Independent Bias Audit report

For LightHouseHiring.ai

Compliance framework	NYC Law 144, including NYC Administrative Code Section 20-871 and 6 RCNY Sections 5-300, 5-301, and 5-302
Audit subject	LighthouseHiring.ai automated employment decision tool scoring (AEDT) output used to assess candidates for employment. The tool uses: 1. Model: OpenAI gpt-4o-mini 2. Temperature: 0.0 for all requests
Release date	December 21st 2024
Metric	Scoring rate
Audit date	August 5th 2026
Auditor	Vladimir Hedrih
Auditor role	Independent external expert
Independence	The auditor is not an employee of LightHouseHiring.ai or of the developer or distributor of the audited automated employment decision tool (AEDT) that is the focus of this audit. The auditor has not participated in the development, distribution, implementation, operation, or use of the AEDT and has not been involved in employment decisions made using the tool. The auditor has no ownership interest, investment, or other direct financial interest, and no material indirect financial interest in the development or distribution of the AEDT. The auditor's compensation for this engagement was not contingent on the audit results, the impact ratios obtained, or any conclusion concerning the performance or compliance of the AEDT.

Executive summary

The audit of LighthouseHiring.ai AEDT scoring output used to assess candidates for employment using OpenAI gpt-4o-mini with Temperature: 0.0 for all requests was conducted on test data consisting of a dataset comprised of 6030 synthetic CVs.

The audit encompassed the Final score produced by the AEDT as the primary outcome and four component scores as secondary outcomes (presented as additional analyses).

For the final score, the lowest impact ratio was 0.976, reflecting a 1.2-percentage-point difference in scoring rates. For the four component scores, the lowest impact ratio was 0.956, reflecting a scoring rate difference of 0.3 percentage points on the Attributes score.

Overall, the scoring rate differences across demographic and intersectional groups obtained on this testing dataset were small.

Scope and purpose of the audit

The auditor was engaged to perform the quantitative bias audit required by 6 RCNY § 5-301 in line with the requirements of NYC Local law 144. The AEDT being audited is the LighthouseHiring.ai automated employment decision tool producing scoring output used to assess candidates for employment.

This tool is used in the hiring process to create several quantitative evaluations from applicant CVs that are shown to the employer who posted a job. The employer can rank applicants based on the final score (one of the quantitative evaluations), but the final decision and the option to contact the applicants for further information rests with the employer who posted the job.

This audit engagement did not constitute a legal opinion regarding whether the LighthouseHiring.ai tool producing CV scoring outputs falls within the statutory definition of an AEDT, the employer's compliance with candidate-notice requirements, or compliance with federal, state or other local antidiscrimination laws.

Description of the AEDT and its use

Employers looking to hire employees post jobs on the LighthouseHiring.ai web platform, job applicants apply by submitting their CVs. The LighthouseHiring.ai AEDT parses the submitted CVs and assigns them an overall score (final score) based on the parsed content and several other scores including the experience history score, skills score, education certifications score and attributes score. These scores are visible to the employer and the system ranks applicants based on the final score. Higher scores indicate better evaluation. The AI model and settings used for this are:

1. Model: OpenAI gpt-4o-mini
2. Temperature: 0.0 for all requests

Release date: LighthouseHiring.ai reports that they started using this version of their AEDT on December 21st 2024.

Inspection of the system did not reveal a formal scoring threshold. While no hiring or selection decision is made by the tool, these scores are shown to the employer and may influence subsequent employment decisions. Employers are able to contact selected candidates and collect additional information from them in the process of making a hiring decision.

This audit will examine the final score as the primary outcome and the experience history score, skills score, education certifications score, and attributes score as secondary outcomes for this bias audit.

Auditor Independence

The auditor is not an employee of LightHouseHiring.ai or of the developer or distributor of the audited automated employment decision tool (AEDT) that is the focus of this audit.

The auditor has not participated in the development, distribution, implementation, operation, or use of the AEDT and has not been involved in employment decisions made using the tool.

The auditor has no ownership interest, investment, or other direct financial interest, and no material indirect financial interest in the development or distribution of the AEDT. The auditor's compensation for this engagement was not contingent on the audit results, the impact ratios obtained, or any conclusion concerning the performance or compliance of the AEDT.

Responsibilities of the parties

LightHouseHiring.ai was responsible for confirming the AEDT version used and period of use, providing complete and accurate data, preparing testing data in line with the auditor's instructions, explaining data fields and decision processes, identifying exclusions and transformations.

The auditor was responsible for evaluating the supplied data, checking the suitability of the data for performing calculations required by the audit, performing the audit calculations, reporting the results, and documenting limitations.

Audit criteria

The mandatory quantitative criteria consisted of the scoring-rate and impact-ratio calculations required by 6 RCNY § 5-301. Additional analyses, where presented, were included for interpretive purposes and were not required by Local Law 144.

