version 2.0 | Nov. 4, 2026

JOELLIE AI

SUMMARY OF BIAS AUDIT RESULTS

Audit of **Amplify Screen** for New York City's Local Law 144

Prepared for
TAL Healthcare

In accordance with
New York City's Local Law 144
UGESP (1978)
ISAE 3000

joellie.ai

contact@joellie.ai

Table of Contents

1. Letter from the Lead Auditor 1
    1.1 Company Responsibilities 1
    1.2 Auditor Responsibilities 2
    1.3 Independence 3
    1.4 Audit Conclusions 3
2. System Description 4
3. Audit Summary 4
    3.1 Out of Scope 4
    3.2 Data & Provenance 5
4. Findings
    4.1 Amplify Screen 6
        4.1.1 Sex, sorted by scoring rate 6
        4.1.2 Race/ethnicity, sorted by scoring rate 6
        4.1.3 Race/Ethnicity × Sex Intersections 7
5. Audit Methodology
    5.1 Scoring configuration change and pooling 8
    5.2 Demographic Inference 8
    5.3 Fractional Assignment 9
    5.4 Small Categories 100
    5.5 Measures 10
6. Disclaimer 12
7. Terms & Definitions 13

1. Letter from the Lead Auditor

From: Joellie AI LLC

To: TAL Healthcare - the ‘Company’

Re: Bias Audit findings on Amplify Screen - the ‘System’

Date: Nov. 4th, 2026

We have independently audited Amplify Screen (the ‘System’), an automated interview tool provided by Bullhorn (the ‘Provider’) and used by TAL Healthcare (the ‘Company’). The Company engaged us.

The audit covers the System as configured and operated by the Company, on the Company’s own historical data from January 1, 2026 to October 8, 2026. It does not extend to other deployers or other configurations of the System.

The goal of this audit is to evaluate disparate impact for the EEO-1 Component 1 demographic categories of race/ethnicity, sex, and their intersections on a dataset adequate to support that evaluation, by a method whose assumptions and limitations are disclosed.

This was a direct engagement, not an attestation engagement over the Company’s own bias tests. We derived the demographic estimates and computed the impact ratios contained in this report. The figures presented are our own computations derived from the Company’s historical data.

The System was audited on the basis that it is an automated employment decision tool (AEDT) under NYC Local Law No. 144 of 2021. This report makes no determination as to whether the System is, in fact, an AEDT under this law.

1.1 Company Responsibilities

Company is responsible for:

- Providing complete and accurate candidate-level data covering the period under audit, including scores, system versions, job identifiers, candidate names and geography, date and time of scoring, and any-self declared sex and race/ethnicity (see § 5.2).
- Providing remote access for the audit team to observe the scoring system in the configuration in which it operates in production. The access was for observation only and was not used to generate scores.
- Representing accurately how the System is used in employment decisions, including any configurable thresholds and any manual review that intervenes between score and candidate outcome.
- Providing access to personnel able to explain the System's operation.

1.2 Auditor Responsibilities

No generally accepted standard exists for the auditing of algorithmic and autonomous systems. This engagement was conducted with reference to three authorities, each serving a distinct purpose: ISAE 3000 (Revised), *Assurance Engagements Other than Audits or Reviews of Historical Financial Information*, which governs the conduct of the engagement; Local Law 144 and the implementing rules of the NYC Department of Consumer and Worker Protection (DCWP), which provides the criteria against which the System was evaluated; and the Uniform Guidelines on Employee Selection Procedures, 29 CFR Part 1607, the reference point for adverse-impact analyses under Title VII.

The DCWP rules require impact ratios to be calculated and published but set no threshold for interpreting them. This report applies the four-fifths convention from the Uniform Guidelines, under which a ratio below 0.80 is generally regarded as evidence of adverse impact.

