



Ashby Criteria Evaluation Model Audit Report

September 21, 2026

Table of Contents

1. REPORT SUMMARY

- 1.1 Audit information
- 1.2 Results overview

2. ABOUT WARDEN AI

- 2.1 Company summary
- 2.2 Independence statement
- 2.3 Company information

3. AUDIT SCOPE

- 3.1 System details
- 3.2 Data details
- 3.3 Audit details

4. AUDIT RESULTS

5. METHODOLOGY

- 5.1 Methodology overview
- 5.2 Disparate impact analysis
- 5.3 Grading system

6. DISCLAIMER

- 6.1 Disclaimer

Report summary

Warden AI is engaged by Ashby to perform independent, ongoing bias audits of the Criteria Evaluation Model system. This report summarizes the most recent audit, generated using Warden AI's assurance platform and reviewed by our audit team.

The audit is structured according to Warden's bias audit framework, which is designed to align with emerging regulatory and industry standards that emphasize the role of anti-bias testing in mitigating discrimination risks related to protected characteristics. Warden's approach supports these aims while promoting fairness and accountability in automated decision-making.

For this audit, the system was evaluated using historical data provided by Ashby. A disparate impact analysis was performed to assess the system's behavior based on this data.

This report reflects the system's behavior under controlled test conditions at the time of evaluation and should not be interpreted as a formal certification or compliance determination. The findings do not predict outcomes for any specific employer, job opportunity, or real-world deployment.

Warden's engagement does not include determining whether the audited system is legally classified as an AEDT or other regulated technology. Regulatory applicability depends on the system's particular configuration, deployment and use. The existence of this audit should not be interpreted as an admission or determination that any particular law applies.

Audit information

System audited:	Criteria Evaluation Model
Audit frequency:	Monthly
Latest audit date:	September 21, 2026
Sample size:	10,839,987



Report summary

Results overview

Group	Group bias
Sex bias	Clear
Race/Ethnicity bias	Clear
Intersectional (Sex X Race/Ethnicity) bias	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 Technologies, Inc. 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.

Company information

Registered address:

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

Website:

<https://warden-ai.com>

Contact:

contact@warden-ai.com

Audit scope

AUDIT DATE

September 21, 2026

AUDIT FREQUENCY

Monthly

DATA SOURCE

Historical data

SAMPLE SIZE

10,839,987

TEST INPUTS

Job criteria, Candidate resume

TEST OUTPUTS

Match score

BIAS DETECTION TECHNIQUES

1

CHECKS

3

System details

Ashby's Criteria Evaluation Model helps organisations improve their hiring process by identifying qualified candidates against user-defined job criteria. Every applicant is evaluated against the same set of criteria, and each criterion returns its own outcome.

The model does not aggregate these criterion-level outcomes into an applicant-level score. For this audit, each application is scored as the share of its criteria that it met, counting an outcome of "undecided" as half. Where an application was re-evaluated, only each criterion's most recent outcome counts.

Data details

This audit was conducted using historical candidate data provided by the customer to evaluate system behavior on past production data. The dataset reflects prior outcomes of the system and is analyzed as provided, subject to standard quality and suitability checks performed by Warden.

Any filtering or pre-processing applied before data transfer to Warden, such as exclusion of withdrawn applications or internal test records was defined by the customer.

Demographic attributes were included in the historical dataset and have been sourced from candidates, as specified by the customer.

Audit scope

Audit details

Audits are performed on a regular cadence to continuously monitor potential bias risks. The most recent audit was completed on *September 21, 2026* and evaluates potential bias across the following 3 protected classes:

- Sex, Race/Ethnicity, and Intersectional (Sex X Race/Ethnicity).

The following bias detection technique was applied:

- **Disparate impact analysis:** This analysis evaluates if the system is adversely impacting a protected group compared to another protected group. An impact ratio of 0.8 or higher is considered acceptable.

For the disparate impact analysis, the scoring rate method was applied. Scores above the median score were classified as selected outcomes, while scores below the median were classified as not selected.

Audit results

Sex bias

Group bias (Disparate impact analysis)

Result: Clear

Sample size: 3,613,329

Group	Samples	Selected	Scoring rate	Impact ratio	Result
Female	1,347,301	508,147	37.72%	0.95	Clear
Male	2,266,028	901,725	39.79%	1.00	Clear

The results indicate equitable outputs across all groups.