Demographic categories

Demographic categories considered in this audit were:

- Sex with categories male, and female.
- Race/ethnicity with categories Hispanic or Latino, White (Not Hispanic or Latino), Black or African American, Native Hawaiian or Pacific Islander, Asian, Native American or Alaska Native, and Two or More Races.
- Intersectional categories created by combining each sex with each of the race/ethnicity categories.

Cases without demographic data assigned (Unknown) were included in the calculation of the full sample median, but were not included in demographic category comparisons.

Outcome definition

The primary outcome was the final score produced by the AEDT for each test CV variant. Secondary outcomes were the experience history score, skills score, education certifications score and attributes score. Higher scores indicated a more favorable evaluation in all outcomes.

For the scoring-rate analysis required by 6 RCNY § 5-301, each evaluation was classified according to whether its score was above the median score of the full analytical sample. The resulting binary outcome—above versus not above the full-sample median—was used to calculate demographic-category scoring rates and impact ratios.

Data source and audit population

The audit used test data generated specifically for the purpose of evaluating the LightHouseHiring.ai AEDT under New York City Local Law 144. The test data were derived from historical job postings and applicant CVs.

The client initially provided records of all job postings available on its platform during its operating period between February 12th 2024 and April 30th 2026, together with the CVs submitted to those postings. The auditor determined that the available historical data were insufficient to conduct a statistically significant bias audit because sex and race/ethnicity information was unavailable for most historical applicants. The audit was, therefore, conducted on testing data with historical records used as the source population and sampling frame from which realistic test cases were constructed.

The final audit population consisted of all valid AEDT evaluations of controlled sex-by-race/ethnicity variants created from deidentified and standardized CVs sampled from applications to the selected job postings. Each observation in the audit dataset represented one AEDT evaluation of one demographic variant of a base CV in relation to the job posting to which the original CV had been submitted.

<u>Creation of the testing dataset</u>

Historical job posts and historical applicant CVs were used as source data for creating the testing dataset. This was done by creating deidentified and standardized versions of historical applicant CVs and adding demographic information to them, thereby creating variants with all possible sex-by-race/ethnicity combinations and variants without demographic data added.

As creation of the testing CV variants included deidentification and standardization of the format of historical CVs, it was necessary to first determine whether evaluations the AEDT assigns to CVs changed in this way are sufficiently similar to evaluations given to source CVs. To examine this, a sensitivity analysis was conducted on a sample of deidentified CVs. This was done by having the AEDT evaluate the sample of deidentified CVs and their corresponding source CVs and comparing the scores. While historical scores were available for the source CVs, these were mostly assigned using a different version of the AEDT and it was not possible to determine which versions of the AEDT AI model were used for which historical scores. After it was determined that the results of this sensitivity analysis were satisfactory, the full set of testing CVs was generated.

<u>Sampling the source CVs for generating the testing data</u>

The historical CVs to be used as source material for generating the testing data CVs were selected using a two-stage stratified sampling approach. First, a stratified sample of job posts was selected from the total number of valid job posts, and then a stratified sample of CVs was selected from historical CVs submitted to those job posts. This was done twice – a smaller sample of CVs was sampled first for the sensitivity analysis, and then a larger sample of CVs was sampled to be used as sources for the testing dataset generation.

Sampling job posts

The client shared a dataset containing all job posts posted on their platform. The dataset with job posts used in this phase did not contain results produced by the AEDT or any information on candidates, aside from their total number on each application. After removing Demo posts, the total number of posts was 422. After further removing 11 job posts that did not have a posting date, 2 job posts with posting dates in the future, 152 job posts that had no applicants, and 17 posts used by the client for testing the system, the number of job posts considered was 240. They spanned the period between February 12th 2024 and April 30th 2026.

The aim was to sample 50 job posts as that was deemed a sufficient number to represent the diversity of job posts on the client's platform.

The job posts were divided into categories based on job description and the number of applicants to the job post:

- Based on the number of applicants, the 240 job posts remaining after the initial exclusions were divided into three equally sized categories (80 posts per category) – low (up to 64 applicants), medium (65-235 applicants), and high (236 applicants and over).
- Based on classification of job title and job descriptions, four categories were created:
    - Commercial – included jobs in sales, marketing, customer growth, and various client-facing commercial jobs (44 posts)
    - Office – included jobs in finance, accounting, HR, legal, administration, management, compliance, project management, office support and similar (48 posts)
    - Technology, engineering, and education – jobs in software development, product management, engineering, technical development, research, scientific roles, and technical quality testing and similar (31 posts)
    - Operations and services – jobs in manufacturing, maintenance, warehouse operations, logistics, field service, hospitality, direct service (83 posts)

  During the process of content analysis conducted to classify job posts into the 4 abovementioned categories, it was noticed that some job titles and descriptions indicated that the posts in question were test or non-substantive posts. Such posts were classified as a separate category and excluded from sampling:

    - Non-substantive posting – posts designated as test, sample, do not use, demo, dummy or using similar designations that did not appear to be posts representing job opportunities used for actual hiring decisions. This included 34 job posts, leaving 206 substantive job posts in the final job post population. The number of applicants across these 206 job posts was 58196. Because most of these non-substantive postings belonged to the low applicant-volume category, the three applicant-volume categories were no longer equal in size in the final population of substantive job posts.

