International Journal of Research and Applied Innovations (IJRAI)

| ISSN: 2455-1864 | www.ijrai.org | editor@ijrai.org | A Bimonthly, Scholarly and Peer-Reviewed Journal |

||Volume 8, Issue 6, November–December 2025||

DOI:10.15662/IJRAI.2025.0806023

The Impact of Structured Validation and Audit Frameworks on the Fairness and Efficiency of AI-Driven Hiring Systems

Safeer Ahmad

MS Industrial and Organizational Psychology from Missouri State University Department of Psychology and SHRM-CP, USA

ABSTRACT: Organizations increasingly use artificial intelligence (AI) to screen and rank job applicants, yet these systems can produce disparate outcomes across protected groups and may add operational friction when they require manual review or compliance documentation. Prior work has proposed algorithmic auditing and assurance frameworks, but empirical evidence linking audit intensity to both fairness and hiring efficiency remains limited within the same operational context. Building on internal algorithmic auditing guidance (Raji et al., 2020) and audit systematization in recruitment (Kazim et al., 2021), this study evaluates whether structured validation and audit frameworks are associated with improved fairness and reduced time to hire. Using an applicant-tracking-system dataset from a logistics supply-chain employer (N = 1,906 active applicants; n = 400 hires), we compared a pre-AI manual baseline (2023) with two 2024 AI screening configurations: a compliance-only audit and an assurance-level audit. Fairness was operationalized with adverse impact ratios (AIRs; UGESP, 1978), and efficiency was operationalized as time to hire (days). Results indicated that assurance-level auditing coincided with substantial improvements in AIRs (e.g., minority AIR increased from 0.12 in the manual baseline to 0.85 under assurance) while also reducing time to hire (M difference = 12.87 days relative to the baseline). Logistic and linear models controlling for job family supported these patterns. Findings suggest that structured, higher-intensity audit frameworks can be associated with simultaneous gains in fairness and process efficiency, warranting replication in multi-site, longitudinal studies under emerging audit mandates (e.g., NYC Local Law 144).

KEYWORDS: algorithmic hiring, AI audit, adverse impact, AI assurance, time to hire, recruitment analytics

I. INTRODUCTION

AI-enabled hiring tools are increasingly used to screen resumes, rank candidates, and support employment decision-making. Although automation is often adopted to improve speed and consistency, research has documented that algorithmic screening can reproduce or amplify inequities when training data, feature choices, or deployment practices encode structural disparities (Raghavan et al., 2020). In response, scholars have called for systematic auditing approaches that document design decisions, risk assumptions, and failure modes across the full system lifecycle (Raji et al., 2020). Parallel governance efforts, such as the NIST AI Risk Management Framework (NIST, 2023) and AI management system standards (ISO/IEC 42001:2023), emphasize that trustworthy AI requires proactive risk assessment, monitoring, and accountability mechanisms rather than ad hoc post hoc checks.

In employment contexts, fairness concerns are not only ethical but also legally salient. The Uniform Guidelines on Employee Selection Procedures (UGESP, 1978) describe how employers should evaluate whether selection procedures create adverse impact, including the four-fifths rule as a practical screening criterion. More recently, regulatory efforts such as New York City’s Local Law 144 have mandated independent bias audits for certain automated employment decision tools (AEDTs), shifting many organizations from voluntary auditing toward compliance-driven evaluation. However, early analyses of audit disclosures suggest variability in rigor, method choice, and interpretability, raising questions about whether compliance-only audits are sufficient to meaningfully reduce disparate outcomes in practice (Filippi et al., 2023; Groves et al., 2024; Wright et al., 2024).

A central unresolved question is whether structured validation and audit frameworks can deliver measurable fairness gains without eroding the efficiency benefits that motivate AI adoption. Existing literature often treats fairness and performance as a trade-off problem (Kleinberg et al., 2017), and much prior work focuses on technical metrics rather than operational outcomes such as time to hire or process throughput. Further, while recruitment audit frameworks have been proposed and systematized (Kazim et al., 2021), empirical studies rarely compare different audit intensities within the same organizational setting. The present study addresses this gap by evaluating two levels of structured auditing i-e, compliance-only versus assurance-level implemented alongside AI screening within a logistics supply-chain employer, and by quantifying associations with both fairness (adverse impact ratios) and efficiency (time to hire).

