Warden AI

ConverzAI NYC LL144 Audit Report

Generated from
ConverzAI AI Assurance Dashboard

October 27, 2025

Table of Contents

Report Summary

Warden AI is engaged by ConverzAI to perform ongoing bias audits of ConverzAI's AI system. This bias audit report has been created by Warden AI's auditing platform and reviewed by the Warden AI team.

The report covers a subset of the overall audit that relates to the requirements of the NYC Local Law 144. The methods used meet the specific requirements for conducting a bias audit of automated employment decision tools (AEDT) as published in the final rules of the NYC Department of Consumer and Worker Protection (DCWP).

A Disparate Impact Analysis was conducted to identify the adverse impact on persons of protected groups separated by sex and race/ethnicity as mandated by the Local Law 144. Warden's independent data set of real candidate profiles was used to perform the audit, due to a lack of access to historical data.

This bias audit is meant for demonstration purposes and does not indicate the bias audit results of ConverzAI's tools for any particular employer or job opportunity.

Audit information

System tested:	ConverzAI - AI Candidate Matching
Audit frequency:	Monthly
Latest audit date:	October 27, 2025
Sample size:	18,007

About Warden AI

Company summary

At Warden AI, our mission is to reduce societal discrimination through fair and transparent AI. We provide third-party oversight into AI systems, building trust and increasing adoption.

We are an independent AI auditor and assurance platform that performs ongoing audits to ensure AI systems are fair, explainable, and transparent. Our team brings extensive experience across AI, regulation, and research, including industry and academia, to deliver our solution.

Our system integrates with the AI system under test, allowing for continuous testing and monitoring. Our methodology employs a combination of bias detection techniques and uses both our proprietary datasets and historical data from the system.

Independence statement

Warden AI Ltd is an independent AI audit and assurance provider. Fees associated with our service are solely for our evaluation and their payment is not related to the outcome of the results.

Our services are strictly limited to testing and monitoring the trustworthiness of AI systems. We do not form part of the solution or in any way affect how the system under test works.

The nature of our auditing methods are the same for all systems of the same use-case that we audit, and we do not customize our service for each system.

Company information

Registered address:
600 Congress Ave, 14th Floor,
Austin, TX 78701, United States

Website:
https://warden-ai.com

Contact:
contact@warden-ai.com

System and Audit Details

System tested

Name:
ConverzAI - AI Candidate Matching

Description:
ConverzAI's AI Candidate Matching is part of their Candidate Engagement platform responsible for matching candidates to job roles, increasing the likelihood of successful placements while reducing the risk of mismatches. To assess a candidate's suitability for a job, the platform sends a set of job-related questions and the candidate's responses are used to calculate an overall score. The score reflects how well the candidate's skills, preferences, expectations, and other factors line up with the job details.

Inputs:
- Job Criteria
- Candidate responses

Outputs:
- Score (0 to 100)

Audit details

Audit frequency	Monthly
Latest audit	October 27, 2025
Data	Historical data of candidate names and calculated matching scores
Integration	Bulk export of ConverzAI's historical data to Warden's platform
Techniques	Group bias: Disparate Impact Analysis

Disparate impact metric	Scoring rate which is calculated from the matching scores produced by the AI system

Results

Impact ratio calculation for:
Sex, Race/Ethnicity, and Intersectional (Sex x Race/Ethnicity)

Sample size
18,007

Sex

Group	# Applicants	# Selected	Scoring rate	Impact ratio
Female	5,604	2,875	0.51	1.00
Male	8,684	4,209	0.48	0.94

Race/Ethnicity

Group	# Applicants	# Selected	Scoring rate	Impact ratio
Asian	3,931	1,818	0.46	0.92
Black or African American	4,710	2,355	0.50	0.99
Hispanic or Latino	2,064	1,036	0.50	1.00
White	7,302	3,677	0.50	1.00

Results

Intersectional (Sex x Race/Ethnicity)

Group	# Applicants	# Selected	Scoring rate	Impact ratio
Asian / Female	1,060	519	0.49	0.92
Asian / Male	1,234	525	0.43	0.80
Black / Female	1,120	594	0.53	1.00
Black / Male	3,009	1,478	0.49	0.93
Hispanic / Female	737	377	0.51	0.96
Hispanic / Male	1,100	540	0.49	0.93
White / Female	2,651	1,366	0.52	0.97
White / Male	3,235	1,625	0.50	0.95

Groups representing less than 2% of individuals are excluded from analysis.