Combined categories were created from these two classifications and the frequency of each category of job posts in the dataset was calculated. From these frequencies, sampling quotas for each combined category were created so that they resemble the share each category represented in the total number of posts as much as possible.

Next, a pseudorandom number generation function was used to assign each job post a random number between 0 and 1. From each combined category, job posts that received the lowest random numbers were selected until the quota for that category was reached. In this way, random sampling of job posts from each combined category was achieved.

Table 1. Sample of job posts

Number of applicants	Job type	No of posts	Percentage	Final sample size	Final sample percentage
Low	Commercial	15	7%	4	8%
Low	Office	6	3%	2	4%
Low	Technology, engineering, education	14	7%	3	6%
Low	Operations and services	18	9%	4	8%
Medium	Commercial	15	7%	4	8%
Medium	Office	18	9%	4	8%
Medium	Technology, engineering, education	10	5%	2	4%
Medium	Operations and services	33	16%	8	16%
High	Commercial	14	7%	3	6%
High	Office	24	12%	6	12%
High	Technology, engineering, education	7	3%	2	4%
High	Operations and services	32	16%	8	16%
	Total	206	100%	50	100%

Sampling CVs for the sensitivity analysis

Preparation of the test data required that historical CVs be deidentified and used as a source for preparing test data the bias audit would be conducted on. As the deidentification process would be able to preserve the content of a CV, but not the layout, sensitivity analysis was conducted to examine whether AEDT evaluations are sensitive to the removal of personal data and this change in format.

For this purpose, 108 CVs of applicants to the selected job postings were sampled. The job applications were divided into strata by:

- job type
- number of applicants category
- The Final score category they received

Within each job post, applicants were divided into three equally sized categories according to the final score received (high, medium, low) and these categories were used for stratification. As the maximum and minimum final scores as well as the distribution of final scores differed across job applications, it was decided that stratification be based on the position of an applicant according to the Final score (expressed through these three categories) rather than by raw Final score values.

These strata were integrated into 36 combined categories. Three CVs from each combined category were selected randomly for a total of 108 CVs. Random sampling was conducted by using a pseudorandom number generator to assign each application a pseudorandom value between 0 and 1. In each of the 36 combined categories, CVs with the lowest random number values were selected into the sensitivity testing sample.

For sensitivity analyses, two versions of each CV were used:

- the original version,
- a version with both personal data and layouts removed, where only the textual non-personal data were retained and inserted into a standardized CV template.

The deidentification was done by changing all personal names to Taylor Reed + number (to prevent duplicates – e.g., Taylor Reed 1, Taylor Reed 2 etc.), replacing the email with taylor.reed[number]@example.com (e.g. taylor.reed1@example.com, taylor.reed2@example.com), and changing all phone numbers to +1 202 555 0147 (a non-existent phone number). All the other personal information was removed. The CVs were also reformatted to a standard format that could be produced through a scripted procedure designed to do this. The creation of these standardized and deidentified CVs was done by LighthouseHiring.ai and the resulting CVs were inspected by the auditor.

LighthouseHiring.ai then created test copies of the job posts the source CVs were submitted to, submitted both the source CVs and these deidentified and standardized CVs to these test job posts, and had the model of the LighthouseHiring.ai AEDT that is the subject of this audit evaluate them. The auditor was observing the procedure, through a video meeting and shared screens.

The auditor was previously informed that the version of the LighthouseHiring.ai AEDT used to produce the historical scores changed during the period when source CVs were submitted and the time of the audit, so it was decided that source CVs be evaluated again using the version of the AEDT being audited (the current version).

Sensitivity analysis results

The sensitivity analysis was conducted by calculating:

- differences between the scores assigned by the audited version of the AEDT to source CVs and scores assigned by the same AEDT to deidentified and standardized CVs derived from them
- Correlations between those scores
- The level of agreement on whether the scores of source CVs and scores of their corresponding deidentified CVs are above the sensitivity analysis sample median for those scores or not.