II. LITERATURE REVIEW

Research on algorithmic hiring and auditing has developed along three complementary streams: (a) foundational arguments for internal and external auditing to close accountability gaps, (b) formalization of audit and assurance processes specific to recruitment systems, and (c) regulatory-driven implementation and critique of bias audit regimes. First, foundational work emphasizes that auditing must be end-to-end and documentation-based to be actionable. Raji et al. (2020) proposed an internal algorithmic auditing framework spanning data collection, model development, testing, deployment, and monitoring, with stage-specific artifacts that enable traceability and accountability. Field scans of the auditing ecosystem further note that the audit market is heterogeneous and that audit quality depends on clear standards, independence, and transparent reporting practices (Costanza-Chock et al., 2022). Complementing these governance perspectives, technical work has clarified that commonly used statistical fairness criteria can be mutually incompatible except under restrictive conditions, implying that audit programs must specify which fairness definitions and legal standards they prioritize (Kleinberg et al., 2017). Toolkits such as AI Fairness 360 operationalize these concepts by providing bias metrics and mitigation methods across the machine-learning pipeline (Bellamy et al., 2018), but open-source tools do not automatically translate into scalable organizational governance without process integration.

Second, recruitment-specific frameworks describe how auditing can be operationalized within hiring workflows. Kazim et al. (2021) systematized audit stages for algorithmic recruitment (e.g., purpose definition, data and feature review, validation, fairness testing, and monitoring), aligning technical checks with organizational governance. Case-based evidence also illustrates how organizations may combine fairness constraints with performance requirements; for example, Wilson et al. (2021) described a candidate-screening system that incorporated fairness evaluation and external review as part of product development. Nevertheless, much of this work emphasizes conceptual assurance processes rather than quantifying whether different levels of audit rigor change operational hiring outcomes such as time to hire.

Third, emerging regulation has increased the salience of audit design choices. New York City’s Local Law 144 requires annual independent audits and candidate notifications for covered AEDTs, yet analyses of the early audit regime highlight concerns about definitional scope, methodological variability, and the risk that compliance may prioritize minimum reporting over substantive fairness improvement (Filippi et al., 2023; Groves et al., 2024; Wright et al., 2024). These critiques suggest that compliance-only audits may not adequately address deeper sources of bias (e.g., construct validity, proxy features, and organizational process effects) and motivate more comprehensive assurance-level approaches aligned with risk management frameworks (NIST, 2023).

Across these streams, a key empirical gap remains: few studies quantify how audit intensity relates to both fairness outcomes and efficiency outcomes within the same operational hiring context. The current study addresses this gap by comparing a manual baseline with two audit-intensity conditions applied to AI screening, using adverse impact ratios (UGESP, 1978) and time to hire as concrete, measurable outcomes.

Hypotheses

Hypothesis 1 (Fairness). Compared with a compliance-only bias audit, an assurance-level validation and audit framework will be associated with higher adverse impact ratios (AIRs; i.e., ratios closer to 1.00) for protected groups, indicating reduced disparity in selection outcomes.

Hypothesis 2 (Efficiency). Compared with the compliance-only bias audit and the manual baseline, an assurance-level validation and audit framework will be associated with shorter time to hire (days) among hired candidates.

III. METHOD

Participants

Data were drawn from a single logistics supply-chain organization’s applicant tracking system (ATS). To support confidentiality and reproducibility, the analytic dataset is a de-identified, distribution-preserving replica of the organization’s ATS extract. Applicants were included if they applied to one of three focal job families (Warehouse, Logistics Coordinator, or Driver) during the study windows and had complete demographic and outcome data. Applicants who withdrew before the final decision were excluded from fairness analyses but were retained for descriptive attrition reporting. The final analytic sample for selection outcomes included N = 1,906 active applicants; the efficiency analysis included n = 400 hires with observed time-to-hire values.

Design: The present study employed a **quasi-experimental, nonequivalent-groups comparative design** to evaluate differences in fairness and efficiency across three hiring conditions within a single logistics supply chain organization (Manual23, AI24-Comp, and AI24-Assur).

Because conditions were **not randomly assigned**, the observed differences across Manual23, AI24-Comp, and AI24-Assur may reflect, in part, **rival explanations** associated with time and context rather than the audit frameworks alone. In particular, **history effects** (e.g., changes in labor market conditions, applicant availability, organizational policy, or recruiting resources across periods) and **maturation/implementation effects** (e.g., increasing recruiter familiarity with the AI workflow, process refinements unrelated to auditing, or vendor model updates) could influence both time-to-hire and selection outcomes. In addition, **selection and job-mix confounding** may occur if the distribution of job families, locations, requisition urgency, or qualification requirements differed across conditions, which could create aggregate differences in selection rates and AIR even when within-job processes are stable (i.e., an aggregation/stratification risk). Finally, **instrumentation effects** are possible if operational definitions or data capture procedures (e.g., timestamps used to compute time-to-hire, recording of hiring outcomes, or demographic self-report completeness) changed between conditions. To mitigate these threats, the study reports applicant pool composition by condition and recommends stratified or covariate-adjusted analyses (e.g., job family, location, and seasonality) when such variables are available; nevertheless, causal conclusions should be interpreted cautiously.

