# Bias audit: summary of results

**AI Interview, Spotted Zebra**

Independent bias audit conducted by VerifyWise under New York City Local Law 144 of 2021 and the implementing rules of the Department of Consumer and Worker Protection.

| | |
|---|---|
| Date of bias audit | 10 September 2026 |
| Valid until | 10 September 2027 |
| Tool audited | Spotted Zebra AI Interview |
| Distribution date of the AEDT | February 2026 |
| Independent auditor | VerifyWise |

## What the tool does and what was measured

The AI Interview conducts a structured interview and scores each candidate from 1 to 5 on six skills. The six scores are combined with even weighting into an overall percentage, calculated as (sum of the six scores minus 6) divided by 24, times 100.

The tool does not select or reject candidates. Spotted Zebra confirmed that no automated progression or rejection occurs and that a person applies any cutoff. This audit therefore reports a **scoring rate**, being the proportion of candidates in each category who scored above the median for the sample, as provided for in section 5-301(c) of the rules.

**Median score for the full sample: 83.3%**

## Source and explanation of the data

This audit used **test data**, as permitted by section 5-302(b) of the rules where insufficient historical data is available to conduct a statistically significant bias audit.

**Why historical data was not used.** Across all use of the AI Interview to 11 September 2026, 281 candidates disclosed a gender identity and 50 disclosed an ethnicity. Of those 50, only three ethnicity categories reach a reportable count of 10 or more.

Local Law 144 requires impact ratios for sex, for race and ethnicity and for the intersections of the two, which is 14 categories across the seven race and ethnicity groups. Fifty disclosures spread across 14 categories average under four candidates each, and several categories have none. Rates calculated on numbers that small carry no information. The audit therefore used test data, as the rules permit where historical data is insufficient for a statistically significant bias audit. As disclosure volumes grow, future audits of this tool are expected to use Spotted Zebra's own historical data.

**How the test data was generated and obtained.** VerifyWise generated 5,000 synthetic candidate interview transcripts for a single Account Executive role. The role definition was written independently by VerifyWise and supplied to Spotted Zebra, who generated the interview questions and skills from it using their standard process. Each synthetic candidate was assigned a sex, race or ethnicity, age band and disability status drawn from a balanced allocation, together with a qualification profile ranging from weak to strong. Transcripts were generated from those profiles by a large language model against a fixed configuration and validated before delivery. Demographic attributes were held by VerifyWise and never disclosed to Spotted Zebra, who scored the transcripts without knowing which candidate belonged to which group. Scores were returned keyed by candidate identifier and joined back to the demographic data by VerifyWise.

Of 5,000 planned candidates, 4,822 transcripts were successfully generated and delivered, and 4,742 received a complete score.

**Number of individuals assessed who fall within an unknown category: 0.** Every synthetic candidate carries an assigned sex and race or ethnicity by construction.

**Number of individuals assessed but excluded from the calculations: 80**, being candidates for whom the scoring pipeline did not return a score. These are distributed evenly across every protected category, at 1.66% of each sex and between 1.17% and 2.03% of each race and ethnicity category.

## Results

### Sex categories

| Category | Candidates | Scored above median | Scoring rate | Impact ratio |
|---|---|---|---|---|
| Female | 2,368 | 1,181 | 49.87% | 1.000 |
| Male | 2,374 | 1,165 | 49.07% | 0.984 |

### Race and ethnicity categories

| Category | Candidates | Scored above median | Scoring rate | Impact ratio |
|---|---|---|---|---|
| Asian | 681 | 345 | 50.66% | 1.000 |
| White | 678 | 340 | 50.15% | 0.990 |
| Native Hawaiian or Other Pacific Islander | 679 | 338 | 49.78% | 0.983 |
| Black or African American | 678 | 334 | 49.26% | 0.972 |
| Two or More Races | 678 | 333 | 49.12% | 0.969 |
| Hispanic or Latino | 677 | 330 | 48.74% | 0.962 |
| American Indian or Alaska Native | 671 | 326 | 48.58% | 0.959 |

### Intersectional categories of sex and race or ethnicity

| Category | Candidates | Scored above median | Scoring rate | Impact ratio |
|---|---|---|---|---|
| Female, Native Hawaiian or Other Pacific Islander | 340 | 179 | 52.65% | 1.000 |
| Male, Asian | 341 | 178 | 52.20% | 0.991 |
| Female, White | 339 | 172 | 50.74% | 0.964 |
| Male, Black or African American | 337 | 168 | 49.85% | 0.947 |
| Female, Two or More Races | 344 | 171 | 49.71% | 0.944 |
|  | 332 | 165 | 49.70% | 0.944 |

| Category | Candidates | Scored above median | Scoring rate | Impact ratio |
|---|---|---|---|---|
| Female, Hispanic or Latino | | | | |
| Male, White | 339 | 168 | 49.56% | 0.941 |
| Female, Asian | 340 | 167 | 49.12% | 0.933 |
| Female, Black or African American | 341 | 166 | 48.68% | 0.925 |
| Male, American Indian or Alaska Native | 339 | 165 | 48.67% | 0.925 |
| Male, Two or More Races | 334 | 162 | 48.50% | 0.921 |
| Female, American Indian or Alaska Native | 332 | 161 | 48.49% | 0.921 |
| Male, Hispanic or Latino | 345 | 165 | 47.83% | 0.908 |
| Male, Native Hawaiian or Other Pacific Islander | 339 | 159 | 46.90% | 0.891 |

No category was excluded under the 2% threshold in section 5-301(d). The smallest intersectional category represents 7.00% of the scored sample.

## Interpretation

**No impact ratio falls below 0.80.** The lowest is 0.891, for Male candidates recorded as Native Hawaiian or Other Pacific Islander in the intersectional analysis.

A ratio of 0.80 is the conventional threshold, drawn from the federal four-fifths rule, at which a difference in outcomes is treated as warranting further examination. Local Law 144 does not set a permitted floor and does not itself reference that rule. It requires that these figures be calculated and published, which is what this summary does.

Two points a reader should weigh alongside the figures.

**The candidate population was balanced by design.** Each protected category contains approximately equal numbers of candidates and an equal spread of qualification profiles. Where a scoring rate is measured against the sample median, a balanced population tends to produce impact ratios near 1.00. These results are therefore evidence that the tool's scoring does not respond to the demographic signal carried in the transcripts. They are not a measurement of how the tool performs on Spotted Zebra's live applicant population, whose composition differs.

**The tool discriminates strongly on qualification.** Synthetic candidates were built to five qualification levels and mean scores across those levels were 49.5%, 65.4%, 83.7%, 95.6% and 98.4%. A tool with no discriminating power would also produce impact ratios near 1.00, so this gradient establishes that the tool was operating as intended when it returned no demographic difference.

The full audit report contains the method, the further analyses conducted and the limitations of this audit.