



Endorsed AI Audit Report

Generated from
Endorsed - AI Assurance Dashboard

May 22, 2025

Table of Contents

1. REPORT SUMMARY

<u>1.1 Audit information</u>	<u>2</u>
<u>1.2 Results summary</u>	<u>2</u>

2. ABOUT WARDEN AI

<u>2.1 Company summary</u>	<u>3</u>
<u>2.2 Independence statement</u>	<u>3</u>
<u>2.3 Company information</u>	<u>3</u>

3. SYSTEM AND AUDIT DETAILS

<u>3.1 System tested</u>	<u>4</u>
<u>3.2 Audit details</u>	<u>4</u>

4. RESULTS

<u>4.1 Sex bias</u>	<u>5</u>
<u>4.2 Race/ethnicity bias</u>	<u>6</u>
<u>4.3 Intersectional bias (Sex X Race/Ethnicity)</u>	<u>7</u>

5. METHODOLOGY

<u>5.1 Methodology overview</u>	<u>8</u>
<u>5.2 Disparate impact analysis</u>	<u>9</u>
<u>5.4 Grading System</u>	<u>10</u>

6. DISCLAIMER	<u>11</u>
----------------------	------------------

Report Summary

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

A Disparate Impact Analysis was conducted to identify the adverse impact on persons of protected groups separated by sex and race/ethnicity. The audit was performed on historical system data of real candidates who have been processed by the AI system.

Audit information

System tested:	Endorsed - AI Applicant Reviewer
Audit frequency:	Monthly
Latest audit date:	May 22, 2025
Sample size:	685,368



Results summary

Group	Group bias
Sex bias	Clear
Race/Ethnicity bias	Clear
Intersectional bias (Sex X Race/Ethnicity)	Clear

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 that is under test, allowing for continuous testing and monitoring. Our methodology employs a combination of bias detection techniques and uses our proprietary datasets and/or 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:

Warden AI Ltd, 71-75 Shelton Street,
London WC2H 9JQ, United Kingdom

Website:

<https://warden-ai.com>

Registered company number:

15321282

Contact:

contact@warden-ai.com

System and Audit Details

System tested

Name:

Endorsed - AI Applicant Reviewer

Description:

Endorsed’s AI Applicant Reviewer helps screen thousands of resumes in minutes using generative AI, improving quality-of-hires, minimizing unconscious bias, and significantly reducing applicant review time.

Hiring criteria can be defined in plain English, and the AI Applicant Reviewer evaluates each resume against these criteria, generating an overall score along with a clear explanation of the AI’s reasoning.

Inputs:

Job criteria
Candidate resume

Outputs:

Score (0 to 100)

Audit details

Audit frequency	Monthly
Latest audit	May 22, 2025
Data	Historical data of candidate names and scores
Integration	Bulk export of historical data to Warden’s platform
Techniques	Group bias: Disparate Impact Analysis
Disparate impact metric	Scoring rate which is calculated from the scores produced by the AI system

Results

The results of this audit are broken down into each bias category and each bias detection technique that was included in the audit. For more information about the techniques and results please see the methodology section below.

Sex bias

Group bias (Disparate Impact Analysis)

Evaluates if a protected group is adversely impacted by comparing the scoring rate of the protected group to the best-performing group.

Result:*Clear***Samples:***685,368*

Sex	Samples	# Selected	Scoring rate	Impact ratio	Result
Female	229,874	109,410	47.6%	0.95	Clear
Male	455,494	227,468	49.9%	1.00	Clear

The test results indicate equitable outputs across both sexes.

Results

Race/ethnicity bias

Group bias (Disparate Impact Analysis)

Evaluates if a protected group is adversely impacted by comparing the scoring rate of the protected group to the best-performing group.

Result:*Clear***Samples:**

685,368

Race/Ethnicity	Samples	# Selected	Scoring rate	Impact Ratio	Result
Asian	304,126	154,175	50.7%	1.00	Clear
Black	51,937	24,561	47.3%	0.93	Clear
Hispanic	89,037	38,174	42.9%	0.85	Clear
White	516,575	255,076	49.4%	0.97	Clear

The test results indicate mostly equitable outputs across race/ethnicities.

Results

Intersectional bias (Sex X Race/Ethnicity)

Group bias (Disparate Impact Analysis)

Evaluates if a protected group is adversely impacted by comparing the scoring rate of the protected group to the best-performing group.

Result:

Clear

Samples:

685,368

Race/Ethnicity	Sex	Samples	# Selected	Scoring rate	Impact Ratio	Result
Asian	Female	59,197	29,708	50.2%	0.97	Clear
	Male	91,929	47,391	51.6%	1.00	Clear
Black	Female	10,652	4,844	45.5%	0.88	Clear
	Male	18,257	8,858	48.5%	0.94	Clear
Hispanic	Female	23,740	10,125	42.6%	0.83	Clear
	Male	50,792	21,919	43.2%	0.84	Clear
White	Female	135,566	64,362	47.5%	0.92	Clear
	Male	292,546	148,354	50.7%	0.98	Clear

The test results indicate mostly equitable outputs across sex x race/ethnicities.

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.

This rigorous approach enables us to accurately report on the level of bias in the system and build trust with the system's users and stakeholders.

Black box testing

We use black-box testing techniques to evaluate AI systems. This approach examines the system's outputs in response to specific inputs without needing to understand the internal workings.

This enables us to make systematic judgements across different AI systems with different underlying models.

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 also choose to have an audit performed on-demand between scheduled audits if they have a significant product update.

Adherent 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.

Methodology

Disparate impact analysis

Disparate Impact Analysis evaluates whether a protected demographic group is adversely affected compared to other groups. This is achieved by comparing the scoring rate of this group to that of the best-performing group. The goal is to ensure that the AI system does not disproportionately disadvantage any specific group based on inherent characteristics such as race, ethnicity, or sex.

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.

Methodology

Grading system

We use a grading system to highlight the extent to which any bias issues were detected in the audit results.

We use a 'traffic light' system (Clear, Consider, Concern) to indicate whether no issues were found; a potential / minor issue was found; or a definite / major issue was found.

Grade	Description	Disparate impact analysis	Counterfactual analysis
Clear	No issues detected	Impact ratio $\geq 80\%$	Consistency score $\geq 95\%$
Consider	Potential or minor issue(s) detected	$60\% \leq \text{Impact ratio} < 80\%$	$90\% \leq \text{Consistency score} < 95\%$
Concern	Definite or major issue(s) detected	Impact ratio $< 60\%$	Consistency score $< 90\%$

Disclaimer

This AI Assurance Report has been prepared by Warden AI Ltd. to provide an independent audit of the AI system developed by the AI provider in question, based on our proprietary methodologies and datasets. The results and conclusions presented in this report reflect our best 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 of bias 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 audited 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

Endorsed AI Audit Report