Materials

Audit conditions: Three screening conditions were defined. (a) Manual baseline (2023): resume screening and progression decisions were made without AI decision support. (b) AI + compliance-only audit (2024): AI resume screening was used alongside a compliance-oriented bias audit focused on group selection-rate monitoring and required disclosures. (c) AI + assurance-level audit (2024): AI resume screening was used alongside an expanded audit framework that included data quality checks, documented validation rationale aligned with UGESP principles, subgroup fairness testing, and post-deployment monitoring. The compliance-only and assurance-level configurations were applied to different requisitions in 2024 based on a pre-specified risk tiering of job families and tool-use contexts.

Measures: Fairness was operationalized using adverse impact ratios (AIRs), defined as the selection rate for a protected group divided by the selection rate for a reference group, with the four-fifths rule (AIR < 0.80) used as a practical threshold for potential adverse impact (UGESP, 1978). Efficiency was operationalized as time to hire, defined as the number of days between application date and offer acceptance date among hired candidates.

Procedure

The ATS provided applicant-level records including job family, screening condition, demographic indicators (gender; protected-minority status), hire outcome, withdrawal status, and when applicable, time to hire. Analyses proceeded in four stages. First, descriptive statistics characterized the applicant pools by condition. Second, efficiency outcomes were evaluated using one-way analysis of variance (ANOVA) and Welch’s t tests for planned comparisons, supported by diagnostic checks (Q–Q plot) and heteroskedasticity-robust standard errors. Third, fairness outcomes were evaluated using selection rates and AIRs by condition, supplemented by chi-square tests of association between demographic group membership and hire outcomes within each condition. Fourth, to reduce confounding from job-family composition differences across conditions, generalized linear models (logistic regression) predicted hiring outcomes from condition, demographic status, and their interaction while controlling for job family; linear regression predicted time to hire from condition while controlling for job family.

IV. RESULTS

Applicant pool characteristics are summarized in Tables 1–2. Condition sample sizes were 949 (manual baseline), 484 (AI + compliance-only), and 473 (AI + assurance-level).

Table 1
Active Applicant Demographics by Condition (Gender)

Condition	Male, n (%)	Female, n (%)	Nonbinary, n (%)	Total, n
Manual23	639 (67.3%)	298 (31.4%)	12 (1.3%)	949
AI24-Comp	315 (65.1%)	167 (34.5%)	2 (0.4%)	484
AI24-Assur	321 (67.9%)	148 (31.3%)	4 (0.8%)	473

Note. Percentages are within condition; applicants who withdrew prior to a hiring decision were excluded from the analytic sample for selection outcomes.

Table 2
Active Applicant Demographics by Condition (Protected-Minority Status)

Condition	Nonminority, n (%)	Protected minority, n (%)	Total, n
Manual23	558 (58.8%)	391 (41.2%)	949
AI24-Comp	277 (57.2%)	207 (42.8%)	484
AI24-Assur	265 (56.0%)	208 (44.0%)	473

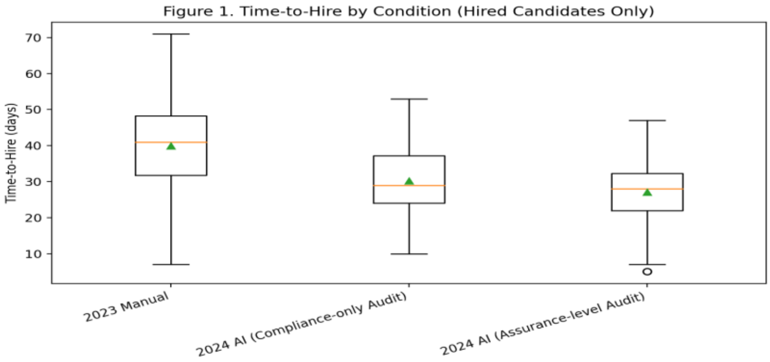
Time to hire differed across conditions (Table 3; Figure 1). A one-way ANOVA indicated a significant condition effect on time to hire, $F(2, 397) = 58.33$, $p < .001$, $\eta^2 = 0.227$. Planned Welch's *t* tests showed that hires under the compliance-only audit were faster than the manual baseline (*M* difference = 9.71 days, 95% CI [7.21, 12.21]), $t(249.06) = 7.65$, $p < .001$, $d = 0.86$. Hires under the assurance-level audit were also faster than the manual baseline (*M* difference = 12.87 days, 95% CI [10.50, 15.24]), $t(268.36) = 10.70$, $p < .001$, $d = 1.16$. A Q–Q plot of residuals suggested approximate normality for time-to-hire values (Figure 5); robust standard errors were used to address potential heteroskedasticity.