This includes the following groups for which no data is available:

- Native Hawaiian or Pacific Islander
- Native American or Alaska Native
- Two or more

Methodology

Methodology overview

Our methodology for evaluating AI systems is designed to ensure fairness and transparency. Our comprehensive approach includes ongoing auditing, multiple bias detection techniques, the use of diverse datasets, and human oversight.

Ongoing audits

AI systems change frequently (often monthly, weekly, or even daily). Our audits are performed on a regular basis at the frequency detailed in this report. The exact frequency is determined with the AI provider based on the nature of their system and their propensity for product updates. In addition to the scheduled evaluations, the AI provider can choose to have an audit performed on-demand between scheduled audits.

Adherence to NYC Local Law 144

Our bias auditing approach is in adherence with NYC Local Law 144 of 2022. While our full auditing framework goes beyond the requirements of this law, we also meet the specific requirements for conducting a bias audit of automated employment decision tools (AEDT) as published in the final rules of the NYC Department of Consumer and Worker Protection (DCWP).

Our Disparate Impact Analysis identifies any adverse impact on persons of protected groups separated by sex and race/ethnicity as mandated by the Local Law 144.

Diverse datasets

Our auditing framework uses a mixture of data. We have our own proprietary datasets which provide an independent benchmark of the AI system. Our dataset is formed of real data sourced from real people where consent has been provided.

Where applicable, we also use both historical and live data to provide context for the system's long-term performance and its current real-time operations.

All datasets are ethically sourced and we adhere to high standards of data collection practices. Some of our evaluations require datasets that contain elements of personal information to test specific AI functionalities. In such instances, we ensure that consent has been explicitly obtained for the use of this information.

Methodology

Disparate impact analysis

Disparate Impact Analysis evaluates whether a protected demographic group is adversely affected compared to other groups.

We assessed the AI system using the guidance published by the NYC Department of Consumer and Worker Protection. As the tested system produces a continuous score we've measured the impact ratio using the scoring rate method.

Scoring rate

Scoring rate is a measure used to evaluate the proportion of individuals in a specific group who receive favorable outcomes from the AI system.

To calculate a group's scoring rate, we divided the number of individuals who received a score above the sample's median score by the total number of individuals with the group.

$$\text{Scoring rate} = \frac{\text{Number of individuals within group with score above the sample's median score}}{\text{Total number of individuals within group}}$$

Impact ratio

The impact ratio is a metric used to measure potential adverse impact on a group by comparing its scoring rate to the highest scoring group.

$$\text{Impact ratio} = \frac{\text{Scoring rate for the group}}{\text{Scoring rate of the highest scoring group}}$$

An impact ratio of 1 indicates no adverse impact, whereas a lower ratio indicates a higher likelihood of adverse impact. According to the four-fifths rule, an impact ratio of 0.8 (80%) or higher is considered acceptable, indicating that the AI system's outcomes are equitable across different demographic groups.

Disclaimer

This AI Assurance Report has been prepared by Warden AI Ltd. to provide an independent evaluation of the AI system developed by the AI provider in question, based on our proprietary methodologies and datasets. The evaluations and conclusions presented in this report reflect our best professional judgments derived from the information available at the time of evaluation. While we strive for accuracy and completeness, we cannot guarantee that our evaluation is exhaustive or that there are no errors.

Our methodology is designed to identify potential issues related to fairness, robustness, and other trust factors in the AI system under examination. However, our approach, like any evaluation methodology, has its limitations. It is important to understand that our findings do not guarantee the absence of any bias, flaws, or limitations within the evaluated AI system. Instead, they indicate that, based on our specific testing framework and within the scope of our analysis, no significant issues were identified.

This report is intended for informational purposes only and should not be interpreted as a guarantee of the system's performance, fairness, or suitability for any specific purpose or use case. Warden AI Ltd. disclaims any liability for any decisions made or actions taken based on the information provided in this report. By using this report, the reader agrees to assume all risks associated with such decisions or actions and agrees to hold Warden AI Ltd. harmless against any claims, damages, or liabilities that may arise from the use of the evaluated AI system.

Warden AI

ConverzAI NYC LL144 Audit Report