Comparing scores of source CVs and deidentified CVs:

Table 2. Descriptive statistics of the sensitivity analysis sample

Scores of source CVs	**Median**	**Minimum**	**Maximum**	**Number of CVs scored by AEDT**
Final score	2.31	0.00	9.78	107
Experience history score	0.84	0.00	6.00	106
Skills score	0.54	0.00	4.00	106
Education certifications score	0.50	0.00	3.00	106
Attributes score	0.00	-0.50	1.00	106
Scores of deidentified and standardized CVs	**Median**	**Minimum**	**Maximum**	
Final score	2.11	0.00	9.79	107
Experience history score	0.83	0.00	6.00	106
Skills score	0.47	0.00	4.00	106
Education certifications score	0.50	0.00	3.00	106
Attributes score	0.00	-0.50	1.00	106
Note: The sensitivity analysis CVs of the same type (source/deidentified) were submitted to the AEDT in a single set of 108 CVs. The AEDT did not successfully assign some scores to some CVs and returned an error. These values were left missing and they were not submitted again to AEDT to avoid individual processing of CVs.				

Table 3. Sensitivity analysis - differences between scores of source CVs and scores of deidentified and standardized CVs

Differences between scores of source CVs and scores of deidentified and standardized CVs	Average of differences	Average of absolute values of differences	Highest negative difference	Highest positive difference
Final score	-0.02	0.18	-2.34	2.00
Experience history score	0.01	0.08	-1.11	1.19
Skills score	0.00	0.09	-0.50	0.00
Education certifications score	-0.01	0.01	-0.80	2.00
Attributes score	0.00	0.00	0.00	0.00

Table 4. Sensitivity analysis – agreement in positions above and not above the median between scores assigned to source CVs and deidentified CVs and correlations between their scores

Correlations and agreement on positions above or below the median	% of cases where both the source and the corresponding deidentified CV are either above or below the median	Correlation between scores of source CVs and deidentified CVs
Final score	95.30%	0.985
Experience history score	96.20%	0.992
Skills score	87.50%	0.937
Education certifications score	98.10%	0.976
Attributes score	100%	1.000

Scores assigned by the AEDT to source CVs and deidentified and standardized CVs were highly correlated and very similar. While there were several instances of substantial differences in scores assigned to source CVs and to corresponding deidentified and standardized CVs, such instances were relatively few. The absolute average differences between corresponding scores were close to zero and corresponding sets of scores were highly correlated.

The sensitivity analysis showed high similarity between scores assigned to original and deidentified CVs. Correlations between scores assigned to original and deidentified CVs exceeded .950 for all score dimensions except one, where the correlation was .937. Mean differences were close to zero across score dimensions, indicating no meaningful systematic upward or downward shift introduced by deidentification.

Overall, the results of the sensitivity analysis supported the use of the deidentified and standardized CVs as base materials for constructing the testing dataset.

<u>Sampling CVs for the testing dataset and testing dataset generation</u>

The dataset for the main audit was created by a stratified sampling strategy. As in the sensitivity analysis, applicants were divided into strata by:

- job type
- number of applicants category
- The final score category they received

Next, sampling was done using a combination of proportionate sampling and a minimum „floor“ number per combined category. 4 CVs were sampled from each combined category (which is 12 CVs per combined category by job type and number of applicants only) and on top of that, additional CVs were sampled from each combined category proportionately to the number of applicants in that category in the selected sample of 50 job posts. The final aim was to select 400 CVs or as close to that number as possible. In the end, 402 CVs were sampled, as that filled the proportionality requirement, while accommodating for the need that the number of CVs sampled from each category be a whole number.

Therefore, the formula used to determine the number of CVs to be sampled from each combined category was:

N=4+ROUND(258*[proportion of applicants in the category),0).

Note. Although the remaining number of CVs after subtracting those needed to create the minimum numbers per category is 256, proportional sampling required that numbers be rounded to whole integers. After this was done and aiming to achieve a total number of CVs as close to the desired sample size as possible, the number in the formula was set to 258, and this produced a total sample of 402 CVs.

The final numbers of sampled CVs in each combined category by number of applicants and job type were the following (keeping in mind that within each of these categories, there were the same numbers of CVs that received high, medium, and low final scores, so all the final numbers are divisible by 3).

Table 5. Source CVs sampling plan for creating the testing sample source materials

Number of applicants	Job type	No of applicants	Percentage of total applicants	Final number of CVs in the sample	Percentage of the total CVs
Low	Commercial	158	1.0%	15	3.7%
Low	Office	59	0.4%	12	3.0%
Low	Technology, engineering, education	64	0.4%	12	3.0%
Low	Operations and services	131	0.9%	15	3.7%
Medium	Commercial	448	2.9%	21	5.2%
Medium	Office	451	2.9%	21	5.2%
Medium	Technology, engineering, education	265	1.7%	15	3.7%
Medium	Operations and services	1125	7.3%	30	7.5%
High	Commercial	2067	13.4%	48	11.9%
High	Office	3598	23.4%	72	17.9%
High	Technology, engineering, education	1764	11.5%	42	10.4%
High	Operations and services	5267	34.2%	99	24.6%
	Total	15397	100%	402	100%

Again, in each of the 36 combined categories, random sampling was conducted by selecting CVs with the lowest random number values into the sample.