Table 3
Time to Hire (Days) Among Hired Candidates by Condition

Condition	n	M	SD
Manual23	200	39.97	12.16
AI24-Comp	100	30.26	9.34
AI24-Assur	100	27.10	8.41

Figure 1

Time to Hire by Screening Condition



Fairness outcomes are summarized using selection rates and adverse impact ratios (AIRs; Tables 4–5; Figures 2–4). For gender, the female-to-male AIR increased from 0.19 in the manual baseline to 0.29 under the compliance-only audit and to 0.66 under the assurance-level audit (Table 4; Figure 4A). Chi-square tests showed a strong association between gender and hiring outcomes in the manual baseline, $\chi^2(1) = 61.86$, $p < .001$, $\varphi = 0.257$. The association remained significant under the compliance-only audit, $\chi^2(1) = 24.29$, $p < .001$, $\varphi = 0.224$, but was not statistically significant under the assurance-level audit, $\chi^2(1) = 3.55$, $p = .059$, $\varphi = 0.087$.

Table 4

Selection Rates and Adverse Impact Ratio (AIR) by Gender and Condition

Condition	Female n	Female hired	Female SR	Male n	Male hired	Male SR	AIR (F/M)
Manual23	298	16	0.054	639	179	0.280	0.192
AI24-Comp	167	13	0.078	315	86	0.273	0.285
AI24-Assur	148	23	0.155	321	76	0.237	0.656

Note. SR = selection rate. AIR values closer to 1.00 indicate smaller disparities; AIR < 0.80 may indicate potential adverse impact (UGESP, 1978).

Figure 2

Gender Selection Rates by Condition

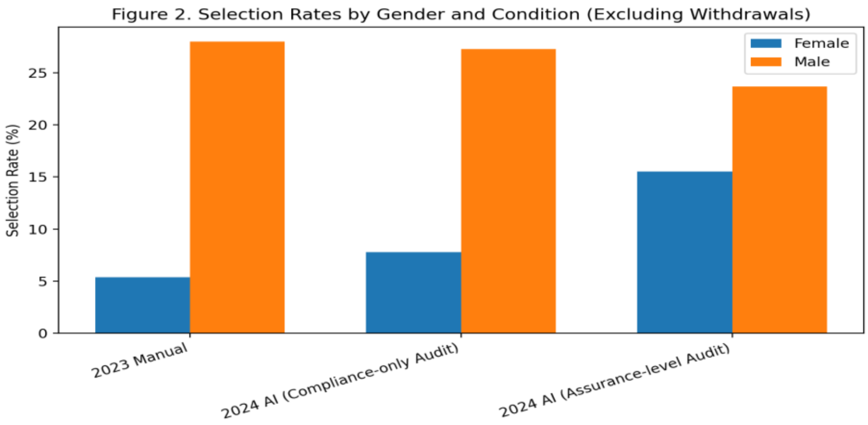
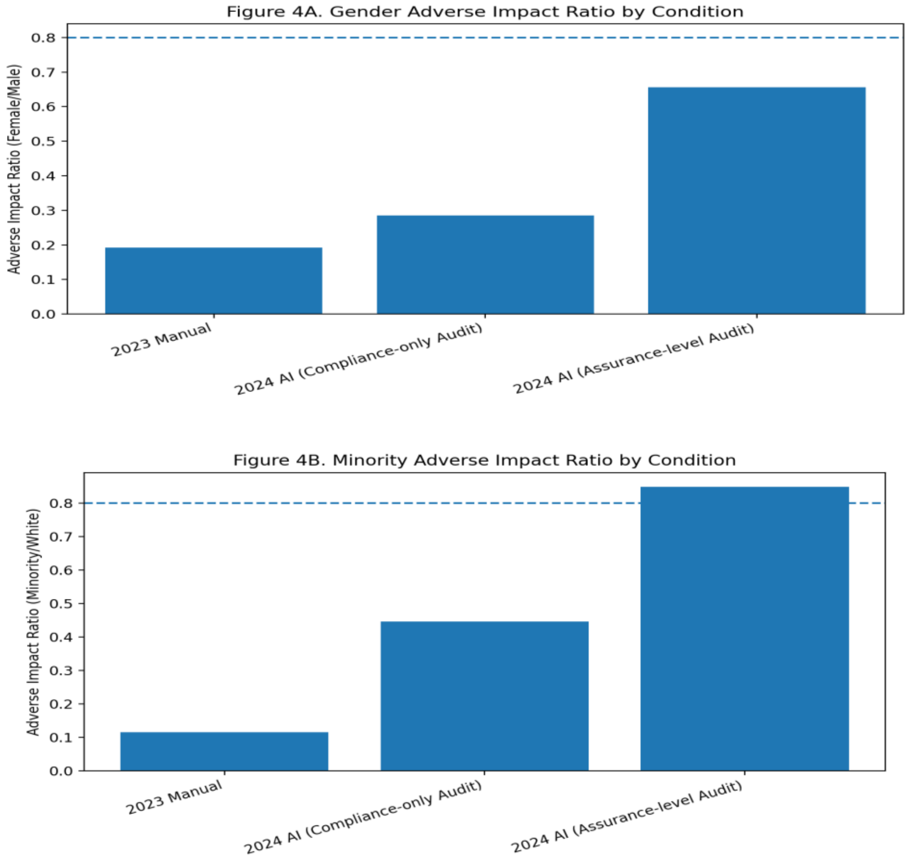


Figure 4

Adverse Impact Ratios by Condition



Note. Panel A shows female-to-male AIR; Panel B shows protected-minority to nonminority AIR.

