Bias Evaluation Report

December 25, 2025

Fastr.ai Proprietary Information – Confidential

Notice

This report may contain confidential and proprietary information. It is intended only for authorized recipients and may be subject to confidentiality obligations under a services agreement or nondisclosure agreement.

The bias evaluation described here (approach, observations, and analysis) is provided to support customers' internal evaluation and adoption planning. It is not intended to satisfy, and should not be used to satisfy, any customer legal obligations. Customers should consult their legal counsel and remain responsible for compliance with applicable laws and regulations.

Where analysis includes data aggregated across tenants or requisitions, results should not be interpreted as uniform or static across all use cases. Model outputs may vary by job context and by the quality and completeness of input data.

Executive Summary

This report summarizes a bias evaluation of the matching model outputs (COLD/WARM/HOT) across sensitive attributes.

We assess whether model outcomes differ meaningfully by gender and by race/ethnicity using (1) chi-square tests of independence, (2) effect size via Cramér's V, (3) rank-fairness (representation among top-k matches vs. pool share), and (4) selection rate (share of candidates receiving WARM or HOT).

Demographic attributes were inferred using established statistical methods (BIFSG) commonly used in academic, governmental, and private sector research for fairness analysis.

Key findings

- Gender: Cramér's V = 0.007 (very small association). Fairness ratios range 0.996–1.004; selection rates range 0.704–0.709.
- Race/Ethnicity: Cramér's V = 0.009 (very small association). Fairness ratios range 0.991–1.016; selection rates range 0.537–0.547.
- Adverse impact (4/5ths rule): For both gender and race/ethnicity, the ratio of the lowest to highest selection rate exceeds the EEOC's 80% threshold (gender: 0.994; race/ethnicity: 0.982), confirming no adverse impact under this standard.

Table ES-1. Summary by Gender

Gender	Pool share	Top-k share	Fairness Ratio	Selection Rate
Female	0.5	0.498	0.996	0.704
Male	0.5	0.502	1.004	0.709

Table ES-2. Summary by Race/Ethnicity

Race/Ethnicity	Pool share	Top-k share	Fairness Ratio	Selection Rate
Api	0.25	0.254	1.016	0.547
Black	0.25	0.249	0.994	0.537
Hispanic	0.25	0.248	0.991	0.538
White	0.25	0.248	0.991	0.540

Overall, both the association measures and the exposure/selection metrics suggest that results are well-balanced across the evaluated demographic groups in this sample.

Table of Contents

1. Data

The bias assessment was conducted on a statistically significant sample of candidates. The sample represents a subset of the overall candidate population and includes only those cases for which sufficient data were available to support the analysis.

Data Sources

The evaluation draws on two separate data sources that are joined at the candidate level.

Model outputs. For each tenant, the matching model scores every candidate against each job requisition and maps the result to a COLD / WARM / HOT badge according to the same thresholds used in the production system. The analysis is scoped to the top 500 candidates per job requisition.

Candidate profile data. For each tenant, a candidate-level file provides first name, last name, an optional self-reported gender field, and zip code. These fields are used exclusively for demographic inference and computation of variables of interest.

Demographic Attribute Inference

Gender and race/ethnicity were not directly collected and were instead inferred using established statistical methods.

Gender is taken from the self-reported field where available. Where the field is missing, gender is inferred from the candidate's first name using name-based statistical classification. Only unambiguous predictions ("male" or "female") are retained; ambiguous results are treated as missing and excluded from the gender analysis.

Race/ethnicity is inferred for all candidates using the Bayesian Improved Firstname Surname Geocoding (BIFSG) method [2][3]. The method combines first name, last name, and zip code to produce a probability distribution across four groups: White, Black, API, and Hispanic. Each candidate is assigned to the group with the highest probability. Candidates with a maximum group probability below 0.7 are excluded to ensure a sufficient confidence threshold.

Important: Demographic attributes are applied only after model scores have been computed, strictly for the purpose of fairness evaluation. The matching model has no access to any demographic data.

Scoring and Badge Assignment

The scoring tool assigns each candidate a scaled score, which is subsequently mapped to a COLD / WARM / HOT category, where COLD indicates a weak match to the role and HOT indicates the strongest match. Badge thresholds are identical to those used in the production system, ensuring that the bias evaluation reflects real-world model behavior rather than an artificial test setting.

Sampling Strategy

Each analysis is run in two variants to allow both a controlled fairness assessment and a real-world view of model behavior.

Balanced samples. For each demographic attribute, the dataset is downsampled so that every group is represented by an equal number of candidates — matched to the size of the smallest group present (minimum 100 candidates). This stratified random sampling removes the effect of unequal group sizes, isolating the model's behavior from population-level imbalances in the underlying talent pool.