Race/Ethnicity bias

Group bias (Disparate impact analysis)

Result: Clear

Sample size: 3,613,329

Group	Samples	Selected	Scoring rate	Impact ratio	Result
Asian	1,708,707	667,838	39.08%	0.92	Clear
Black or African American	408,985	153,173	37.45%	0.88	Clear
Hispanic or Latino	255,635	94,238	36.86%	0.87	Clear
Multiracial	142,757	60,454	42.35%	1.00	Clear
Native American	13,651	5,492	40.23%	-	N/A
Pacific Islander	6,315	2,369	37.51%	-	N/A
White	1,077,279	426,308	39.57%	0.93	Clear

Groups representing less than 2% of the data are excluded from the impact ratio calculation.

The results indicate equitable outputs across all groups.

Audit results

Intersectional (Sex X Race/Ethnicity) bias

Group bias (Disparate impact analysis)

Result: Clear

Sample size: 3,613,329

Group	Samples	Selected	Scoring rate	Impact ratio	Result
Asian / Female	569,880	214,498	37.64%	0.85	Clear
Asian / Male	1,138,827	453,340	39.81%	0.90	Clear
Black / Female	199,556	76,321	38.25%	0.86	Clear
Black / Male	209,429	76,852	36.70%	0.83	Clear
Hispanic / Female	107,207	39,129	36.50%	0.82	Clear
Hispanic / Male	148,428	55,109	37.13%	0.84	Clear
Multiracial / Female	59,296	23,503	39.64%	-	N/A
Multiracial / Male	83,461	36,951	44.27%	1.00	Clear
Native American / Female	4,250	1,440	33.88%	-	N/A
Native American / Male	9,401	4,052	43.10%	-	N/A
Pacific Islander / Female	2,426	878	36.19%	-	N/A
Pacific Islander / Male	3,889	1,491	38.34%	-	N/A
White / Female	404,686	152,378	37.65%	0.85	Clear
White / Male	672,593	273,930	40.73%	0.92	Clear

Groups representing less than 2% of the data are excluded from the impact ratio calculation.

The results indicate equitable outputs across all groups.

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.

Multiple bias detection techniques

Our bias detection techniques include both disparate impact analysis and counterfactual analysis. These methods help identify any potential biases in the system by comparing the outcomes for different demographic groups and testing hypothetical scenarios where demographic attributes are altered.

Including both techniques ensures a more comprehensive evaluation, as they provide complementary insights into the system's fairness.

Methodology

Hybrid auditing

Our evaluation process combines automated methods with human oversight to ensure accuracy and reliability.

By integrating AI systems with our standardized datasets, we can conduct large-scale and frequent audits. This approach is complemented by human-led data curation and quality assurance processes for creating the datasets. Additionally, our team of experts reviews and validates the results of audits to ensure reliability.

Diverse datasets

Our audits are anchored in proprietary datasets that provide an independent benchmark of the AI system. These are built around counterfactual profiles: matched sets of candidate profiles that are identical in skills, experience, and qualifications, and differ only in demographic indicators and related proxies. The datasets also include job criteria spanning a wide range of occupations, grounded in established occupational taxonomies.

The construction of these datasets draws on a range of evidence, including real profile data, established academic research, public statistics, and domain expert review. Any real profile data used was obtained under an appropriate legal basis, such as the explicit consent of the individuals who provided it. We are committed to maintaining confidentiality and protecting personal data.

Where applicable, we complement this benchmark with the system's own historical and live data, showing how it performs in real-world use and over time.

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 its scoring rate to the highest scoring 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 median score for the job they applied to by the total number of individuals with the group. Scores are not comparable across jobs, because each job sets its own requirements, so each individual is compared with the others who applied to the same job rather than with the whole sample.

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

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

Results are evaluated using a grading system that indicates the severity of any bias detected during the audit.

Grades follow a 'traffic light' model: **Clear** indicates no issues found, **Consider** indicates potential or minor concerns requiring review, and **Concern** indicates significant issues requiring attention.

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

Disclaimer

This report has been prepared by Warden Technologies, Inc. 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 Technologies, Inc. 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.



Ashby Criteria Evaluation Model Audit Report