For protected-minority status, the minority-to-nonminority AIR increased from 0.12 in the manual baseline to 0.45 under the compliance-only audit and to 0.85 under the assurance-level audit (Table 5; Figure 4B). Chi-square tests indicated a strong association between protected-minority status and hiring outcomes in the manual baseline, $\chi^2(1) = 117.05$, $p < .001$, $\varphi = 0.351$, and a smaller but significant association under the compliance-only audit, $\chi^2(1) = 15.36$, $p < .001$, $\varphi = 0.178$. Under the assurance-level audit, the association was not statistically significant, $\chi^2(1) = 0.62$, $p = .431$, $\varphi = 0.036$.

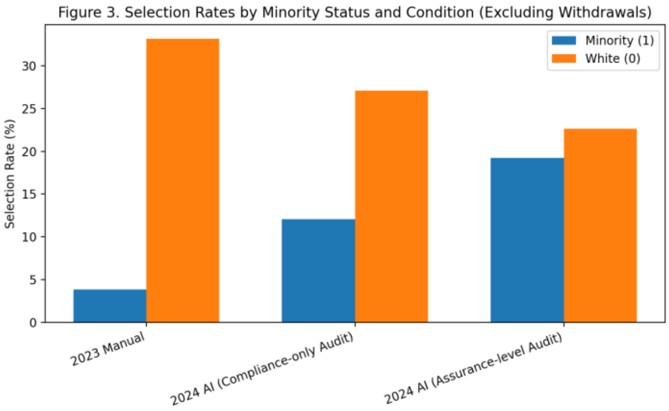
Table 5

Selection Rates and Adverse Impact Ratio (AIR) by Protected-Minority Status and Condition

Condition	Minority n	Minority hired	Minority SR	Nonmin n	Nonmin hired	Nonmin SR	AIR (Min/Non)
Manual23	391	15	0.038	558	185	0.332	0.116
AI24-Comp	207	25	0.121	277	75	0.271	0.446
AI24-Assur	208	40	0.192	265	60	0.226	0.849

Figure 3

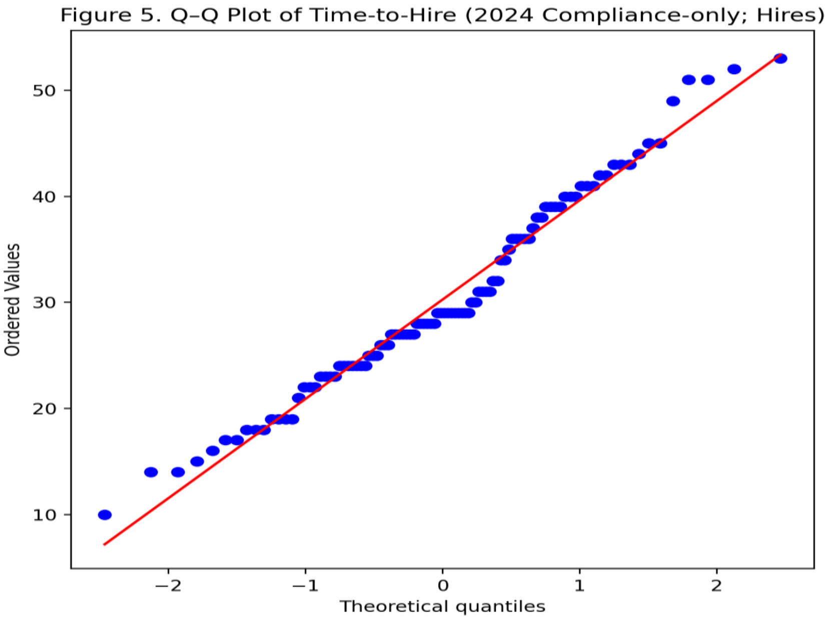
Protected-Minority Selection Rates by Condition



To adjust for differences in job-family composition across conditions, logistic regression models predicted hiring outcomes from condition, demographic status, and their interaction while controlling for job family. In the gender model (excluding nonbinary cases due to small cell sizes), the female disadvantage observed in the manual baseline was large (OR = 0.108, 95% CI [0.061, 0.190], $p < .001$). The assurance-level audit significantly reduced the gender disparity, as indicated by a positive condition-by-female interaction (OR = 5.05, 95% CI [2.24, 11.41], $p < .001$). In the protected-minority model, a strong minority disadvantage in the manual baseline (OR = 0.035, 95% CI [0.019, 0.064], $p < .001$) was substantially attenuated under both audits, with the largest reduction under assurance (interaction OR = 20.81, 95% CI [9.23, 46.92], $p < .001$). A linear regression model for time to hire controlling for job family yielded similar conclusions: relative to the manual baseline, compliance-only auditing was associated with 9.64 fewer days ($p < .001$) and assurance-level auditing with 12.86 fewer days ($p < .001$).