Raw sample. No balancing is applied. The analysis uses the natural demographic distribution of candidates as it exists in the data, subject only to the overall candidate cap. Results from the raw sample reflect the real-world exposure and outcome rates that candidates from each group would experience in live deployment.

Reporting both variants side by side allows readers to distinguish between differences that arise from the model itself and those that reflect the demographic composition of the talent pool.

2. Methodology

Associations between bias-sensitive groups and model outcomes are tested using chi-square tests of independence, and their strength is characterized by Cramér's V.

Furthermore, we complement these statistics with a rank fairness analysis, which compares how different groups are represented among the top-ranked candidates relative to their overall share in the candidate pool. Finally, we measure the Selection Rate defined as the percentage of candidates considered a good match (HOT or WARM) from the given bias-sensitive category [4][5][6].

2.1 Chi-square (X^2)

We used the chi-square (X^2) test to check whether the share of candidates in each outcome category (COLD/WARM/HOT) differs across demographic groups (such as gender or race/ethnicity), compared with what we would expect if group membership had no impact on the results.

The chi-square test statistic is computed as the sum, over all categories, of the squared differences between the observed (O_i) and expected (E_i) counts, divided by the expected counts:

$$X^2 = \Sigma\ (O_i - E_i)^2 / E_i$$

In large samples (large N), even very small differences can become statistically significant (very small p-values), even when those differences are too small to be meaningful in practice. For this reason, we do not rely on chi-square significance tests in our analysis. Instead, we focus on effect

sizes (Cramér's V) to assess the practical importance of any detected association.

2.2 Cramér's V

To quantify the strength of the association detected by the chi-square test, we use Cramér's V, a standard effect-size measure for relationships between categorical variables [1].

$$V = \sqrt{(X^2 / (n \cdot \min(c - 1, r - 1)))}$$

where:

- V is Cramér's V.
- X^2 is the chi-square test statistic.
- n is the total sample size (number of observations).
- r is the number of rows in the contingency table.
- c is the number of columns in the contingency table.

By construction, V takes values between 0 and 1, where higher values indicate a stronger association. As a rough guideline:

- around 0.10 – weak association,
- around 0.30 – moderate association,
- 0.50 or higher – strong association.

2.3 Rank Fairness & Selection Rate

Rank Fairness is assessed by comparing how much visibility different demographic groups receive in ranked results with their share in the overall candidate pool.

- Pool share – for each group, the share in the entire candidate population is calculated.
- Top-k share – the share calculated only among candidates surfaced by the model (i.e., those appearing in the top-k matches for jobs).
- Fairness ratio – ratio(group) = Top-k share / Pool share.
- Selection Rate – the share of candidates that receive a positive outcome (HOT or WARM) among all candidates from that group.

Interpretation:

- ratio ≈ 1.0 – visibility is roughly proportional to presence in the population (neutral exposure).
- ratio < 1.0 – under-represented in the top-k relative to pool share (lower exposure).
- ratio > 1.0 – over-represented in the top-k relative to pool share (higher exposure).

In the bias report, for each protected attribute (e.g., gender, race), pool share, top-k share, and the corresponding exposure ratio are reported, together with the minimum and maximum ratio across groups.

3. Results

In this section, we present the bias analysis results for the matching model.

3.1 Gender Analysis

Cramér's V: 0.007253450120776708

Distribution of candidates by group

Gender	COLD count	COLD %	HOT count	HOT %	WARM count	WARM %
Female	1,904,434	29.61	796,417	12.38	3,731,257	58.01
Male	1,891,086	29.15	830,379	12.80	3,765,927	58.05

Rank Fairness

Gender	Pool share	Top-k share	Fairness Ratio
Female	0.5	0.498	0.996
Male	0.5	0.502	1.004

Selection Rate

Gender	Selection Rate
Female	0.704
Male	0.709

Conclusion: The magnitude of this measure is very low (Cramér's V ≈ 0.007), which suggests that there is no relationship between gender and the resulting ratings. The observed distributions of COLD / WARM / HOT ratings for women and men are almost indistinguishable in practice. Rank-fairness metrics are also practically in ideal parity. Women and men each account for 50% of the overall candidate pool, and their representation among top-k matches is 49.8% and 50.2% respectively (fairness ratios ≈ 0.996 and 1.004). Selection rates are likewise almost identical: 0.704 for women and 0.709 for men.

