Bias Investigation Report – HireScore AI

TalentMatch AI | QA Bias Audit

Executive Summary

The bias investigation of HireScore reveals a **high bias severity**. Multiple bias types are present across the data lifecycle, resulting in systematic disadvantage to female candidates, applicants from non-elite universities, and candidates originating from regions outside Greater Accra and Ashanti. These biases originate from historically skewed hiring data and are amplified by feature selection and model training decisions.

Primary Affected Groups:

1. Female candidates
2. Applicants from non-elite universities
3. Candidates from Northern and other non-dominant regions

Top 3 Recommended Actions:

1. Remove or neutralize proxy features such as university attended, location, and LinkedIn connections.
2. Introduce fairness constraints and rebalance training data.
3. Implement continuous bias monitoring with human review for top-ranked candidates.

Detailed Findings

1. Bias Type Analysis

Historical Bias – Present: Yes
Evidence shows that historical hiring favored males (72%), elite universities, and candidates from Accra and Ashanti. These imbalances are directly reflected in the training dataset and reproduced in current model outputs.

Sampling Bias – Present: Yes
The dataset underrepresents female candidates, non-elite institutions, and applicants from other regions. This lack of representativeness limits the model’s ability to fairly evaluate the full applicant population.

Measurement Bias – Present: Yes
Features such as LinkedIn connections, professional references, and prior company names favor candidates with greater access to professional networks, which are unevenly distributed across gender and region.

Proxy Bias – Present: Yes
University attended, location of employment, LinkedIn connections, and age act as proxies for socioeconomic status, gender, and regional background, indirectly disadvantaging protected groups.

2. Bias Pipeline Mapping

Bias Point #1: Skewed historical hiring records collected without correction.
Bias Point #2: Selection of features closely correlated with protected attributes.
Bias Point #3: Optimization purely for predictive accuracy without fairness constraints.
Bias Point #4: Deployment without post-deployment bias monitoring or human checks.

3. Feature Analysis

University attended, LinkedIn connections, location, and age significantly influence scores and reinforce existing social inequalities. These features disproportionately raise scores for male, Accra-based, and elite-university candidates.

Mitigation Plan

Immediate Actions (This Week)

1. Remove university ranking, LinkedIn connections, and location-based features.
2. Apply temporary score normalization across gender and region.
3. Track weekly metrics: demographic parity ratio, score distribution by group.

Short-Term Actions (1–3 Months)

1. Collect balanced training data from underrepresented regions and institutions.
2. Retrain models using fairness-aware learning techniques.
3. Add mandatory human review for top 5–10% ranked candidates.

Long-Term Actions (6–12 Months)

1. Adopt Equal Opportunity as the primary fairness metric.
2. Require clients to use HireScore as a decision-support tool, not an automated gatekeeper.
3. Disclose scoring factors, fairness audits, and appeal mechanisms to applicants.

Success Metrics & Timeline

Success will be measured by reduced score gaps across demographics, improved fairness metrics, and documented human overrides. Full fairness compliance is targeted within 12 months.