Figure 5

Normal Q–Q Plot of Time-to-Hire Values (Compliance-Only Condition)



V. DISCUSSION

This study examined whether structured validation and audit frameworks used with AI-based resume screening were associated with improved fairness and efficiency in hiring. Across both protected-minority status and gender, adverse impact ratios (AIRs) improved monotonically as audit intensity increased: the compliance-only audit produced meaningful gains relative to the manual baseline, and the assurance-level audit produced the largest reductions in disparity. At the same time, time to hire decreased substantially under both audit conditions, with the largest reduction under assurance. The pattern is notable because debates about algorithmic fairness often emphasize trade-offs between fairness constraints and predictive or operational performance (Kleinberg et al., 2017). In this dataset, higher audit intensity coincided with both improved fairness and faster hiring, suggesting that governance interventions can target process inefficiencies (e.g., rework, escalations, and uncertainty about model use) while also improving equity. The findings align with end-to-end auditing arguments that emphasize early-stage documentation and lifecycle monitoring as mechanisms for reducing downstream harm (Raji et al., 2020). In practical terms, assurance-level auditing likely involved more than reporting selection-rate ratios. Recruitment audit systematizations emphasize that audits should evaluate the construct being measured, data quality, proxy features, and decision thresholds, and should specify a monitoring plan for drift and emerging bias (Kazim et al., 2021). Such activities can plausibly improve both fairness and efficiency by reducing false negatives for qualified candidates in protected groups and by stabilizing decision thresholds that otherwise require manual override. The results also complement case-based evidence that organizations can embed fairness evaluation into candidate-screening product development without sacrificing usability (Wilson et al., 2021).

These results are relevant to current policy debates about bias-audit mandates. Analyses of NYC Local Law 144 audit disclosures suggest that minimum-compliance reporting may vary widely in rigor and may not adequately address deeper design logics, model access, or process context (Filippi et al., 2023; Groves et al., 2024; Wright et al., 2024). The present findings support the view that audit programs that extend beyond minimal selection-rate reporting—toward assurance practices that incorporate validation logic and monitoring—may offer more robust fairness improvements. Importantly, the observed efficiency gains suggest that assurance-level audits need not be interpreted solely as compliance overhead; they may function as process-improvement mechanisms that reduce cycle time. From a measurement perspective, using AIRs provides an interpretable bridge between statistical fairness evaluation and employment-law practice (UGESP, 1978). However, AIRs do not identify which stage(s) of the hiring pipeline generate disparities, nor do they address other fairness constructs (e.g., calibration or equal opportunity). Future studies should triangulate AIRs with additional fairness metrics and pipeline-stage analyses to identify where audit interventions are most effective.

Limitations

Several limitations temper interpretation. First, the design is observational and based on a single organization; audit configurations were not randomly assigned, so unobserved differences between requisitions may contribute to observed effects. Although models controlled for job family, other contextual factors (e.g., labor market conditions, recruiter staffing, or requisition urgency) were not measured. Second, demographic analyses were limited to gender and a binary protected-minority indicator; small subgroup sizes (e.g., nonbinary candidates) limited the ability to assess intersectional effects. Third, fairness was evaluated using selection outcomes rather than downstream job performance or retention; thus, the study does not address whether audit frameworks improve predictive validity or long-term workforce outcomes. Fourth, time to hire captures only one dimension of operational efficiency; additional metrics (e.g., cost per hire, recruiter workload, candidate experience, and quality of hire) are needed for a complete cost–benefit assessment.

Future Directions

Future research should replicate these analyses using multi-site, multi-year datasets and quasi-experimental or randomized rollouts of audit frameworks to strengthen causal inference. A priority is longitudinal evaluation that jointly models diversity outcomes and operational KPIs to estimate the net benefits of audit maturity over time. Studies should also examine how audit frameworks interact with emerging AI modalities used in hiring (e.g., large language models for resume summarization and interview assistance), including audits that explicitly quantify trade-offs between fairness interventions and system performance. Finally, standard-setting work should integrate employment-law concepts of validation and adverse impact with AI governance frameworks (NIST, 2023; UGESP, 1978), to specify minimum evidence requirements for assurance-level audits that are meaningful, comparable, and scalable.

DOI:10.15662/IJRAI.2025.0806023

VI. CONCLUSION

Within the limits of a single-organization observational design, higher-intensity structured auditing was associated with improved fairness measured via adverse impact ratios and reduced time to hire. The results support the premise that assurance-level audit frameworks may contribute to both equitable and efficient AI-enabled hiring, and they motivate rigorous replication under real-world audit mandates.