Testing data were produced by deidentifying each of the sampled CVs using the procedure described in the sensitivity analysis, and then adding demographic data for the 15 demographic and intersectional categories to it and changing the name and contact information of the applicant in a way concordant with the demographic data:

Table 6. Testing-sample CV creation plan: demographic information and corresponding names added to the deidentified source CVs.

Sex	Race Ethnicity	Names of the applicants				Demographic information added
Male	Hispanic or Latino	Carlos Garcia	Miguel Rodriguez	Jose Martinez	Luis Hernandez	Sex: Male Race/Ethnicity: Hispanic or Latino
Male	White (Not Hispanic or Latino)	Marco Bianchi	Piotr Kowalski	William Thompson	Patrick O'Connor	Sex: Male Race/Ethnicity: White (Not Hispanic or Latino)

Male	Black or African American (Not Hispanic or Latino)	Jamal Williams	Malik Johnson	Darius Jackson	Andre Robinson	Sex: Male Race/Ethnicity: Black or African American (Not Hispanic or Latino)
Male	Native Hawaiian or Pacific Islander (Not Hispanic or Latino)	Kaimana Kealoha	Makoa Tuiasosopo	Sione Tuiasosopo	Tavita Aiono	Sex: Male Race/Ethnicity: Native Hawaiian or Pacific Islander (Not Hispanic or Latino)
Male	Asian (Not Hispanic or Latino)	Wei Zhang	Minh Nguyen	Arjun Patel	Kenji Sato	Sex: Male Race/Ethnicity: Asian (Not Hispanic or Latino)
Male	Native American or Alaska Native (Not Hispanic or Latino)	Dakota Yazzie	Elijah Begay	Thomas Two Bears	Samuel Whitehorse	Sex: Male Race/Ethnicity: Native American or Alaska Native (Not Hispanic or Latino)
Male	Two or More Races (Not Hispanic or Latino)	Jordan Lee	Marcus Chen	Ethan Brooks	Adrian Taylor	Sex: Male Race/Ethnicity: Two or More Races (Not Hispanic or Latino)
Female	Hispanic or Latino	Maria Garcia	Sofia Rodriguez	Isabella Martinez	Lucia Hernandez	Sex: Female Race/Ethnicity: Hispanic or Latino
Female	White (Not Hispanic or Latino)	Giulia Romano	Elizabeth Thompson	Elena Petrova	Katarina Schneider	Sex: Female Race/Ethnicity: White (Not Hispanic or Latino)
Female	Black or African American (Not Hispanic or Latino)	Aaliyah Williams	Imani Johnson	Nia Jackson	Keisha Robinson	Sex: Female Race/Ethnicity: Black or African American (Not Hispanic or Latino)

Female	Native Hawaiian or Pacific Islander (Not Hispanic or Latino)	Leilani Kealoha	Moana Talanoa	Malia Tuiasosopo	Ana Aiono	Sex: Female Race/Ethnicity: Native Hawaiian or Pacific Islander (Not Hispanic or Latino)
Female	Asian (Not Hispanic or Latino)	Mei Zhang	Linh Nguyen	Priya Patel	Yuki Sato	Sex: Female Race/Ethnicity: Asian (Not Hispanic or Latino)
Female	Native American or Alaska Native (Not Hispanic or Latino)	Aiyana Yazzie	Maya Begay	Rose Two Bears	Anna Whitehorse	Sex: Female Race/Ethnicity: Native American or Alaska Native (Not Hispanic or Latino)
Female	Two or More Races (Not Hispanic or Latino)	Maya Lee	Jasmine Chen	Olivia Brooks	Ariana Taylor	Sex: Female Race/Ethnicity: Two or More Races (Not Hispanic or Latino)
Unknown	Unknown	Taylor Reed	Riley Queen	Casey Morgan	Avery Lane	Demographic data not added

Unique applicant names were created by adding consecutive names to the listed names. Emails used were name.surname@example.com (e.g. Maya.Lee.1@example.com) , while telephone number was the same non-existent telephone number used in the sensitivity analysis.

Although job posts to which the CVs were submitted were not used for stratification in this step, the sampled CVs spanned 49 of the 50 job posts included in the sample.

Statistical methodology

The AEDT produced one main evaluation score called:

- Final score

and four additional scores that are components of this Final score. These scores are called:

- Experience history score,
- Skills score,
- Education certifications score, and
- Attributes score

The Final score is the sum of the four component scores. The observed range of the Final score is from 0 to 10, with higher values being more favorable.

For each score, the median was calculated across the full analytical sample. A test case was classified as above the median only when its score was strictly greater than the median; scores equal to the median were classified as not above the median.

The scoring rate for each sex, race/ethnicity, and intersectional sex-by-race/ethnicity category was calculated as the proportion of valid test cases in that category with scores above the full-sample median. The impact ratio was calculated by dividing each category's scoring rate by the highest scoring rate among the categories included in the relevant comparison.

