highly automated. Information-processing tasks such as parsing and summarization can also be automated if accuracy is monitored and source information remains accessible. Predictive ranking should generally operate as decision support, with validated job-related criteria, uncertainty information, and structured review. Final rejection or selection decisions, particularly when based on novel, opaque, or high-dimensional assessments, should retain accountable human authority and an accessible route for reconsideration.

Meaningful oversight has at least six properties. First, the reviewer must have competence: understanding what the system predicts, its validated use, known limitations, and relevant fairness indicators. Second, the reviewer must have information: not merely a score, but sufficient underlying evidence and explanation. Third, the reviewer must have time; impossible caseloads turn oversight into rubber-stamping. Fourth, the reviewer must have authority to disagree without penalty. Fifth, overrides and reasons should be documented and monitored for systematic patterns. Sixth, candidates need contestability: an avenue to correct data, request accommodation, or obtain human reconsideration.

Human review should also be independent enough to add information rather than simply repeat the model. If a recruiter sees a high-confidence score before examining the résumé, anchoring may contaminate the supposedly independent judgment. In high-stakes stages, a better design may involve staged review: independent human assessment of job-relevant evidence, followed by comparison with the model and structured reconciliation of disagreements. This approach treats disagreement as diagnostic information rather than as a nuisance to be eliminated.

Finally, oversight must extend upstream and downstream. Upstream oversight includes job analysis, choice of outcome variables, procurement, validation, and threshold setting. Downstream oversight includes monitoring selection rates, quality-of-hire indicators,

false-negative patterns, candidate complaints, accommodation requests, model drift, and human overrides. A human at the final click cannot repair a system whose target variable, training data, or procurement assumptions were flawed from the beginning.

Table 3. Risk-calibrated human oversight model for AI recruitment

Recruitment function	Suggested automation level	Required safeguards
Scheduling, reminders, FAQ	High	Accuracy monitoring; accessible alternative channel; privacy controls
Résumé parsing / data extraction	High–moderate	Source verification; error correction; no silent exclusion from parsing failures
Minimum-qualification screening	Moderate	Job-related rules; accommodation route; periodic false-negative review
Candidate ranking / matching	Moderate	Criterion validation; subgroup monitoring; uncertainty/explanation; human review
AI-scored assessments	Moderate–low	Construct and criterion validity; accessibility; bias testing; independent review
Video/voice/behavioral inference	Low unless strongly validated	Necessity test; subgroup accuracy; privacy review; alternative assessment
Interview support / summarization	Moderate	Human access to original evidence; hallucination/error checks; no unsupported trait inference
Final rejection / selection	Low automation	Accountable human decision; documented rationale; contestability and appeal

5.5 Decision quality as a multi-dimensional outcome

A deeper reading of the literature indicates that “decision quality” is frequently used too loosely. In recruitment, a high-quality decision is not simply one that predicts a later criterion with acceptable accuracy. It must also rest on job-relevant constructs, distribute errors defensibly, remain usable by accountable decision-makers, and withstand scrutiny from applicants and regulators. These dimensions can conflict. A highly standardized model may increase reliability while reducing contextual sensitivity; a more interpretable model may sacrifice some predictive complexity; and a procedure that improves average prediction may worsen subgroup false-negative rates. Organizations should therefore abandon single-metric evaluation and assess predictive validity, reliability, subgroup error, applicant reactions, accessibility, operational efficiency, and contestability together.

This multi-dimensional view changes the meaning of optimization. If an employer optimizes only historical retention, it may prefer applicants resembling workers who previously stayed even where turnover was shaped by unequal caregiving burdens, workplace climate, or promotion opportunities. If it optimizes recruiter acceptance of recommendations, it may learn existing preferences rather than job performance. The objective function is therefore a governance choice. Technical optimization cannot determine which organizational outcomes deserve priority; that judgment must be explicit and defensible.

5.6 The hidden asymmetry of false negatives

False negatives deserve particular attention because recruitment data systematically hide them. Employers observe later performance mainly for people they hire, not the counterfactual performance of qualified people screened out. This selective-label problem creates a structural blind spot in validation. Historical data can reinforce earlier selection boundaries because unconventional candidates rarely hired in the past rarely generated outcome data from which a model could learn their potential.

