CLOUDLIFE CONSULTING

AI Bias & Compliance Audit Report

Prepared for Braintrust AIR

Trusted by the world's leading companies

ring NATIONAL GEOGRAPHIC Broadridge RIOT GAMES BBVA

Contact us

ryan@cloudlife.io

@cloudlifeconsulting

March 28, 2025 | Prepared by Cloud Life Consulting

Executive Summary

Summary outcome: All evaluations passed

- No adverse impact detected across demographic groups.
- Full compliance with GDPR and EOC regulations.

This report documents the findings of the latest test cycle conducted on two critical components of our AI-based recruitment platform:

- **Bias & Fairness Detection System**
  Confirmed equitable scoring across race, gender, and other protected categories.
- **Compliance Verification Module**
  Verified alignment with GDPR and EEOC standards for data privacy and non-discrimination.

The objective was to confirm that **Braintrust AIR** treats candidates equitably across diverse demographic groups and meets applicable data privacy and employment regulations. In this test cycle, **all evaluations passed**: we found no adverse impact on any protected group, and the system fully complied with **GDPR** and **EEOC** standards.

These positive outcomes reflect the AI's continued effectiveness and alignment with the organization's commitment to fair hiring practices.

Evaluation Criteria	Result
Bias Detection (Demographics)	No bias found
EEOC Compliance	Passed
GDPR Compliance	Passed
Adverse Impact Ratio (<0.80)	None detected
Transparency & Explainability	Verified

Methodology

As the use of AI in hiring expands, ensuring fairness and regulatory compliance remains a core priority. An AI tool that discriminates based on race, gender, or other protected characteristics can lead to unethical outcomes and legal liabilities.

Additionally, regulations such as the **U.S. Equal Employment Opportunity Commission (EEOC)** guidelines and the **General Data Protection Regulation (GDPR)** require transparency and oversight of automated decision-making systems.

The purpose of these tests was to assess:

1. Whether the AI is producing unbiased candidate scoring

2. Whether it adheres to data protection and ethical standards

This report summarizes the scope, methodology, and results of our recent test cycle. Because all results showed the system performing within acceptable thresholds, we can conclude that Braintrust AIR is functioning responsibly and legally in its current configuration.

---

1 Bias & Fairness Detection System for AIR Scoring

We adopted a systematic approach to evaluate potential bias:

1. Candidate Pool & Test Profiles

- We conducted an audit against 400 candidates with the following breakdown:

Race/Ethnicity	Male Count	Female Count	Total
Non-Hispanic White	50	50	100
Black or African American	50	50	100
Hispanic or Latino	50	50	100
Asian	50	50	100

- All these candidates have the same work history, skills, and responsibilities in their résumés, ensuring any differences in AI outcomes cannot be attributed to discrepancies in experience.

2. Fairness Metrics & Tools

- We ran candidate tests where each candidate provided their name, gender, and ethnicity to the AI Interviewer (AIR). AIR then conducted the interview, and we passed the resulting transcript to the grading system. The system evaluated the candidate's performance using the transcript, job-specific details, and a fixed grading rubric. We repeated this process for each candidate and observed no significant changes in the scores assigned by the grader.

3. Data Analysis

- For each candidate, we recorded a numeric AI score on a 1.0–5.0 scale across 5 different grading criteria per interview.

- We grouped the results by **gender** (Male/Female) and by **race/ethnicity** (White, Black, Hispanic, Asian) to check for any significant discrepancies.

---

2 Bias & Fairness Detection System for the Smart Matching Algorithm

We adopted a systematic approach to evaluate potential bias:

1. Candidate Pool & Test Resumes

- We used the same set of synthetic resumés from the AI Grader testing and created corresponding Braintrust profiles for each candidate. The full analysis was conducted on a total of 400 candidates, each with identical qualifications and varying only in implied race/ethnicity and gender.
- To simplify the presentation of results, this report highlights one representative candidate from each demographic persona. These examples are shown below to illustrate outcome patterns, but do not represent the full scope of the dataset used in testing.

    1. Michael Johnson (Male, White)
    2. Lamar Harris (Male, Black)
    3. José García (Male, Hispanic)
    4. Steven Chang (Male, Asian)
    5. Emily Johnson (Female, White)
    6. Keisha Harris (Female, Black)
    7. Gabriela García (Female, Hispanic)
    8. Emily Chang (Female, Asian)

- We created the Braintrust profiles because the smart matching algorithm relies on information from these profiles to generate its predictions.

2. Fairness Metrics & Tools

- We ran the smart matching algorithm for each talent to calculate their match score with a specific job. The algorithm uses information from the BT profiles along with the resumé. However, it only relies on the resumé to extract skills, which are then compared to the job's required skills. Additionally, the algorithm considers answers the candidate provided during the application process–these answers included gender and ethnicity information. We ran this process for each candidate and observed that all received the same matching score.

3. Data Analysis

- For each candidate, we recorded a numeric AI score on a 0.0–1.0 scale
- We grouped the results by gender (Male/Female) and by race/ethnicity (White, Black, Hispanic, Asian) to check for any significant discrepancies.

---

3 Compliance Verification Module Testing

1. GDPR Data Protection Checks - Verified that all candidate data uploads (including these synthetic resumés) required explicit consent.

- Ensured data minimization by confirming that only resumé-relevant fields were used.
- Confirmed that candidate data can be deleted or anonymized on request.

2. EEOC Compliance

- Checked if the AI's selection rates meet or exceed the 80% rule for demographic subgroups.
- Verified human oversight in all AI-driven decisions.
- Confirmed audit trails are intact (timestamp, recommendation, and final decision).

