PILOT

ORCAA

Bias Audit Report Summary: Bullhorn Automation

Bullhorn engaged a third-party provider, O'Neil Risk Consulting & Algorithmic Auditing (ORCAA), to undertake a bias audit of the Bullhorn Automation service to evaluate inherent bias. The audit evaluated 49,902 candidates by gender and race/ethnicity to determine the Match Score (e.g. score given to a candidate in terms of the candidate's potential match for a job) and Match Percentage (e.g. percentage of candidates with a Match Score above the average Match Score in a group), among other metrics. The results of the audit showed that the average Match Score and Match Percentage were similar across groups (see Slides 4 and 5). The average Match Score was similar across all groups at approximately 96% (e.g. the average rating of candidates in terms of potential match for a job was approximately 96% in each group) and the average Match Percentage within each group was 100% (e.g. the average percentage of candidates receiving a Match Score at or above the average 96% Match Score, the ScoringRate). Each of the groups assessed in the audit – Female Asian and Pacific Islander (API), Female Black, Female Hispanic, Female White, Male API, Male Black, Male Hispanic, and Male White – scored approximately the same in the Bullhorn Automation product (see Slide 6). Each group assessed received the same MatchPercentage / ScoreRate of 100% and further comparison of each group with the highest rated group (ImpactRatio1) showed that all groups also had a similar result of 90-100% (e.g. all groups were rated similarly to one another).

Version March 2024

PILOT

ORCAA

Bias Audit Report: Bullhorn Automation

Prepared for Bullhorn | September 6, 2023

Audit Scope

Audit client: Bullhorn

Automated Employment Decision Tool: **Bullhorn Automation**, described by Bullhorn as: “a tool that automates common actions for Staffing firms. This includes emails, SMS messages, internal notifications, notes and field updates. As a part of these messages that are automated, customers can also include matching job openings or when a new job opening is added, customers can have their recruiting teams sent a list of potential candidates, subject to a minimum Match Score set by the customer.” To match candidates with job openings, Bullhorn Automation calculates a Match Score representing the extent to which the candidate’s resume fits the job title.

Employment decisions evaluated: Match Score (0-100%)

Protected classes addressed: Race/ethnicity (inferred using first name, surname, and address*), sex (inferred using first name†).

Scope of data: 49,902 candidates that were sent to customers as top potential matches for jobs, subject to the customer’s minimum Match Score. This is Historical data** drawn from 26 customers.

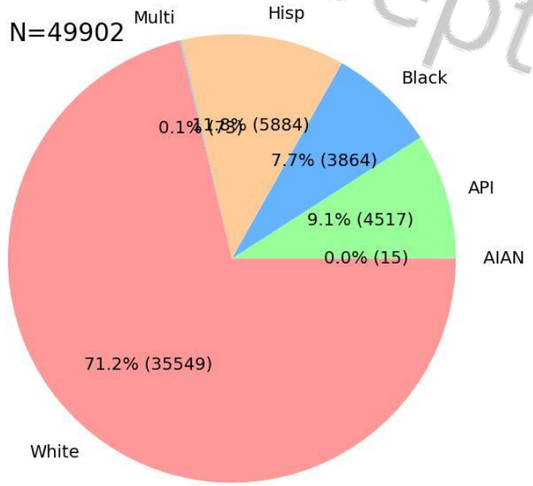
Detail on inferences and data cleaning:

- A gender label was given to all candidates. Possible labels are “male,” “female,” “mostly male,” “mostly female,” “androgynous,” or “unknown.” We counted “mostly male” as “male” and likewise for female. In gender analyses we exclude “androgynous” (696) and “unknown” (8,789).
- A race/ethnicity label was given to all candidates. Geocoding failed for 29,235 candidates, meaning the address was missing or invalid; in these cases the inference was based on first name and surname only. In race/ethnicity analyses we exclude the following groups since each has <2% of all data: race/ethnicity “Multiracial” (73) and “Native American or Alaska Native” (15).