The asymmetry has fairness consequences. A false positive creates a visible organizational cost; a false negative primarily harms the rejected applicant and can remain invisible to the employer. Governance based only on post-hire outcomes can therefore underweight exclusion errors. Human review should be concentrated on borderline rejection cases, unusual career trajectories, employment gaps, non-traditional credentials, and cases where the model operates outside the population on which it was validated.

5.7 Human–AI disagreement as diagnostic information

Most recruitment interfaces implicitly treat agreement between recruiter and algorithm as desirable. A stronger architecture treats disagreement as information. When an accountable reviewer and a validated model reach different conclusions, either the human has noticed contextual evidence unavailable to the model, the model has detected a pattern the human overlooked, or one of the two is relying on an invalid cue.

Structured reconciliation can improve both. Reviewers should record the evidence supporting disagreement; recurring patterns can

reveal model drift, weak constructs, biased human overrides, or missing information.

This approach is preferable to both blind deference and unrestricted intuition. Algorithm aversion can cause useful predictions to be ignored after visible errors, whereas automation bias can make quantified recommendations appear more authoritative than their evidence warrants. Calibrated reliance requires interfaces and training that communicate intended use, uncertainty, and limitations. A ranking without uncertainty can falsely imply meaningful differences between candidates whose predicted scores are practically indistinguishable.

5.8 Generative AI and the changing information environment

Generative AI creates a qualitatively different challenge from conventional predictive models. Applicants can use large language models to draft résumés, cover letters, interview answers, and portfolio narratives, while employers can use the same class of models to summarize applications, generate questions, and recommend candidates. Recruitment is therefore becoming an interaction between algorithmically shaped applicant signals and algorithmically assisted employer interpretation. This weakens the assumption that application documents directly measure communication ability and increases construct contamination.

The appropriate response is not a blanket prohibition on candidate use of generative tools, which could be difficult to enforce and could disadvantage legitimate accessibility or language-support use. Selection should instead reduce reliance on cheaply generated stylistic proxies and increase the weight of job-relevant demonstrations, structured work samples, transparent interviews, and verification of consequential claims. Employers should also protect AI-assisted screening against adversarial manipulation while avoiding disproportionate surveillance.

5.9 Institutional power and accountability

AI recruitment redistributes power among employers, applicants, recruiters, and technology vendors. Applicants usually possess the least information about model design and the least practical ability to refuse its use without abandoning the opportunity. Vendors may possess the strongest technical knowledge while bearing limited direct responsibility for the employment outcome.

Recruiters may be accountable for outputs they did not design and cannot inspect. This asymmetry means that formal consent is an inadequate ethical foundation for high-stakes automated assessment.

Duties should instead follow control and knowledge. Vendors should provide validation evidence, documentation, change notifications, subgroup-testing support, and audit access. Employers should establish job relevance, lawful use, accessibility, monitoring, and final accountability. Recruiters should have training and authority to challenge outputs. Candidates should have intelligible notice and routes for correction, accommodation, and reconsideration. Responsibility can be distributed, but it should never be diffused until no actor is answerable.

6. Discussion

The review supports a conditional rather than deterministic account of AI in recruitment. AI can improve recruitment decisions when it disciplines information processing around validated job-relevant evidence, increases consistency, reduces administrative overload, and creates auditable records. It can worsen decisions when organizations mistake historical prediction for merit, optimize poorly chosen labels, deploy unvalidated proxies, or allow opaque scores to displace professional judgment. The same technology can therefore function as a debiasing instrument in one architecture and a discrimination multiplier in another.

This conditionality helps reconcile apparently conflicting findings in the literature. Studies that emphasize efficiency are often examining throughput, administrative workload, or structured screening. Studies that emphasize injustice or distrust often examine candidate-facing, high-discretion stages such as interviewing or final evaluation. The contradiction is partly resolved when recruitment is disaggregated by task. AI is comparatively well suited to repetitive acquisition and analysis of structured information; it is less obviously suited to interpreting ambiguous narratives, relational cues, unusual career pathways, disability accommodations, or contested notions of “fit.”