Within the scope of this engagement, the following procedures (amongst others) were performed:

- Inspection of the candidate-level records supplied by Company, and reconciliation of record counts from source to analysis population.
- Independent derivation of demographic estimates from name and geography using externally published reference data.
- Independent construction of the impact-ratio tables presented in this report for the System, and estimation by bootstrap resampling of the uncertainty attaching to each ratio.
- Verification that a single calculation engine, confirmed byte-identical by abstract-syntax-tree comparison, was applied to quantitative result tables, and that the build is deterministic, re-running from source to byte-identical output.
- To provide reasonable assurance that the reported results are free from material misstatement.
- Interviews with stakeholders (Company users) on system operation and procedures followed.
- Observation of system settings, features, automated workflows, operation and score provenance.

Stratified system version release analysis was performed in the course of the engagement and is documented separately. The tables in this report present the computed impact ratios only.

We believe the procedures performed provide a reasonable basis for the audit conclusions.

1.3 Independence

Our role in this audit satisfies the definition of an 'independent auditor' under the New York City Department of Consumer and Worker Protection (DCWP) rules governing AEDTs (6 RCNY § 5-300). Specifically, regarding Bullhorn as the System provider and the Company as the deployer of the System ('Covered Entities'), we: (i) have not been involved in developing, distributing, or using the audited System; (ii) held no employment relationship with the Company or any Covered Entity at any point during the bias audit; and (iii) maintained no direct financial interest in the Company, nor any material indirect financial interest in any Covered Entity, throughout the audit.

Fees for this engagement are fixed, are payable for the conduct of the audit, and are unrelated to the opinion rendered.

Our conclusions are based solely on the evidence obtained, evaluated against the criteria set out in this report.

1.4 Audit Conclusions

The audit conclusion rests on the following, which sections 4 and 5 of this report set out in more detail:

- No impact ratio below the four-fifths threshold.
- Impact is measured on scoring rate against the sample median. The median is the benchmark the DCWP rules prescribe for tools that score candidates (6 RCNY § 5-300); it is not a threshold the Company applies.
- Demographic characteristics are inferred for Component 1 categories based on Company's historical data. Self-declared data were available only for a subset of candidacies between 31 July and 8 October 2026, and were used to verify inference accuracy (see § 5.2).

Signed: Edward Feldman **Date:** Nov. 4, 2026

2. System Description

Joellie AI was engaged to audit Amplify Screen, an automated interview tool that screens candidates at the initial phase of the job application process. Once a job applicant completes a screening (by voice or chat), the System generates a Screening Summary which contains:

- an analysis of a candidate's responses
- strengths and potential gaps identified by the screener
- resume-based strengths and potential gaps related to the job description
- a complete transcript
- a list of top skills demonstrated, including skills not found on the resume
- an overall screening score (0-100)

Candidates who complete the screening never see a copy of their summary or score.

The overall screening score is a weighted average of two component scores:

- the interview transcript score (0-100)
- a candidate resume score (0-100)

Component score weights can be customized in system settings based on user preference. For example, if weighted 65% for the resume and 35% for the interview transcript, a candidate scoring 73 on the resume and 93 on the interview transcript receives (0.65 x 73) + (0.35 x 93) = 80.

The median score for the sample population becomes the reference point for calculating the 'scoring rate' for each of the protected demographic categories included in the analysis. The scoring rate is the rate at which candidacies from a particular demographic category receive a score above the sample population's median score. Scoring rates for each of the demographic groups analysed appear in the impact ratio tables in section 4.

3. Audit Summary

3.1 Out of Scope

1. Protected categories other than race/ethnicity, sex, and their intersections have not been assessed. Age, disability, veteran status, national origin, religion, sexual orientation, marital status, pregnancy, and citizenship status were not tested as the data does not support further inferences.
2. The audit measures scoring rate only. It does not measure who was interviewed, offered or hired, and no finding here should be restated as a claim about those outcomes.

3. The audit does not certify that the tools are ‘bias-free’. It reports measured disparity on the categories tested, in the population supplied, at a point in time.