3.2 Race/Ethnicity Analysis

Cramér's V: 0.008692754192465213

Distribution of candidates by group

Race/Ethnicity	COLD count	COLD %	HOT count	HOT %	WARM count	WARM %
Api	1,486,058	45.30	300,205	9.15	1,494,414	45.55
Black	1,493,812	46.31	270,375	8.38	1,461,605	45.31
Hispanic	1,483,943	46.21	287,636	8.96	1,439,827	44.83
White	1,471,724	45.97	275,979	8.62	1,453,922	45.41

Rank Fairness

Race/Ethnicity	Pool share	Top-k share	Fairness Ratio
Api	0.25	0.254	1.016
Black	0.25	0.249	0.994
Hispanic	0.25	0.248	0.991
White	0.25	0.248	0.991

Selection Rate

Race/Ethnicity	Selection Rate
Api	0.547
Black	0.537
Hispanic	0.538
White	0.540

Conclusion: The association metric has a very small value (Cramér's V ≈ 0.009), which confirms that badge distribution (COLD/WARM/HOT) is highly consistent across racial/ethnic groups. Rank-fairness indicators remain very close to ideal parity, and selection rates are tightly clustered around ~0.54 across groups.

4. Conclusions

Summary of findings

Across both gender and race/ethnicity, the analyses consistently show that model outputs are highly similar across groups. In all cases, the association measure (Cramér's V) takes very small values — 0.007 for gender and 0.009 for race/ethnicity — well below the threshold commonly associated with even a weak practical association (≈ 0.10). This indicates that group membership has no practically meaningful impact on the distribution of COLD / WARM / HOT outcomes

Rank-fairness ratios remain close to 1.0 across all groups, confirming that no demographic group receives disproportionately higher or lower visibility in the ranked shortlists presented to recruiters. Selection rates — the share of candidates receiving a positive WARM or HOT outcome — are likewise tightly clustered, with a maximum spread of 0.5 percentage points by gender and 1.0 percentage point across racial/ethnic groups. All selection-rate comparisons satisfy the EEOC's 4/5ths rule.

Consistency across sampling variants

The findings hold consistently across both the balanced and raw sample variants. In the balanced sample, where all demographic groups are represented equally by design, the model produces near-identical outcome distributions — demonstrating that the matching model itself does not introduce differential treatment when candidates are equivalently represented. In the raw sample, which reflects the natural demographic composition of the talent pool, outcomes remain similarly uniform, indicating that population-level imbalances in the underlying data do not translate into biased model behavior whilst ranking.

The agreement between both variants is an important quality signal: it suggests that the results are not an artefact of the sampling approach, but a genuine property of the model.

Limitations and considerations

Several limitations should be kept in mind when interpreting these results. First, demographic attributes — gender and race/ethnicity — were inferred statistically rather than self-reported. While the inference methods used (gender imputation and BIFSG) are established and widely applied, they are probabilistic and may introduce misclassification for a subset of candidates, particularly those from underrepresented name origins or geographic areas.

Second, the analysis is scoped to the top 500 candidates per job requisition, which reflects the recruiter-facing shortlist. Bias patterns that might exist further down the ranking — beyond the shortlist — are outside the scope of this evaluation.

Forward outlook

Bias evaluation is treated as a continuous process rather than a one-time assessment. The methodology and thresholds applied in this report will be reviewed periodically to ensure they remain aligned with evolving best practices in algorithmic fairness. Results from future evaluation cycles will be compared against the baseline established here to detect any regressions introduced by model updates or changes in the underlying candidate population.

5. Bibliography

[1] Agresti, A. (2012). Categorical Data Analysis (3rd ed.). Hoboken, NJ: Wiley.

[2] Elliott, M. N., et al. (2009). Using the Census Bureau's surname list to improve estimates of race/ethnicity and associated disparities. Health Services and Outcomes Research Methodology, 9(2), 69–83.

[3] Imai, K., & Khanna, K. (2016). Improving ecological inference by predicting individual ethnicity from voter registration records. Political Analysis, 24(2), 263–272. [BIFSG methodology]

[4] Singh, A., & Joachims, T. (2018). Fairness of Exposure in Rankings. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD).

[5] Biega, A. J., Gummadi, K. P., & Weikum, G. (2018). Equity of Attention: Amortizing Individual Fairness in Rankings. In Proceedings of the 41st International ACM SIGIR Conference.

[6] Singh, A., & Joachims, T. (2019). Policy Learning for Fairness in Ranking. In Advances in Neural Information Processing Systems 32 (NeurIPS).

Appendix A

As an extension of our analysis, we provide an example breakdown of results and sampling approaches over two anonymized tenants. The purpose is to demonstrate that even when sensitive-variable distributions are skewed, the model's outputs remain aligned with the underlying pool distribution without systematic preference.