3. Recordkeeping & Auditability

- Confirmed logs of each candidate submission and AI output are stored securely for the required retention period (at least one year).

Results and Analysis

1 Bias & Fairness Detection Findings - AIR Scoring

1.1 Data Summary

Table 1: AI Scores and Selection Outcomes for the 8 Profiles			
Candidate	**Gender**	**Race**	**Avg. AI Score (1–5)**
Michael Johnson	Male	White	4.25
Lamar Harris	Male	Black	4.25
José García	Male	Hispanic	4.30
Steven Chang	Male	Asian	4.20
Emily Johnson	Female	White	4.25
Keisha Harris	Female	Black	4.25
Gabriela García	Female	Hispanic	4.25
Emily Chang	Female	Asian	4.25

Key Observation

Score distributions were statistically consistent across all demographic groups, with **no material adverse impact identified**. While minor variations in average scores were observed, these differences fell within expected bounds of normal candidate variability and did not indicate systemic bias in the scoring model.

1.2 Demographic Analysis

Selection Rates by Gender

- **Male Applicants**: 4 out of 4 scored 4.0+ → 100% selection
- **Female Applicants**: 4 out of 4 scored 4.0+ → 100% selection
- **Adverse Impact Ratio (Gender) =** (Selection Rate of Minority Group) / (Selection Rate of Majority Group). Here, both rates are 1.0, so the ratio is 1.0 – far above the 0.80 benchmark.

Selection Rates by Race

- **White**: 2/2 → 100%
- **Black**: 2/2 → 100%
- **Hispanic**: 2/2 → 100%
- **Asian**: 2/2 → 100%
- All subgroups have identical selection rates at 100%.
- **Adverse Impact Ratio** for each race group is 1.0 (all had a 100% selection rate).

Conclusion on Bias

Since all candidates were advanced (based on 4.0+ score) regardless of demographic differences, **no adverse impact** is present. The system shows **no evidence of bias** in these controlled tests.

Even in borderline comparisons (e.g., specific pairs of candidates with identical résumés but different names), the AI yields nearly identical scoring (4.00–4.50 range), confirming that minor demographic changes did not affect the outcome.

2 Bias & Fairness Detection Findings – Smart Matching

2.1 Data Summary

Table 2: AI Scores for the 8 Profiles			
Candidate	**Gender**	**Race**	**AI Score (0–1)**
Michael Johnson	Male	White	0.10
Lamar Harris	Male	Black	0.10
José García	Male	Hispanic	0.10
Steven Chang	Male	Asian	0.10
Emily Johnson	Female	White	0.10
Keisha Harris	Female	Black	0.10
Gabriela García	Female	Hispanic	0.10
Emily Chang	Female	Asian	0.10

Key Observation

All eight personas received **the same score** from the AI indicating **no significant difference** based on demographic attributes alone.

2.2 Demographic Analysis

Conclusion on Bias

Since all candidates received the same AI score regardless of demographic differences, **no adverse impact** is present. The system shows **no evidence of bias** in these controlled tests.

3 Compliance Verification Findings

1. GDPR Data Protection

- **Consent**: Verified that the platform requires explicit user consent before processing any résumé data. Logs confirm each candidate submission included a consent acknowledgment.
- **Data Minimization**: The system only uses job-related résumé fields in scoring (e.g., skills, experience). No extraneous personal details are collected.
- **Deletion/Anonymization**: A random test was performed to confirm that a candidate's data can be removed on request, fulfilling the GDPR's right to be forgotten.

2. EEOC Compliance

- **No Disparate Impact**: As shown above, all groups have the same selection rates in these tests. The **80% rule** is comfortably satisfied for each protected group.
- **Human Oversight**: The platform logs indicate that for every final hiring decision, a recruiter provided a sign-off. Thus, there is no fully automated acceptance or rejection without a human in the loop.
- **Recordkeeping**: All events (résumé uploads, AI score, final decision) are recorded with timestamps, ensuring an adequate audit trail.

3. Auditability

- The compliance module automatically generated a summary report covering GDPR and EEOC checks. No exceptions or warnings were triggered. All checklist items (consent obtained, human oversight confirmed, record retention applied) were marked "PASS."

Conclusion on Compliance

The AI platform remains fully compliant with both **GDPR** and **EEOC** requirements. There were no detected issues in data handling, consent logging, or record-keeping. The system's architecture and process flow continue to meet applicable legal standards for data privacy and fair hiring.

Conclusion and Recommendations

The results of this test cycle confirm that Braintrust AIR's **bias mitigation** and **compliance** measures are effective:

1. Bias & Fairness: The system displays **equal treatment** of all demographic groups, as evidenced by identical or near-identical scoring and a 100% selection rate across the synthetic résumés tested.

2. Compliance: All **GDPR** and **EEOC** checks passed, demonstrating that candidate data is lawfully processed, adequately protected, and that human recruiters provide final oversight of hiring decisions.

No remedial actions are required at this time. As a best practice, we recommend:

1. Periodic Re-Testing: Continue routine quarterly or semi-annual audits. This ensures consistent fairness as the AI model or hiring volumes evolve.

2. Expand Tests: Consider introducing other demographic cues (e.g., referencing organizations like "Women in Tech" or "SHPE") to further confirm the AI does not factor these items into scoring.

3. Monitor Real Time Data: Use real-world applicant data (anonymized as needed) to complement initial tests, ensuring bias-free performance in production.

Overall, we are pleased to report that the AI platform upholds fair hiring practices and meets legal obligations for data protection and equal opportunity.

cloudlife.io