3.2 Data & Provenance

The Company provided candidate-level records for the full audit population covering the period January 1, 2026 to October 8, 2026. Every record included a candidate ID, submission and job identifier, screener summary score and screening dates, candidate names and geography. The unit of analysis is the candidacy, one candidate’s application to one job ID, as the dataset includes individuals who have applied for more than one position.

The Company also provided self-declared sex and race/ethnicity from a voluntary candidate survey deployed on July 20, 2026 (first-record collected 7/31/26). These cover 944 candidacies (about 30% of the audit population), of which 164 have no declared sex and 237 with no declared race/ethnicity. They were used to verify the inferred demographics, not to replace them. Survey answers were mapped to the LL144 categories, though some do not map exactly; “Two or more races”, in particular, is defined differently in the survey and in the inference.

Records were extracted by the audit team from the Company’s Bullhorn instance through its read-only API and reconciled against the dataset the Company provided. Where a candidate had more than one attempt at the same job, the latest attempt and overall score was retained. Records generated by Company personnel for the purposes of conducting operational tests were confirmed by the Company and excluded from the analysis.

The audit covers 3,107 candidacies, each scored by the System’s composite summary score (see § 2).

Amplify Screen composite score	Number of records
Race/ethnicity resolved	3,107
Sex-resolved *	2,715

**Sex is estimated from each candidate’s first name as a probability, using US Social Security Administration (SSA) birth records for 1960–2005. Names used for both sexes receive a probability between 0 and 1 and remain in the analysis. The 392 candidacies without a sex estimate are those with no usable first name: the name was blank, given only as an initial, or does not appear in the SSA records. They are excluded from the sex and intersectional analyses. All but 7 have a race estimate and remain in the race/ethnicity analysis.*

4. Findings

4.1 Amplify Screen

In the following tables, N is the expected count under fractional assignment (see § 5.4). Scoring rate is the weight-summed share of candidacies scoring strictly above the sample median of 78.0. Impact ratio is against the highest-scoring rated group.

4.1.1 Sex, sorted by scoring rate

Sex	N candidacies	Scoring rate	Impact ratio
Male	1,222.4	0.4794	1.0000
Female	1,492.6	0.4750	0.9909

4.1.2 Race/ethnicity, sorted by scoring rate

Race / ethnicity	N candidacies	Scoring rate	Impact ratio
White	1,322.4	0.5017	1.0000
Asian	707.5	0.4616	0.9202
Black or African American	456.7	0.4570	0.9109
Two or more races	168.3	0.4493	0.8956
Hispanic or Latino	428.0	0.4390	0.8750
American Indian or Alaska Native (AIAN)	24.1	0.4888*	*Excluded < 2%*
Native Hawaiian or Pacific Islander (NHOPI)	0.0†	-	*Excluded < 2%*

* *The scoring rate for AIAN is reported for transparency but is based on too few candidacies for the rate to be statistically reliable, which is why the subgroup is excluded from the four-fifths impact ratio comparison.*

† *N is the sum of applicants' probabilities of belonging to the category, each paired with that candidacy's score. The main inference method groups Native Hawaiian/Pacific Islander (NHOPI) with Asian, and American Indian/Alaska Native (AIAN) with 'other', so both were separated out for each applicant (§ 5.4). AIAN was estimated from 2020 Census name tables; it has a scoring rate but, at under 2% of the sample, no impact ratio. NHOPI was estimated from a list of Pacific*

Islander surnames. No applicant's surname appears on it, so the NHOPI estimate is zero and no scoring rate or impact ratio can be calculated. This reflects the method's limits: NHOPI applicants not on the list are counted as Asian.