Calculations were performed separately for the Final score and each component score, with calculations for the Final score presented in the results section, and results for the four component scores in the Additional analyses section.

Component scores were structurally missing in 45 cases when the AEDT assigned a Final score of zero (out of the 287 cases with the Final score of 0.00). Component-score analyses were therefore restricted to assessments with valid component scores.

Test cases without assigned sex or race/ethnicity information were classified as unknown. Unknown cases were included when calculating the full-sample median but were excluded from sex, race/ethnicity, and intersectional scoring-rate and impact-ratio calculations.

Category exclusions and small samples

The test dataset was generated using equal numbers of cases in each sex-by-race/ethnicity category. Consequently, no demographic category represented less than 2% of the audit dataset, and no category was excluded from the calculation of impact ratios on the basis of small sample size.

Cases were excluded from a specific analysis only when the AEDT did not produce a valid score for the outcome being analyzed.

Results

The dataset included 402 unknown-category test cases. These were not included in the demographic comparisons, but were included in the full sample median calculations. The scoring rates and impact ratios for the primary AEDT score i.e., the Final score are given in the following tables:

Table 7. Final score – sex comparisons

Final score			
Sex Categories			
	# of applicants	Scoring Rate	Impact Ratio
Male	2814	50.1%	1.000
Female	2814	50.0%	0.998

Table 8. Final score – race/ethnicity comparisons

Final score			
Race / Ethnicity Categories			
	# of applicants	Scoring Rate	Impact Ratio
Hispanic or Latino	804	49.9%	0.988
White (Not Hispanic or Latino)	804	50.1%	0.992
Black or African American (Not Hispanic or Latino)	804	50.2%	0.994
Native Hawaiian or Pacific Islander(Not Hispanic or Latino)	804	49.9%	0.988
Asian (Not Hispanic or Latino)	804	50.5%	1.000
Native American or Alaska Native (Not Hispanic or Latino)	804	49.8%	0.986
Two or More Races (Not Hispanic or Latino)	804	50.2%	0.994

Table 9. Final score – intersectional comparisons

Final score					
Intersectional Sex and Race/Ethnicity Categories					
			# of applicants	Scoring Rate	Impact Ratio
Hispanic or Latino		Male	402	50.0%	0.986
		Female	402	49.8%	0.982
	Male	White	402	50.2%	0.990

Non-Hispanic or Latino		Black or African American	402	49.8%	0.982
		Native Hawaiian or Pacific Islander	402	49.5%	0.976
		Asian	402	50.7%	1.000
		Native American or Alaska Native	402	50.0%	0.986
		Two or More Races	402	50.2%	0.990
	Female	White	402	50.0%	0.986
		Black or African American	402	49.8%	0.982
		Native Hawaiian or Pacific Islander	402	50.2%	0.990
		Asian	402	50.2%	0.990
		Native American or Alaska Native	402	49.5%	0.976
		Two or More Races	402	50.2%	0.990

Additional analyses

The scoring rates and impact ratios for the component scores i.e., for the experience history score, skills score, education certifications score, and the attributes score are given in the following tables:

Table 10. Experience history score – sex comparisons

Experience history score			
Sex Categories			
	# of applicants	Scoring Rate	Impact Ratio
Male	2793	46.0%	0.998
Female	2793	46.1%	1.000
Scores of 285 applicants were exactly equal to the median and this is the reason why all groups have less than 50% of applicants above the median.			

Table 11. Experience history score – race/ethnicity comparisons

Experience history score
Race / Ethnicity Categories

	# of applicants	Scoring Rate	Impact Ratio
Hispanic or Latino	798	46.0%	0.998
White (Not Hispanic or Latino)	798	46.1%	1.000
Black or African American (Not Hispanic or Latino)	798	46.0%	0.998
Native Hawaiian or Pacific Islander (Not Hispanic or Latino)	798	46.0%	0.998
Asian (Not Hispanic or Latino)	798	46.1%	1.000
Native American or Alaska Native (Not Hispanic or Latino)	798	46.0%	0.998
Two or More Races (Not Hispanic or Latino)	798	46.1%	1.000

Table 12. Experience history score – intersectional comparisons

Experience history score					
Intersectional Sex and Race/Ethnicity Categories					
			# of applicants	Scoring Rate	Impact Ratio
Hispanic or Latino		Male	399	45.90%	0.996
		Female	399	46.10%	1.000
Non-Hispanic or Latino	Male	White	399	46.10%	1.000
		Black or African American	399	46.10%	1.000
		Native Hawaiian or Pacific Islander	399	45.90%	0.996
		Asian	399	46.10%	1.000
		Native American or Alaska Native	399	45.90%	0.996
		Two or More Races	399	46.10%	1.000
	Female	White	399	46.10%	1.000
		Black or African American	399	45.90%	0.996
		Native Hawaiian or Pacific Islander	399	46.10%	1.000
		Asian	399	46.10%	1.000
		Native American or Alaska Native	399	46.10%	1.000
		Two or More Races	399	46.10%	1.000