A second implication is that fairness cannot be reduced to the absence of protected attributes or to a single statistical test. Recruitment fairness has at least four layers: measurement fairness (does the tool measure the intended construct comparably?); allocation fairness (how are errors and opportunities distributed?); procedural fairness (is the process consistent, transparent, and open to voice?); and substantive/legal fairness (does the process comply with anti-discrimination, disability, and data-protection obligations?).

A system can satisfy one layer and fail another. For example, equal selection rates do not prove construct validity, while a highly predictive system can still be procedurally opaque or inaccessible.

A third implication concerns the comparison standard. Human recruitment is not a neutral benchmark. Historical evidence demonstrates discrimination in conventional hiring, and unstructured interviews are vulnerable to inconsistency. Responsible evaluation should therefore compare AI-assisted processes with credible alternative processes using common outcomes: predictive validity, adverse impact, applicant reactions, cost, time, accessibility, and error correction. This prevents both technological solutionism and nostalgia for unaided human judgment.

The fourth implication is that “human oversight” should be treated as an organizational capability. Puranam (2021) frames human–

AI collaboration as an organization-design problem, which is particularly apt for recruitment. Effective oversight requires role clarity between HR, hiring managers, data/IT teams, legal/compliance functions, and vendors. It requires training, escalation rules, audit access, and resources. Without these, placing a person nominally in the loop can become a governance theatre that leaves the algorithm's practical authority unchanged.

Regulatory developments increasingly reflect this socio-technical understanding. The EU AI Act treats recruitment systems as high-risk, not because every algorithmic hiring tool is discriminatory, but because the context creates significant potential effects on rights and livelihoods. New York City's bias-audit regime moves governance toward measurable external scrutiny, though audits alone cannot guarantee fairness. The EEOC's guidance emphasizes that existing civil-rights duties continue to apply when technology is introduced. The ICO's recent work adds a particularly important operational point: where a decision is said to involve humans, that involvement must be meaningful and consistently applied. These regimes differ, but they converge on documentation, risk assessment, transparency, monitoring, and accountable human control.

The emergence of large language models (LLMs) adds a new layer of complexity. LLMs can summarize applications, draft interview questions, generate candidate communications, and assist with matching. They can also hallucinate, infer unsupported attributes, respond inconsistently to semantically equivalent résumés, and reproduce biases embedded in pre-training data.

Recent research on LLM-based résumé screening is already focused on debiasing while preserving performance. For employers, the practical lesson is that generative fluency should not be confused with psychometric validity. Any LLM output that affects candidate ranking should be treated as a measurement instrument requiring validation, version control, monitoring, and human verification.

Candidate adaptation also deserves more attention. Applicants increasingly use AI to write résumés, optimize keywords, practice interviews, and generate application materials. This creates an algorithm–algorithm interaction: employer systems rank materials partly shaped by candidate-facing systems. The result may be strategic homogenization, where success depends on understanding platform conventions rather than demonstrating job capability. Future recruitment research should therefore study equilibrium effects, not merely the accuracy of one model in isolation.

Another underdeveloped issue is distributional error. Organizations commonly monitor aggregate accuracy, yet hiring is an asymmetric decision problem. False negatives deny potentially qualified candidates an opportunity and may never become visible to the employer; false positives are more likely to be observed after hiring. This asymmetry can make a system appear successful while systematically excluding unconventional but capable applicants. Human oversight should therefore be especially attentive to borderline rejection cases, subgroup false-negative rates, and out-of-distribution candidates.

Taken together, the evidence suggests a principle of constrained augmentation: AI should extend human capacity where its comparative advantages are strongest, while organizational design should constrain automation where errors are difficult to detect, rights are strongly affected, constructs are ambiguous, or contextual judgment is essential. This principle is more defensible than either full automation or mandatory human control at every stage.

7. Practical Framework for Responsible AI Recruitment

Organizations considering AI recruitment should begin with the decision problem rather than the technology. A documented job analysis should identify the constructs that matter, the evidence available to measure them, and the consequences of error. Procurement should then ask whether the proposed system measures those constructs and whether a simpler, more transparent method could achieve the same objective.

Before deployment, employers should require criterion-related and construct-validity evidence appropriate to the role and population, subgroup performance data, accessibility testing, privacy and data-flow documentation, model/version information, and clear statements of intended use. Vendor assurances that a tool is “bias free” or “objective” should not be accepted without methodology and results. Where sensitive attributes can lawfully be used for auditing, organizations should monitor them separately from operational decision features so that disparate outcomes can be detected.