4.1.3 Race/Ethnicity × Sex Intersections

Race / Ethnicity	Sex	N candidacies	Scoring rate	Impact ratio
Hispanic or Latino	Male	144.1	0.4234	0.8338
Hispanic or Latino	Female	230.9	0.4448	0.8760
Non-Hispanic or Latino - White	Male	617.4	0.5078	1.0000
Non-Hispanic or Latino - Asian	Male	246.5	0.4581	0.9022
Non-Hispanic or Latino - Black or African American	Male	130.8	0.4745	0.9345
Non-Hispanic or Latino - Two or More Races	Male	65.6	0.4400	0.8665
Non-Hispanic or Latino - AIAN	Male	9.0	0.5109*	*Excluded — below 2%*
Non-Hispanic or Latino - NHOPI	Male	0.0	-	*Excluded — below 2%*
Non-Hispanic or Latino - White	Female	605.6	0.4943	0.9734
Non-Hispanic or Latino - Asian	Female	299.9	0.4787	0.9428
Non-Hispanic or Latino - Black or African American	Female	258.8	0.4591	0.9042
Non-Hispanic or Latino - Two or More Races	Female	81.7	0.4659	0.9176
Non-Hispanic or Latino - AIAN	Female	11.6	0.4738*	*Excluded — below 2%*
Non-Hispanic or Latino - NHOPI	Female	0.0†	-	*Excluded — below 2%*

* † *See § 4.1.2*

5. Audit Methodology

5.1 Scoring configuration change and pooling

The Screener's summary score combines the interview transcript score and the résumé score (see § 2). Before 13 April 2026 the two components were weighted 50/50; from that date they were weighted 35/65. For the analysis, the first candidacy scored on or after 13 April 2026 (submission_id 51766) marks the boundary between the two configurations. Of the 3,107 candidacies analysed, 2,239 (72%) were scored under the 35/65 weighting.

The change lowered summary scores. Comparing the same jobs before and after, scores fell by about 11 points, and comparing the same candidates, by about 9.5 points. The median score was 84 under the 50/50 weighting and 75 under 35/65; this difference also reflects differences in the jobs and candidates scored in each period. The candidate pools differ in composition too: based on inferred demographics, proportionally more women and more Asian and Black candidates were scored under 35/65.

As the DCWP rules require (6 RCNY § 5-300, "scoring rate"), the primary results compare every candidacy with the median of the full sample (78), across both configurations. Because scores under 35/65 are lower, candidacies scored under that weighting are more likely to fall below the full-sample median, and so are the groups over-represented among them. The primary results therefore partly reflect which configuration scored a candidacy, not only how the tool ranks groups. Results comparing each candidacy with the median of its own configuration, and results for each configuration separately, are reported in a supplementary findings report for this audit.

5.2 Demographic inference

Sex and race/ethnicity were estimated as probabilities for each candidacy (see § 5.3). The probabilities of being female or male were estimated from the given name, using the US Social Security Administration's national baby names data for 1960–2005, with estimates for rare names smoothed towards 0.5. Race/ethnicity probabilities were estimated with wru 3.1.0, using fully Bayesian Improved Surname Geocoding with first names (fBIFSG). This combines surname, given name and the most detailed 2020 Census geography available for the candidacy, with standard BIFSG or names alone where geography was unavailable or the fully Bayesian model could not be applied. wru returns five categories: Hispanic or Latino, White, Black or African American, Asian, and Other. American Indian or Alaska Native (AIAN) was separated from Other using the 2020 Census surname and first-name tables, and the remainder is reported as Two or more races. This is an approximation, because wru's Other is not defined as multiracial. Native Hawaiian

or Pacific Islander (NHOPI) was separated from Asian using a compiled list of Pacific Islander surnames (see § 5.4).