REFERENCES

1. 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. (2018). AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. arXiv. https://arxiv.org/abs/1810.01943
2. Costanza-Chock, S., Raji, I. D., & Buolamwini, J. (2022). Who audits the auditors? Recommendations from a field scan of the algorithmic auditing ecosystem. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22). Association for Computing Machinery. https://doi.org/10.1145/3531146.3533213
3. Filippi, G., Zannone, S., Hilliard, A., & Koshiyama, A. S. (2023). Local Law 144: A critical analysis of regression metrics. arXiv. https://arxiv.org/abs/2302.04119
4. Groves, L., Metcalf, J., Kennedy, A., Vecchione, B., & Strait, A. (2024). Auditing work: Exploring the New York City algorithmic bias audit regime. In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24). Association for Computing Machinery. https://doi.org/10.1145/3630106.3658959
5. International Organization for Standardization & International Electrotechnical Commission. (2023). Information technology—Artificial intelligence—Management system (ISO/IEC 42001:2023). ISO.
6. Kazim, E., Koshiyama, A. S., Hilliard, A., & Polle, R. (2021). Systematizing audit in algorithmic recruitment. Journal of Intelligence, 9(3), 46. https://doi.org/10.3390/jintelligence9030046
7. Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. In Proceedings of the 8th Innovations in Theoretical Computer Science Conference (ITCS 2017) (Article 43). Schloss Dagstuhl–Leibniz-Zentrum für Informatik. https://doi.org/10.4230/LIPIcs.ITCS.2017.43
8. 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
9. New York City Department of Consumer and Worker Protection. (n.d.). Automated employment decision tools (AEDT). Retrieved December 29, 2025, from https://www.nyc.gov/site/dca/about/automated-employment-decision-tools.page
10. 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 (FAccT '20) (pp. 469–481). Association for Computing Machinery. https://doi.org/10.1145/3351095.3372828
11. Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAccT '20) (pp. 33–44). Association for Computing Machinery. https://doi.org/10.1145/3351095.3372873
12. U.S. Equal Employment Opportunity Commission, U.S. Department of Labor, U.S. Department of Justice, & U.S. Civil Service Commission. (1978). Uniform guidelines on employee selection procedures (29 C.F.R. § 1607). https://www.ecfr.gov/current/title-29/subtitle-B/chapter-XIV/part-1607
13. Wilson, C., Ghosh, A., Jiang, S., Mislove, A., Baker, L., Szary, J., Trindel, K., & Polli, F. (2021). Building and auditing fair algorithms: A case study in candidate screening. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445928
14. Wright, R., Hilliard, A., Koshiyama, A. S., & Filippi, G. (2024). Null compliance? Measuring bias audits under NYC Local Law 144. arXiv. https://arxiv.org/abs/2406.01399

International Journal of Research and Applied Innovations (IJRAI)

| ISSN: 2455-1864 | www.ijrai.org | editor@ijrai.org | A Bimonthly, Scholarly and Peer-Reviewed Journal |

||Volume 8, Issue 6, November–December 2025||

DOI:10.15662/IJRAI.2025.0806023

Appendix A

Audit Framework Operationalization Matrix

This matrix defines, at an implementation level, what compliance-only auditing (AI24-Comp) versus assurance-level auditing (AI24-Assur) meant in this study. The purpose is to make the audit-intensity independent variable replicable and falsifiable.

For each control area, the organization should document whether the control is implemented and retain the listed evidence artifacts. AI24-Assur requires documented completion of all assurance-level items for the defined release cycle. AI24-Comp is satisfied when the compliance-only items are completed.

Table A1

Audit Framework Operationalization Matrix (Compliance-Only vs Assurance-Level)

Control area	**AI24-Comp (Compliance-only audit)**	**AI24-Assur (Assurance-level audit)**	**Minimum evidence artifacts (retain)**
Audit scope and tool definition	Tool classified as AI screening/AEDT; scope limited to required selection-rate comparisons.	Formal scope statement includes system boundary, model(s), vendor modules, human-in-the-loop points, and downstream workflow impacts.	Audit scope memo; system boundary diagram; process map.
Governance and accountability	Named business owner; documented compliance responsibility.	Named owner plus independent reviewer; RACI for data/model/HR/legal; sign-off gate before deployment.	RACI chart; sign-off form; meeting minutes.
Data provenance and documentation	Basic dataset description (fields used, timeframe).	Data lineage documented; representativeness checks; missingness analysis; demographic self-report handling plan.	Data sheet; lineage map; missingness report.
Job-relatedness linkage	Statement of job relevance for major predictors.	Documented job analysis or competency mapping linking predictors to job requirements (validity rationale).	Job analysis summary; predictor-to-competency mapping.
Model documentation	High-level model description and intended use.	Model card with training/evaluation data, limitations, subgroup performance, known risks, intended users; version log.	Model card; version log.
Baseline performance evaluation	Overall performance check (business KPI or vendor metric).	Performance and stability evaluation across job families/roles (or justified single-role scope); documented thresholds.	Performance report; threshold rationale; stability diagnostics.
Fairness metrics	AIR for gender and minority status computed and reported.	AIR plus at least one supporting metric (e.g., subgroup selection-rate differences; error-rate parity if scores available); intersectional check if feasible.	Fairness dashboard; metric definitions; subgroup tables.
Pre-deployment remediation	If AIR is concerning, recommended actions documented.	If AIR < .80, remediation required before release or formal risk acceptance with mitigation plan.	Remediation log; risk acceptance form.
Decision thresholds and override	Threshold documented; basic guidance to recruiters.	Threshold and rationale documented; override policy; override logging and periodic review.	Threshold memo; override policy; override audit log.