During use, recruiters should receive training in both the tool and its limits. Interfaces should present uncertainty and relevant evidence rather than a single authoritative score. Human review should be structured, with reasons for overrides and escalation rules for ambiguous cases. Candidates should be told when consequential automation is used, what type of information is evaluated, how to request accommodation, and how to seek correction or reconsideration.

After deployment, governance should be continuous. Selection-rate analyses, criterion validity, subgroup errors, candidate complaints, recruiter overrides, drift, and changes in job content should be reviewed periodically. Material model updates should trigger revalidation. High-risk tools should have a named business owner and an independent review function capable of suspending use when evidence is inadequate.

Most importantly, organizations should evaluate the complete hiring funnel. A fair résumé screener cannot compensate for discriminatory sourcing; a transparent model cannot repair an inaccessible assessment; and a well-audited ranking tool can still be undermined by unstructured final interviews. Responsible AI recruitment is therefore a property of the end-to-end process.

Table 4. Governance checklist for employers

Governance question	Minimum evidence before/while using the system
Purpose	Defined recruitment problem; necessity and proportionality documented
Job relevance	Job analysis and construct mapping; criterion-related evidence
Data	Provenance, representativeness, quality, retention, lawful basis
Fairness	Subgroup selection/error analysis; intersectional checks where feasible
Accessibility	Alternative process and reasonable accommodation pathway
Transparency	Candidate notice; recruiter explanation; documented limitations
Human oversight	Competent reviewer, time, information, authority, documented overrides
Vendor governance	Audit rights, change notification, validation evidence, incident support
Monitoring	Drift, adverse impact, quality of hire, complaints, overrides, periodic revalidation
Contestability	Correction, reconsideration, appeal/escalation, human contact

8. Research Agenda

First, future studies should move beyond perceptions and short-term laboratory scenarios toward longitudinal field evidence linking AI-assisted recruitment to actual job performance, retention, diversity, applicant behavior, and organizational outcomes. The literature contains many claims about improved quality of hire but substantially less independent evidence than the strength of those claims implies.

Second, research should compare decision architectures rather than isolated tools. Experiments should test human-first, AI-first, blinded parallel review, consensus, and exception-review designs. This would clarify when human judgment adds independent signal and when it merely anchors on algorithmic output.

Third, fairness research should report error distributions, not only selection rates. False-negative rates, calibration, accessibility outcomes, and intersectional performance are particularly important because excluded applicants are usually invisible in post-hire validation datasets.

Fourth, more work is needed outside North American and Western European legal and labor-market contexts. Recruitment norms, demographic categories, language, disability frameworks, data availability, and institutional protections differ substantially across countries. Models trained in one labor market may not transfer fairly or validly to another.

Fifth, LLM-based recruitment requires dedicated psychometric and audit methods. Research should test stability across prompts and model versions, sensitivity to names and demographic cues, hallucinated résumé facts, adversarial optimization by applicants, and whether explanations accurately reflect the model's decision process.

Sixth, candidate agency should become a core outcome. Studies should examine whether applicants understand AI notices, can meaningfully challenge data, can obtain accommodations, and alter their behavior in response to automated systems. A recruitment system is not fair merely because its designers can explain it to auditors; it must also be navigable by the people whose opportunities it affects.

9. Limitations

This review integrates a rapidly expanding and heterogeneous literature. Technologies labelled “AI recruitment” differ substantially in data, model type, autonomy, and stage of use, which limits direct comparison. Publication bias may favor novel harms or positive efficiency findings. Many empirical studies use hypothetical applicant scenarios rather than consequential hiring settings, while vendor systems are often proprietary and difficult to independently validate. Legal requirements also vary by jurisdiction and continue to evolve. Accordingly, this review does not claim exhaustive coverage or provide a PRISMA study-flow count; its conclusions should be interpreted as a structured critical synthesis of identifiable secondary evidence.