The self-declared demographic survey described in § 3.2 allows the inferred demographics to be checked against candidates' own answers. Inferred sex matches declared sex for 592 of 623 respondents (95.0%). For race/ethnicity, among the 691 respondents with a comparable answer and grouped into five categories, the inferred totals are close to the declared totals for Black (−0.1%) and Hispanic (−3%) candidates. They undercount Asian candidates by about 15–16% and overcount White candidates by about 10%. About two-thirds of the Asian shortfall comes from respondents whose first name the inference model does not recognise. Candidacies with an unrecognised first name make up 9.1% of the full population and score much like the rest: 48.3% are above the median of their own scoring period, against 47.4% of other candidacies. Misclassifying them therefore changes group sizes but has little effect on scoring rates. Applying the respondents' error rates to the full population suggests an overall Asian undercount of about 18%. This figure is indicative only, because it assumes non-respondents are misclassified at the same rates as respondents; it is reported as a sensitivity result, and the figures in table 4.1.2 have not been adjusted for it. For candidacies in the survey window, impact ratios calculated from declared and from inferred demographics agree within sampling error (§ 4.2.4): any differences between them are no larger than would be expected from chance variation in which candidates happened to be in the window.

The validation has limits. It covers only candidates who answered the survey, all of whom applied from July 31 onwards. Nearly all respondents in the race/ethnicity validation had tract- or ZIP-level geography, whereas about half of the candidacies completed before the survey began have only city- or state-level geography, or none, and the inference's accuracy for those candidacies was not measured. The validation therefore gives no direct evidence about the roughly 70% of candidacies completed before 31 July.

5.3 Fractional assignment

All demographic estimates are fractional. Each candidacy with an estimate carries a probability for each of the seven race/ethnicity categories and for each sex, and contributes that probability to the corresponding group. Intersectional groups use the product of the sex and race/ethnicity probabilities, which treats the two estimates as independent. No candidacy is assigned to its single most likely category. Doing so would discard the uncertainty that inference carries and would inflate the groups the reference data favours for a given name and area at the expense of the others. Group sizes are therefore non-integer and are reported as expected counts.

5.4 Small categories

The DCWP rules allow an independent auditor to "exclude a category that represents less than 2% of the data being used for the bias audit from the required calculations for impact ratio" (6 RCNY § 5-301(d)). Where a category is excluded, the summary of results must give the auditor's justification, together with the number of applicants and the scoring rate for that category. In this audit, the 2% threshold is applied within each block (sex, race/ethnicity and intersectional), to candidacies with an estimate in that block.

American Indian or Alaska Native (AIAN) and Native Hawaiian or Pacific Islander (NHOPI) fall below the threshold, as do their intersectional cells, and are not assigned impact ratios. AIAN was estimated for each candidate from the 2020 Census name tables, so a scoring rate could be calculated for it. NHOPI was estimated from a list of Pacific Islander surnames. No candidate's surname appears on the list, so the estimated NHOPI count is zero and no scoring rate can be calculated.

These figures should not be read as evidence that AIAN and NHOPI applicants are almost absent, because both rest on name evidence of limited reach. Most AIAN surnames are shared with other groups, so the Census name tables give little evidence about any individual, and the separation of AIAN from wru's Other category draws on names alone. NHOPI applicants are identified only if their surname is on the Pacific Islander list; others are counted as Asian or another group. The NHOPI count of zero is therefore a lower bound, and the AIAN count probably understates representation. Neither should be read as a measure of how many AIAN or NHOPI candidates applied.

5.5 Measures

Scoring rate: The DCWP rules define the scoring rate as "the rate at which individuals in a category receive a score above the sample's median score" (6 RCNY § 5-300). The median is taken over all candidacies in the sample (78), including those without a demographic estimate, and "above" means strictly above. Each candidacy with an estimate counts towards a category in proportion to its estimated probability of belonging to it. A candidacy estimated 70% likely to be Asian adds 0.7 to the size of the Asian group and, if it scored above the median, 0.7 to the Asian group's count above the median. A group's scoring rate is its count above the median divided by its size.