Transparency and candidate notice	Required notices (if applicable) and internal disclosure.	Candidate-facing notice and internal transparency (tool capabilities/limits); accommodation contact path.	Notice template; communications record.
Monitoring cadence	Periodic reporting (e.g., quarterly).	Regular monitoring schedule (e.g., monthly) including drift triggers and fairness re-checks.	Monitoring plan; drift triggers; monthly outputs.
Change management	Informal tracking of changes.	Formal change control: versioning, rollback plan, release notes, re-validation triggers.	Change tickets; release notes; rollback plan.
Incident response	Escalation path identified.	Incident playbook for bias/quality issues; severity levels; post-incident review.	Incident playbook; incident tickets; post-incident review template.
Independence and assurance	Conducted internally for compliance.	Independence via separate reviewer function and/or external validation of methodology when feasible.	Reviewer attestation; external review letter (if used).
Documentation completeness	Required compliance artifacts retained.	Full audit trail retained for replication and regulator/third-party review.	Audit binder/index; artifact checklist.

Note. AI24-Comp indicates completion of compliance-only controls. AI24-Assur indicates completion of assurance-level controls with documentation retained for auditability. AEDT = automated employment decision tool; AIR = adverse impact ratio.

Appendix B

Measures and Operational Definitions

This appendix defines study variables, coding rules, and the computations used for fairness and efficiency outcomes.

Independent Variable

Audit framework condition (categorical; three levels). Manual23 = 2023 manual screening workflow; AI24-Comp = 2024 AI screening with compliance-only auditing; AI24-Assur = 2024 AI screening with assurance-level auditing (see Appendix A). Recommended coding for analysis: Manual23 = 0, AI24-Comp = 1, AI24-Assur = 2.

Dependent Variables

Fairness: Adverse impact ratio (AIR)

AIR is a group-level screening indicator of potential adverse impact. It is computed as the protected group's selection rate divided by the reference group's selection rate (Equal Employment Opportunity Commission et al., 1978).

SR_g = (Number hired in group g) / (Number of active applicants in group g)

Gender AIR uses female as the protected group and male as the reference group.

AIR_gender = SR_female / SR_male

Minority-status AIR uses minority applicants as the protected group and White applicants as the reference group.

AIR_minority = SR_minority / SR_white

Reporting convention: AIR values below .80 are commonly flagged for review under the four-fifths rule heuristic (Equal Employment Opportunity Commission et al., 1978). AIR was computed using active applicants only (withdrawn applicants excluded).

Efficiency: Time-to-hire (days)

Time-to-hire is defined as the number of calendar days from application submission to offer acceptance/hire decision for candidates who were hired (hired = 1).

Time-to-hire = Date(offer accepted/hire decision) - Date(application submitted)

Time-to-hire analyses used hires only (hired = 1). All timestamps should be recorded consistently across conditions.

Supporting and Derived Variables
Hiring outcome (binary): 1 = hired; 0 = not hired.
Active applicant indicator: active applicants remain in the selection process; withdrawn applicants are excluded from AIR computations.

Demographic coding: gender categories are female and male for AIR computations; nonbinary applicants should be reported descriptively when cell sizes are too small for stable ratio estimates. Minority status is coded as 1 = minority and 0 = White.

Statistical Tests and Effect Sizes
Time-to-hire comparisons use Welch's t tests for pairwise comparisons across conditions to reduce sensitivity to unequal variances, with Hedges' g reported as an effect size. Hiring-outcome differences by group (gender and minority status) are examined using chi-square tests of independence within condition, with Cramer's V reported as an association effect size. Normality and outliers for time-to-hire are assessed using Shapiro-Wilk tests, skewness/kurtosis summaries, and Q-Q plots.

Inclusion and Exclusion Criteria
Included: applicants to the focal organization's defined roles during the study windows with valid hiring outcomes recorded; hires with valid timestamps for time-to-hire.

Excluded: withdrawn applicants for AIR/selection-rate analyses; cases with missing or invalid timestamps for time-to-hire; subgroup cells too small for stable AIR estimation (reported descriptively).