10. Conclusion

Artificial intelligence is changing recruitment decision-making by increasing the speed, scale, and formalization with which employers evaluate candidates. The evidence does not support a simple conclusion that AI makes hiring either objective or discriminatory. Its effects depend on what is predicted, how data were produced, whether constructs are job-relevant, how errors

are distributed, how recruiters interact with outputs, and whether candidates can understand and challenge consequential decisions. AI can reduce administrative burden and some forms of human inconsistency; it can also encode historical inequality, obscure responsibility, and make exclusion faster and harder to contest. The most defensible model is therefore neither full automation nor a return to unaided intuition. It is constrained human–AI augmentation: automate routine acquisition and processing, use validated models as decision support, preserve meaningful human authority for consequential judgments, and surround the entire process with auditability, accessibility, transparency, monitoring, and contestability. In recruitment, the question is not whether a human appears somewhere in the workflow. The question is whether the socio-technical system as a whole produces decisions that are valid, fair, explainable, and accountable.

11. Recommendations

The evidence supports a stronger policy position than simply advising employers to “use AI responsibly.” Organizations should adopt a presumption against autonomous exclusion in consequential recruitment decisions. AI may automate administrative tasks and support evidence processing, but an applicant should not be rejected solely because an opaque or weakly validated model assigns a low score. Where automation materially determines access to an interview, assessment stage, or job offer, the employer should be able to demonstrate job relevance, criterion and construct validity, acceptable subgroup performance, accessibility, and a meaningful route to human reconsideration. If those conditions cannot be demonstrated, the system should not be used for exclusion.

Validation should become a continuing organizational obligation rather than a document obtained once from a vendor. Employers should validate the system for the occupation, population, language, and context in which it is deployed and re-examine performance after material changes to the model, data, job, or labor market. A generic vendor accuracy figure is insufficient because recruitment validity is context-dependent. Monitoring should include overall selection rates, false-negative and false-positive patterns, intersectional subgroup outcomes where feasible and lawful, recruiter overrides, candidate complaints, accommodation requests, and model drift. A system that cannot be monitored at this level should be considered unsuitable for high-stakes selection.

Human oversight should be redesigned as a professional control function. The reviewer must understand intended use and limitations, have access to the underlying job-relevant evidence, receive uncertainty information rather than only a categorical recommendation, have enough time to conduct a genuine review, and possess clear authority to reverse the system. Organizations should audit reviewers as well as models. If overrides systematically benefit or disadvantage groups, if reviewers almost never disagree with the model, or if review times are implausibly short, those patterns should trigger investigation. The goal is demonstrably calibrated reliance, not maximum human discretion or maximum automation.

Employers should prohibit weakly justified inferential features in consequential decisions. Facial expression, vocal characteristics, personality inference from video, social-media traces, and other behavioral proxies should not be used merely because a vendor can generate a score. The evidential standard should rise as a feature becomes more remote from observable job behavior and more intrusive to privacy. Where a structured work sample, validated test, structured interview, or documented qualification measures the relevant construct more directly, the less intrusive and better validated method should be preferred.

Candidate rights should be designed into recruitment rather than added after complaints. Applicants should receive clear notice when AI materially contributes to evaluation, an intelligible description of the information assessed, a mechanism to correct inaccurate data, an accessible accommodation route, and a channel through which a consequential recommendation can be reconsidered by a competent person. These mechanisms should not require technical expertise or legal escalation. Transparency without a practical ability to correct or challenge a decision is disclosure, not accountability.

Procurement governance requires equally strong reform. Employers should not purchase high-stakes recruitment technology unless contracts provide validation evidence, intended and prohibited uses, relevant subgroup-testing information, data-protection terms, material-change notifications, incident cooperation, and audit rights. Procurement should involve HR, occupational or organizational psychology expertise, data protection, legal/compliance, information security, and the people who will operate the system. Commercial confidentiality should not prevent an employer from explaining or defending its own employment decisions. Organizations should separate efficiency from effectiveness. Reductions in time-to-screen, recruiter workload, or cost per application are useful operational outcomes, but they are not evidence that hiring decisions improved. Effectiveness evaluation should include predictive validity, cautiously interpreted quality and retention outcomes, diversity and subgroup error, applicant withdrawal and acceptance, accessibility, complaints, and the opportunity cost of excluding capable candidates. Senior management should receive both categories of evidence so that efficiency gains cannot conceal deterioration in fairness or decision quality.