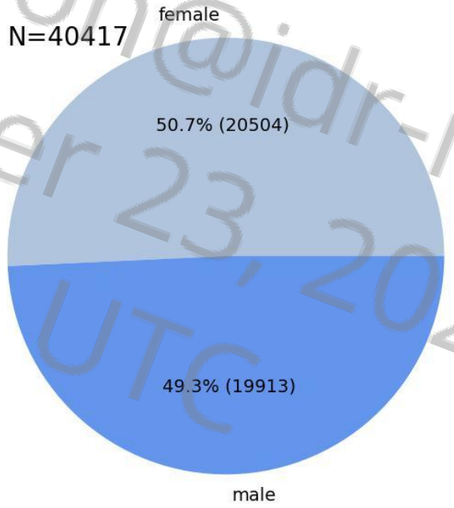
* We used BIFSG, leveraging US Census data; see Voicu for details
† We used this open-source methodology by Jörg Michael
** We believe this is “Historical data” as defined in final rules for Local Law 144. However, in light of FAQ III.6 it may be considered "Test data" since candidates' race/ethnicity and gender were inferred. If so, then the explanation for using this Test data is that the data used for purposes of the audit did not contain complete self-reported demographic data of candidates assessed by the Bullhorn Automation product.

1. Sourcing Analysis

Candidates reviewed: Race/ethnicitv

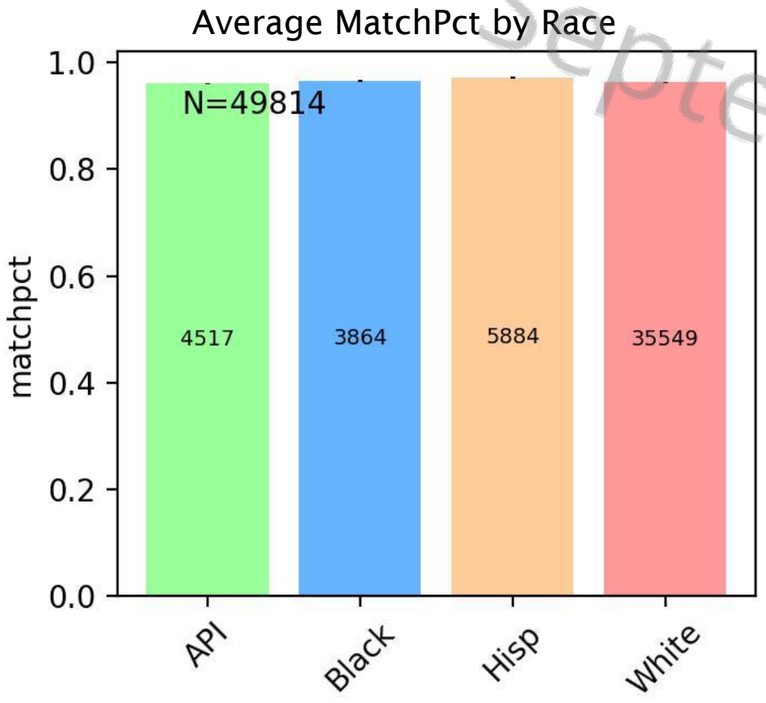


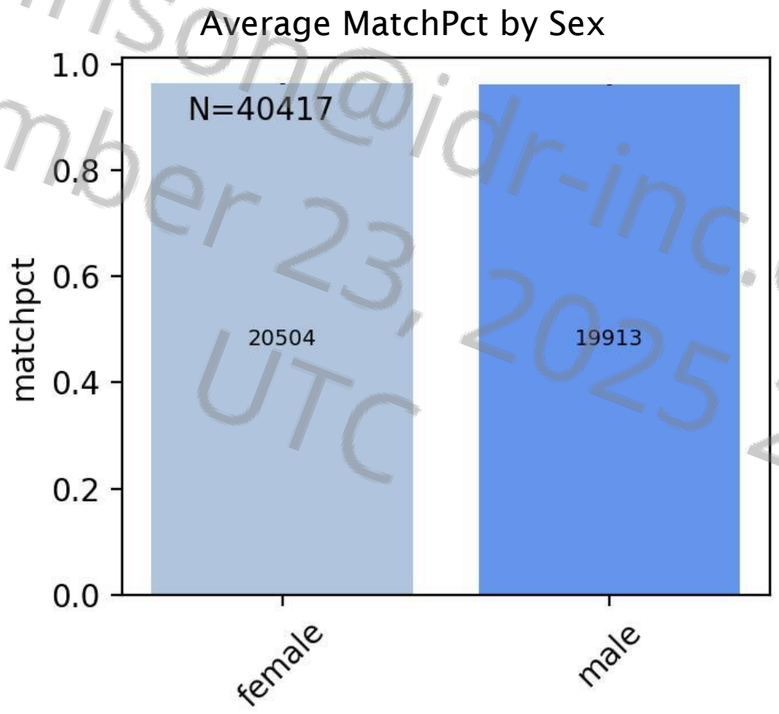
Candidates reviewed: sex



The pie charts show the breakdown by inferred race/ethnicity (left) and inferred gender (right) of candidates that were assessed by Bullhorn Automation.

2. Disparate Impact Analysis: Match Percentage





This bar chart shows the average Match Percentage for each race and gender group of candidates. The black “whiskers” at the top of each bar show a 95% confidence interval of the estimate for that group.

Average scores are high (~0.96) and similar across groups.

3. Required Tables (1 of 2)

inferred_race	n_Applied	ScoringRate	ImpactRatio	ScoringRate1	ImpactRatio1
API	4517	0.00	NaN	0.60	0.91
Black	3864	0.00	NaN	0.62	0.94
Hisp	5884	0.00	NaN	0.66	1.00
White	35549	0.00	NaN	0.60	0.91

inferred_gender	n_Applied	ScoringRate	ImpactRatio	ScoringRate1	ImpactRatio1
female	20504	0.00	NaN	0.62	1.00
male	19913	0.00	NaN	0.61	0.99

Column definitions:

1. n_applied: How many candidates applied
2. ScoringRate: share of candidates with scores above the median score
3. ImpactRatio: ScoringRate for this group, divided by ScoringRate for the highest-ScoringRate group
   Note: Since the median MatchPct for every demographic group was 100, all ScoringRates were 0, and ImpactRatio could not be calculated. Therefore we also show:
4. ScoringRate1: proportion of candidates in this group with MatchPct=100 (i.e., median or higher)
5. ImpactRatio1: ScoringRate1 for this group, divided by ScoringRate1 for the highest-ScoringRate1 group

We omit inferred_race groups with <2% of the data: Multiracial (n=57, ScoringRate=0, ScoringRate1=0.67) and AIAN (n=15, ScoringRate=0, ScoringRate1=0.8).

3. Required Tables (2 of 2)

inferred_gender	inferred_race	n_Applied	ScoringRate	ImpactRatio	ScoringRate1	ImpactRatio1
female	API	820	0.00	NaN	0.61	0.92
	Black	1526	0.00	NaN	0.63	0.95
	Hisp	2357	0.00	NaN	0.66	1.00
	White	15769	0.00	NaN	0.61	0.92
male	API	1291	0.00	NaN	0.60	0.90
	Black	990	0.00	NaN	0.63	0.96
	Hisp	2431	0.00	NaN	0.64	0.97
	White	15165	0.00	NaN	0.60	0.91

Column definitions:

1. n_applied: How many candidates applied
2. ScoringRate: share of candidates with scores above the median score
3. ImpactRatio: ScoringRate for this group, divided by ScoringRate for the highest-ScoringRate group

   Note: Since the median MatchPct for every demographic group was 100, all ScoringRates were 0, and ImpactRatio could not be calculated. Therefore we also show:
4. ScoringRate1: proportion of candidates in this group with MatchPct=100 (i.e., median or higher)
5. ImpactRatio1: ScoringRate1 for this group, divided by ScoringRate1 for the highest-ScoringRate1 group

We omit inferred_race groups with <2% of the data: Multiracial (n=57, ScoringRate=0, ScoringRate1=0.67) and AIAN (n=15, ScoringRate=0, ScoringRate1=0.8).

PILOT

ORCAA

Disclaimer: The information provided herein is Bullhorn Inc's ("Company's") Confidential Information and should not be further disseminated. The information provided is for informational purposes only and Company and its affiliated companies disclaim all legal representations and warranties. The report is believed to be accurate as of the report date but is subject to change without notice.