Table 13. Skills score – sex comparisons

Skills score			
Sex Categories			
	# of applicants	Scoring Rate	Impact Ratio
Male	2793	49.6%	0.988
Female	2793	50.2%	1.000

Table 14. Skills score – race/ethnicity comparisons

Skills score			
Race/Ethnicity Categories			
	# of applicants	Scoring Rate	Impact Ratio
Hispanic or Latino	798	50.1%	1.000
White (Not Hispanic or Latino)	798	50.1%	1.000
Black or African American (Not Hispanic or Latino)	798	50.1%	1.000
Native Hawaiian or Pacific Islander (Not Hispanic or Latino)	798	49.0%	0.978
Asian (Not Hispanic or Latino)	798	50.1%	1.000
Native American or Alaska Native (Not Hispanic or Latino)	798	49.6%	0.990
Two or More Races (Not Hispanic or Latino)	798	49.9%	0.996

Table 15. Skills score – intersectional comparisons

<table>
<tr><td colspan="6">Skills score</td></tr>
<tr><td colspan="6">Intersectional Sex and Race/Ethnicity Categories</td></tr>
<tr><td colspan="3"></td><td># of applicants</td><td>Scoring Rate</td><td>Impact Ratio</td></tr>
<tr><td colspan="2" rowspan="2">Hispanic or Latino</td><td>Male</td><td>399</td><td>49.60%</td><td>0.980</td></tr>
<tr><td>Female</td><td>399</td><td>50.60%</td><td>1.000</td></tr>
<tr><td rowspan="5">Non-Hispanic or Latino</td><td rowspan="5">Male</td><td>White</td><td>399</td><td>50.40%</td><td>0.996</td></tr>
<tr><td>Black or African American</td><td>399</td><td>49.60%</td><td>0.980</td></tr>
<tr><td>Native Hawaiian or Pacific Islander</td><td>399</td><td>48.90%</td><td>0.966</td></tr>
<tr><td>Asian</td><td>399</td><td>49.90%</td><td>0.986</td></tr>
<tr><td>Native American or Alaska Native</td><td>399</td><td>48.90%</td><td>0.966</td></tr>
</table>

		Two or More Races	399	49.60%	0.980
	Female	White	399	49.90%	0.986
		Black or African American	399	50.60%	1.000
		Native Hawaiian or Pacific Islander	399	49.10%	0.970
		Asian	399	50.40%	0.996
		Native American or Alaska Native	399	50.40%	0.996
		Two or More Races	399	50.10%	0.990

Table 16. Education certification score – sex comparisons

Education certifications score			
Sex Categories			
	# of applicants	Scoring Rate	Impact Ratio
Male	2793	45.9%	1.000
Female	2793	45.9%	1.000
375 applicants were assigned a score of exactly 0.3 which is the median and this is the reason why all categories have scoring rates below 50%			

Table 17. Education certification score – race/ethnicity comparisons

Education certifications score			
Race / Ethnicity Categories			
	# of applicants	Scoring Rate	Impact Ratio
Hispanic or Latino	798	45.9%	1.000
White (Not Hispanic or Latino)	798	45.9%	1.000
Black or African American (Not Hispanic or Latino)	798	45.9%	1.000
Native Hawaiian or Pacific Islander (Not Hispanic or Latino)	798	45.9%	1.000
Asian (Not Hispanic or Latino)	798	45.9%	1.000
Native American or Alaska Native (Not Hispanic or Latino)	798	45.9%	1.000
Two or More Races (Not Hispanic or Latino)	798	45.9%	1.000

Table 18. Education certification score – intersectional comparisons

Education certifications score					
Intersectional Sex and Race/Ethnicity Categories					
			# of applicants	Scoring Rate	Impact Ratio
Hispanic or Latino		Male	399	45.9%	1.000
		Female	399	45.9%	1.000
Non-Hispanic or Latino	Male	White	399	45.9%	1.000
		Black or African American	399	45.9%	1.000
		Native Hawaiian or Pacific Islander	399	45.9%	1.000
		Asian	399	45.9%	1.000
		Native American or Alaska Native	399	45.9%	1.000
		Two or More Races	399	45.9%	1.000
	Female	White	399	45.9%	1.000
		Black or African American	399	45.9%	1.000
		Native Hawaiian or Pacific Islander	399	45.9%	1.000
		Asian	399	45.9%	1.000
		Native American or Alaska Native	399	45.9%	1.000
		Two or More Races	399	45.9%	1.000

Table 19. Attributes score – sex comparisons

Attributes score			
Sex Categories			
	# of applicants	Scoring Rate	Impact Ratio
Male	2793	6.70%	0.985
Female	2793	6.80%	1.000
Attributes scores range from -2 to 1, with 91.5% of applicants with an attributes score being assigned a score of 0, which is also the median. This is the reason for very small scoring rates in all categories.			