For generative AI, organizations should assume that both sides of the labor market will increasingly use AI. Recruitment should shift away from signals that can be cheaply generated or optimized and toward evidence closely connected to job performance. Structured work samples, transparent skill demonstrations, verification of material claims, and well-designed structured interviews should carry more weight than stylistic qualities of application documents. Employers should distinguish deceptive manipulation from legitimate language, accessibility, or drafting assistance and avoid rules that punish candidates merely for using contemporary productivity tools.

Finally, governance should operate at the level of the complete recruitment system. A technically fair screening model cannot compensate for discriminatory sourcing, inaccessible assessments, poorly trained interviewers, or arbitrary final decisions. Every

organization deploying AI in hiring should maintain an accountable owner for the end-to-end process, an inventory of AI uses, risk classification by recruitment stage, periodic independent review, incident and complaint procedures, and a defined threshold for suspending a system. Where an organization cannot explain why the technology is necessary, what evidence establishes validity, how affected candidates can challenge errors, and who is accountable for failures, the appropriate recommendation is non-deployment until those questions are answered.

Declarations

Conflict of Interest

The author declares no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

No new datasets were generated or analysed in this study; the review is based exclusively on secondary sources.

Ethical Approval

Not applicable. This study is based exclusively on secondary sources and does not involve human participants or primary data collection.

Author Contributions

The author was responsible for the conceptualization, literature review, analysis, writing, and revision of the manuscript.

References

1. Acikgoz, Y., Davison, H.K., Compagnone, M. and Laske, M. (2020) 'Justice perceptions of artificial intelligence in selection', *International Journal of Selection and Assessment*, 28(4), pp. 399–416. doi:10.1111/ijsa.12306.
2. Barocas, S. and Selbst, A.D. (2016) 'Big data's disparate impact', *California Law Review*, 104(3), pp. 671–732. doi:10.15779/Z38BG31.
3. Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bannetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. and Herrera, F. (2020) 'Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI', *Information Fusion*, 58, pp. 82–115. doi:10.1016/j.inffus.2019.12.012.
4. Bertrand, M. and Mullainathan, S. (2004) 'Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination', *American Economic Review*, 94(4), pp. 991–1013. doi:10.1257/0002828042002561.
5. Black, J.S. and van Esch, P. (2020) 'AI-enabled recruiting: What is it and how should a manager use it?', *Business Horizons*, 63(2), pp. 215–226. doi:10.1016/j.bushor.2019.12.001.
6. Buolamwini, J. and Gebru, T. (2018) 'Gender shades: Intersectional accuracy disparities in commercial gender classification', *Proceedings of Machine Learning Research*, 81, pp. 77–91.
7. Chen, A., Han, F., Zhang, X. and Lu, Y. (2025) 'Cracking the AI recruitment code: Striving for transparency in finding the right person–job fit', *Information & Management*, 62(5), 104156. doi:10.1016/j.im.2025.104156.
8. Chen, Z. (2023) 'Ethics and discrimination in artificial intelligence-enabled recruitment practices', *Humanities and Social Sciences Communications*, 10, 567. doi:10.1057/s41599-023-02079-x.
9. Dadaboyev, S.M.U., Abdullayeva, J., Abbosova, N., Suleymenova, A. et al. (2025) 'Role of artificial intelligence in employee recruitment: Systematic review and future research directions', *Discover Global Society*, 3, 99. doi:10.1007/s44282-025-00246-w.
10. Dietvorst, B.J., Simmons, J.P. and Massey, C. (2015) 'Algorithm aversion: People erroneously avoid algorithms after seeing them err', *Journal of Experimental Psychology: General*, 144(1), pp. 114–126. doi:10.1037/xge0000033.
11. European Parliament and Council of the European Union (2024) 'Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)', *Official Journal of the European Union*.
12. Information Commissioner's Office (2026) *Recruitment rewired: An update on the ICO's work on the fair and responsible use of automation in recruitment*. Information Commissioner's Office.
13. Jarrahi, M.H. (2018) 'Artificial intelligence and the future of work: Human–AI symbiosis in organizational decision making', *Business Horizons*, 61(4), pp. 577–586. doi:10.1016/j.bushor.2018.03.007.
14. Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. and Mullainathan, S. (2018) 'Human decisions and machine predictions', *Quarterly Journal of Economics*, 133(1), pp. 237–293. doi:10.1093/qje/qjx032.
15. Köchling, A. and 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, pp. 795–848. doi:10.1007/s40685-020-00134-w.
16. Lavanchy, M., Reichert, P., Narayanan, J. and Savani, K. (2023) 'Applicants' fairness perceptions of algorithm-driven hiring procedures', *Journal of Business Ethics*, 188, pp. 125–150. doi:10.1007/s10551-022-05320-w.