Tenant A

Gender results – balanced sample

Gender	COLD count	COLD %	HOT count	HOT %	WARM count	WARM %
Female	166,603	19.67	71,540	8.45	608,869	71.88
Male	163,284	19.10	77,474	9.06	614,230	71.84

Rank Fairness

Gender	Pool share	Top-k share	Fairness Ratio
Female	0.500	0.502	1.004
Male	0.500	0.498	0.995

Selection Rate

Gender	Selection Rate
Female	0.803
Male	0.809

Gender results – raw sample

Gender	COLD count	COLD %	HOT count	HOT %	WARM count	WARM %
Female	123,093	16.23	74,090	9.77	561,247	74.00
Male	154,265	16.35	96,001	10.17	693,304	73.48

Rank Fairness

Gender	Pool share	Top-k share	Fairness Ratio
Female	0.557	0.554	0.995
Male	0.443	0.446	1.006

Selection Rate

Gender	Selection Rate
Female	0.838
Male	0.837

Race results – balanced sample

Race/Ethnicity	COLD count	COLD %	HOT count	HOT %	WARM count	WARM %
Api	341,500	84.58	10,514	2.60	51,734	12.81
Black	357,413	83.82	9,826	2.30	59,175	13.88
Hispanic	355,105	83.45	11,518	2.71	58,926	13.85
White	373,731	83.74	10,992	2.46	61,566	13.80

Rank Fairness

Race/Ethnicity	Pool share	Top-k share	Fairness Ratio
Api	0.25	0.237	0.949
Black	0.25	0.251	1.002
Hispanic	0.25	0.250	1.000
White	0.25	0.262	1.049

Selection Rate

Race/Ethnicity	Selection Rate
Api	0.154
Black	0.162
Hispanic	0.166
White	0.163

Race results – raw sample

Race/Ethnicity	COLD count	COLD %	HOT count	HOT %	WARM count	WARM %
Api	45,673	14.99	32,043	10.52	227,017	74.50
Black	8,518	13.55	7,342	11.68	47,022	74.78
Hispanic	11,726	13.00	9,966	11.05	68,531	75.96
White	179,718	14.44	134,580	10.82	929,864	74.74

Rank Fairness

Race/Ethnicity	Pool share	Top-k share	Fairness Ratio
Api	0.187	0.179	0.956
Black	0.035	0.037	1.047
Hispanic	0.052	0.053	1.024
White	0.726	0.731	1.007

Selection Rate

Race/Ethnicity	Selection Rate
Api	0.850
Black	0.865
Hispanic	0.870
White	0.856

Tenant B

Gender results – balanced sample

Gender	COLD count	COLD %	HOT count	HOT %	WARM count	WARM %
Female	82,226	6.32	483,101	37.11	736,537	56.58
Male	82,189	6.34	475,251	36.65	739,196	57.01

Rank Fairness

Gender	Pool share	Top-k share	Fairness Ratio
Female	0.500	0.501	1.002
Male	0.500	0.499	0.998

Selection Rate

Gender	Selection Rate
Female	0.937
Male	0.937

Gender results – raw sample

Gender	COLD count	COLD %	HOT count	HOT %	WARM count	WARM %
Female	85,019	6.34	502,020	37.42	754,508	56.24
Male	78,827	6.27	457,744	36.42	720,382	57.31

Rank Fairness

Gender	Pool share	Top-k share	Fairness Ratio
Female	0.507	0.516	1.018
Male	0.493	0.484	0.982

Selection Rate

Gender	Selection Rate
Female	0.937
Male	0.937

Race results – balanced sample

Race/Ethnicity	COLD count	COLD %	HOT count	HOT %	WARM count	WARM %
Api	60,395	8.71	223,552	32.23	409,740	59.07
Black	55,793	8.88	188,787	30.05	383,711	61.07
Hispanic	58,055	9.28	195,787	31.31	371,493	59.41
White	59,049	9.07	196,483	30.17	395,655	60.76

Rank Fairness

Race/Ethnicity	Pool share	Top-k share	Fairness Ratio
Api	0.25	0.267	1.068
Black	0.25	0.242	0.967
Hispanic	0.25	0.241	0.963
White	0.25	0.251	1.002

Selection Rate

Race/Ethnicity	Selection Rate
Api	0.913
Black	0.911
Hispanic	0.907
White	0.910

Race results – raw sample

Race/Ethnicity	COLD count	COLD %	HOT count	HOT %	WARM count	WARM %
Api	10,773	6.09	65,254	36.89	100,870	57.02
Black	24,622	6.61	139,979	37.60	207,729	55.79
Hispanic	23,493	6.52	130,070	36.12	206,502	57.35
White	103,986	6.16	627,911	37.17	957,311	56.67

Rank Fairness

Race/Ethnicity	Pool share	Top-k share	Fairness Ratio
Api	0.069	0.068	0.987
Black	0.149	0.143	0.961
Hispanic	0.147	0.139	0.944
White	0.635	0.650	1.024

Selection Rate

Race/Ethnicity	Selection Rate
Api	0.939
Black	0.934
Hispanic	0.935
White	0.938

For questions or additional technical details, contact your Fastr.ai account representative.