Impact ratio: Each category's scoring rate divided by the scoring rate of the highest-scoring category in the same block (sex, race/ethnicity or intersectional) (6 RCNY § 5-300). Categories excluded under the 2% provision (§ 5.4) cannot serve as the highest-scoring category. LL144 requires impact ratios to be reported but sets no threshold. For context, the four-fifths rule of the Uniform Guidelines on Employee Selection Procedures (29 CFR § 1607.4(D)), under which a ratio below

0.80 is generally regarded as evidence of adverse impact, is applied here by analogy to scoring rates.

Uncertainty: Two measures of uncertainty were computed. Neither is tabulated in this report, though both are reported in the supplemental findings report. First, a delta-method standard error was calculated for every impact ratio, treating both the category's rate and the reference category's rate as estimates. Second, a 95% bootstrap percentile interval was calculated from 2,000 simulated samples (seed 20261027). Each sample was drawn by selecting candidates at random, with replacement, and including all of each selected candidate's candidacies, so that a person who applied for several jobs is never split across or within samples. The median was recalculated in each simulated sample before scoring rates were computed, so the intervals include uncertainty in the median as well as in the scoring rates. If the reference category were re-selected in each resample, the highest-scoring category could change from one resample to the next, widening the intervals for reasons unrelated to the disparity being measured.

6. Disclaimer

The results in this report are point-in-time. They describe the scores the System produced in production for the candidacies supplied over the audit period, covering both weighting configurations in use during that period (§ 5.1). The audit team observed the System remotely but did not generate any scores. The results do not predict future behaviour, and any retraining or reconfiguration after the analysis date is outside their scope.

The audit measures scoring rates, not outcomes. TAL Healthcare applies no company-wide score threshold. The median used throughout is the benchmark the DCWP rules prescribe for tools that score candidates, but it corresponds to no decision TAL Healthcare actually makes. No finding in this report should be restated as a claim about who was interviewed, offered or hired. The impact ratios describe scoring across all jobs combined; they show whether groups' scoring rates differ, not why.

For the full audit period, demographic characteristics are inferred, not observed, and figures based on them carry the uncertainty of that inference in addition to sampling uncertainty. Self-declared data, available for candidacies completed from 31 July 2026, were used to verify the inference (§ 5.2). Impact ratios are reported to four decimal places for reproducibility, not because they are accurate to that precision.

Responsibility is allocated as follows. Bullhorn, as the developer of the System, is responsible for its design. TAL Healthcare, as the employment agency using the System, is responsible for the accuracy of the data supplied, for giving candidates the notices the law requires, and for publishing this summary of results. Preparing this summary does not by itself meet those obligations.

This report does not certify that the System is free from bias, and makes no determination as to whether it is an AEDT within the meaning of NYC LL144.

7. Terms & Definitions

Term	Definition
Automated Employment Decision Tool (AEDT)	"any computational process, derived from machine learning, statistical modeling, data analytics, or artificial intelligence, that issues simplified output, including a score, classification, or recommendation, that is used to substantially assist or replace discretionary decision making for making employment decisions that impact natural persons." – see § 20-870 of the Code and § 5-300 of the adopted rule for full definition
Candidacy	The unit of analysis for this audit is the candidacy, one candidate's application to one job ID, as the dataset includes individuals who have applied for more than one position.
Disparate impact or Adverse impact	"a selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or 80%) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact" – see 29 CFR Part 1607 of UGESP (1978) for full definition
Fractional assignment	Each candidate contributes a probability of membership to every demographic group, rather than being assigned to the single most likely one.
Impact ratio	"either (1) the selection rate for a category divided by the selection rate of the most selected category or (2) the scoring rate for a category divided by the scoring rate for the highest scoring category. " – see § 5-300 of the adopted rule for full definition
Scoring rate	"the rate at which individuals in a category receive a score above the sample's median score, where the score has been calculated by an AEDT." – see § 5-300 of the adopted rule for full definition
Selection rate	"the rate at which individuals in a category are either selected to move forward in the hiring process or assigned a classification by an AEDT" – see § 5-300 of the adopted rule for full definition