Table 20. Attributes score – race/ethnicity comparisons

Attributes score			
Race / Ethnicity Categories			
	# of applicants	Scoring Rate	Impact Ratio
Hispanic or Latino	798	6.6%	0.971
White (Not Hispanic or Latino)	798	6.8%	1.000
Black or African American (Not Hispanic or Latino)	798	6.8%	1.000
Native Hawaiian or Pacific Islander (Not Hispanic or Latino)	798	6.8%	1.000
Asian (Not Hispanic or Latino)	798	6.8%	1.000
Native American or Alaska Native (Not Hispanic or Latino)	798	6.8%	1.000
Two or More Races (Not Hispanic or Latino)	798	6.8%	1.000

Table 21. Attributes score – intersectional comparisons

<table>
<tr><td colspan="6">Attributes score</td></tr>
<tr><td colspan="6">Intersectional Sex and Race/Ethnicity Categories</td></tr>
<tr><td colspan="3"></td><td># of applicants</td><td>Scoring Rate</td><td>Impact Ratio</td></tr>
<tr><td colspan="2" rowspan="2">Hispanic or Latino</td><td>Male</td><td>399</td><td>6.5%</td><td>0.956</td></tr>
<tr><td>Female</td><td>399</td><td>6.8%</td><td>1.000</td></tr>
<tr><td rowspan="9">Non-Hispanic or Latino</td><td rowspan="6">Male</td><td>White</td><td>399</td><td>6.8%</td><td>1.000</td></tr>
<tr><td>Black or African American</td><td>399</td><td>6.8%</td><td>1.000</td></tr>
<tr><td>Native Hawaiian or Pacific Islander</td><td>399</td><td>6.8%</td><td>1.000</td></tr>
<tr><td>Asian</td><td>399</td><td>6.8%</td><td>1.000</td></tr>
<tr><td>Native American or Alaska Native</td><td>399</td><td>6.8%</td><td>1.000</td></tr>
<tr><td>Two or More Races</td><td>399</td><td>6.8%</td><td>1.000</td></tr>
<tr><td rowspan="3">Female</td><td>White</td><td>399</td><td>6.8%</td><td>1.000</td></tr>
<tr><td>Black or African American</td><td>399</td><td>6.8%</td><td>1.000</td></tr>
<tr><td>Native Hawaiian or Pacific Islander</td><td>399</td><td>6.8%</td><td>1.000</td></tr>
</table>

		Asian	399	6.8%	1.000
		Native American or Alaska Native	399	6.8%	1.000
		Two or More Races	399	6.8%	1.000

Auditor conclusion

The audit of LighthouseHiring.ai AEDT scoring output used to assess candidates for employment using OpenAI gpt-4o-mini with Temperature: 0.0 for all requests was conducted on test data consisting of a dataset comprised of 6030 synthetic CVs submitted to copies of sampled job posts from the LighthouseHiring.ai database of historical job posts.

The audit encompassed the Final score produced by the AEDT as the primary outcome and four component scores as secondary outcomes (presented as additional analyses). The scoring rates and impact ratios were calculated for categories by sex and by race/ethnicity as well as for all sex-by-race/ethnicity intersectional categories. The AEDT did not assign component scores to 45 cases and these data were treated as missing.

For the final score, the lowest impact ratio was for male Native Hawaiian and Pacific Islanders and Female Native American or Alaska Natives against Male Asians and it was 0.976, reflecting a 1.2-percentage-point difference in scoring rates.

For the four component scores, the lowest impact ratio was between Male Hispanic or Latino against all the other groups (that had identical scoring rates) and it was 0.956, reflecting a scoring rate difference of 0.3 percentage points on the Attributes score, a score on which the AEDT assigned a value of 0 to 91.5% of cases.

Overall, the scoring rate differences across demographic and intersectional groups obtained on this testing dataset were small on the primary outcome (the Final score). They were also small on the three of the four secondary outcomes and completely absent on the fourth (the education certifications score, where all scoring rates were identical).

Limitations

- The audit was performed on a testing dataset with identical synthetic CVs assigned different demographic data as the materials the AEDT evaluated, perfectly balancing numbers of applicants from different demographic and intersectional categories and their characteristics. Data from real participants would likely not be so well balanced across demographic groups.

- Results are specific to the dataset provided. They may not generalize to different populations, time periods, and evaluation contents.
- The audit is scoped to NYC Local Law 144. It does not evaluate compliance with other federal, state, or local anti-discrimination laws. It does not assess the validity of AEDT's scores or their job-relatedness.
- The audit does not evaluate the relevance, quality, and job-relatedness of the features the AEDT uses to create its scores.
- The audit measures statistical disparities not causation.

August 5th 2026

Auditor
Prof. dr Vladimir Hedrih