Please note that the criteria presented in this report were constructed to correspond with the requirements of a "bias audit" outlined in NYC Local Law No. 144 of 2021 (the "NYC LL No. 144") and the requirements set forth in the rules announced by the New York City Department of Consumer and Worker Protection ("DCWP") on April 6, 2023 and adopted by the DCWP to implement the NYC AEDT law by amending Chapter 5 of Title 6 of the Rules of the City of New York to add Subchapter T (the "implementation rules" and together with NYC LL No. 144, the "NYC AEDT Law"). Bullhorn Automation was audited as though it were an automated employment decision tool (" AEDT") under the NYC AEDT Law, but Company does not make any determination whether Bullhorn Automation is, in fact, an AEDT under the NYC AEDT Law, which depends in part on how the Customer uses Bullhorn Automation. Note that this report does not pertain to any other laws or regulations that may be applicable to Bullhorn Automation or other Bullhorn Services.

Company is a service provider and not an "Employer" under the NYC AEDT Law in regards to Customer's use of Bullhorn Automation. Customer is the "Employer" under the NYC AEDT Law and, accordingly, it is the responsibility of Customer to ensure that it complies with any applicable requirements under the NYC AEDT Law.

Warden AI

ConverzAI NYC LL 144 Audit Report

Generated from
ConverzAI - AI Assurance Dashboard

October 30, 2024

Table of Contents

1. REPORT SUMMARY 2

2. ABOUT WARDEN AI

2.1 Company summary 3
2.2 Independence statement 3
2.3 Company information 3

3. SYSTEM AND AUDIT DETAILS

3.1 System tested 4
3.2 Audit details 4

4. RESULTS

4.1 Sex bias 5
4.2 Race/Ethnicity bias 5
4.3 Intersectional bias (Sex X Race/Ethnicity) 6

5. METHODOLOGY

5.1 Methodology overview 7
5.2 Disparate impact analysis 8

6. DISCLAIMER 9

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 assess potential adverse impact on protected groups, specifically by sex and race/ethnicity, in compliance with Local Law 144. The audit utilized historical data from real candidates who were processed by the AI system.

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 30, 2024

Samples: 3,328

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:
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 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 reponses

Outputs:
Score (0 to 100)

Audit details

Audit frequency	Monthly
Latest audit	October 30, 2024
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

Disparate impact calculated for:
Sex, Race/Ethnicity, and Intersectional (Sex X Race/Ethnicity)

Total records
3,328

Sex bias

Sex	# Applicants	# Selected	Selection Rate	Impact Ratio
Female	1387	679	49.0%	1.00
Male	1941	902	46.5%	0.95

Race/Ethnicity bias

Race/Ethnicity	# Applicants	# Selected	Selection rate	Impact ratio
Asian	612	275	44.9%	0.93
Black or African American	1153	560	48.6%	1.00
Hispanic or Latino	358	163	45.5%	0.94
White	1205	583	48.4%	1.00

Results

Intersectional bias (Sex X Race/Ethnicity)

Race/Ethnicity	Sex	# Applicants	# Selected	Scoring rate	Impact ratio
Asian	Female	262	125	47.7%	0.94
	Male	350	150	42.9%	0.84
Black Or African American	Female	523	266	50.9%	1.00
	Male	630	294	46.7%	0.92
Hispanic or Latino	Female	162	75	46.3%	0.91
	Male	196	88	44.9%	0.88
White	Female	440	213	48.4%	0.95
	Male	765	370	48.4%	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 also choose to have an audit performed on-demand between scheduled audits if they have a significant product update.

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.

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 binary score we've measured the impact ratio using the selection rate method.

Selection Rate

Selection rate is a measure used to evaluate the proportion of individuals in a specific group who receive favorable outcomes.

To calculate a group's selection rate, we divided the number of individuals that have completed the phone interview by the total number of individuals with the group.

$$\text{Selection Rate} = \frac{\text{Number of individuals within group that have completed the phone interview}}{\text{Total number of individuals within group}}$$

Impact Ratio

The Impact Ratio is a metric used to measure potential adverse impact on a group by comparing its selection rate to the most selected group.

$$\text{Impact Ratio} = \frac{\text{Selection rate for the group}}{\text{Selection rate of the most selected group}}$$

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 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

ConverzAI NYC LL 144 Audit Report