17. Leicht-Deobald, U., Busch, T., Schank, C., Weibel, A., Schafheitle, S., Wildhaber, I. and Kasper, G. (2019) 'The challenges of algorithm-based HR decision-making for personal integrity', *Journal of Business Ethics*, 160(2), pp. 377–392. doi:10.1007/s10551-019-04204-w.
18. Logg, J.M., Minson, J.A. and Moore, D.A. (2019) 'Algorithm appreciation: People prefer algorithmic to human judgment', *Organizational Behavior and Human Decision Processes*, 151, pp. 90–103. doi:10.1016/j.obhdp.2018.12.005.
19. Marler, J.H. and Boudreau, J.W. (2017) 'An evidence-based review of HR Analytics', *International Journal of Human Resource Management*, 28(1), pp. 3–26. doi:10.1080/09585192.2016.1244699.
20. 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(6), Article 115. doi:10.1145/3457607.
21. Moritz, J.M., Pomrehn, L., Steinmetz, H. and Wehner, M.C. (2026) 'A meta-analysis on reactions to algorithmic decision-making in human resource management', *Human Resource Management Review*, 36(2), 101135. doi:10.1016/j.hrmr.2026.101135.
22. New York City Department of Consumer and Worker Protection (2023) *Automated Employment Decision Tools (AEDT): Local Law 144 of 2021 and implementing rules*. New York City Department of Consumer and Worker Protection.
23. Ologunoye, O.T., Adisa, T.A., Gbadamosi, G., Mordi, C. and Chang, K. (2026) 'The resourcing paradox: A systematic review of efficiency and effectiveness in AI-powered recruiting', *Employee Relations: The International Journal*, 48(5), pp. 873–892. doi:10.1108/ER-05-2025-0337.
24. Ore, O. and Sposato, M. (2022) 'Opportunities and risks of artificial intelligence in recruitment and selection', *International Journal of Organizational Analysis*, 30(6), pp. 1771–1782. doi:10.1108/IJOA-07-2020-2291.
25. Parasuraman, R., Sheridan, T.B. and Wickens, C.D. (2000) 'A model for types and levels of human interaction with automation', *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 30(3), pp. 286–297. doi:10.1109/3468.844354.
26. Puranam, P. (2021) 'Human–AI collaborative decision-making as an organization design problem', *Journal of Organization Design*, 10, pp. 75–80. doi:10.1007/s41469-021-00095-2.
27. 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*, pp. 469–481. doi:10.1145/3351095.3372828.
28. Sánchez-Monedero, J., Dencik, L. and Edwards, L. (2020) 'What does it mean to “solve” the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems', *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 458–468. doi:10.1145/3351095.3372849.
29. Schmidt, F.L. and 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), pp. 262–274. doi:10.1037/0033-2909.124.2.262.
30. 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*, pp. 59–68. doi:10.1145/3287560.3287598.
31. Tambe, P., Cappelli, P. and Yakubovich, V. (2019) 'Artificial intelligence in human resources management: Challenges and a path forward', *California Management Review*, 61(4), pp. 15–42. doi:10.1177/0008125619867910.
32. 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*. U.S. Equal Employment Opportunity Commission.
33. 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*. U.S. Equal Employment Opportunity Commission.
34. Woods, S.A., Ahmed, S., Nikolaou, I., Costa, A.C. and Anderson, N.R. (2020) 'Personnel selection in the digital age: A review of validity and applicant reactions, and future research challenges', *European Journal of Work and Organizational Psychology*, 29(1), pp. 64–77. doi:10.1080/1359432X.2019.1681401.



Published by Research and Analysis Journals Journal, this is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026 _