babl

OFFICIAL DOCUMENT
V1.0: **06/26/2026**

PREPARED BY BABL AI INC.

SUMMARY OF BIAS AUDIT RESULTS

Audit of **LANDED's Candidate Vetting Product**
for New York City's Local Law 144

Presented to
LANDED

babl

Table of Contents

Letter of Assurance **2**
- Company Responsibilities 2
- BABL AI Responsibilities 2
- Independence 3
- Opinion 3
- Emphasis of Matters 3

System Description **4**

Audit Summary **5**
- Background 5
- Auditor Responsibilities 5
- Scope & Objective 6
  - Out of Scope 6

Findings **7**
- Disparate Impact 7

Appendix **13**
- Audit Methodology 13
  - The Criterion Audit 13
- Terminologies & Definitions 13

Letter of Assurance

From: **BABL AI Inc.**
630 Fairchild Street
Iowa City, Iowa 52245

To: **Gofriendshop, Inc. dba LANDED**
2261 Market Street,
San Francisco, California, 94114

Re: **Audit Opinion on LANDED's Candidate Vetting Product**

06/26/2026

We have conducted a direct limited assurance engagement of Gofriendshop, Inc. dba LANDED (the "Company") as of 06/26/2026, by performing independent bias testing of the Company's Candidate Vetting Product in accordance with the criteria and audit methodology set forth in this report. The goals of this audit are to:

1. Perform bias testing procedures of the Company's products in accordance with the audit criteria and testing methodology set forth in this report, and
2. Obtain limited assurance over the bias testing results, describing the procedures performed and the consistency of the findings with the audit criteria

Note that the criteria presented in this report were constructed specifically to address the requirements of a "bias audit" outlined in NYC Local Law No. 144 of 2021. The model was audited as though it were an automated employment decision tool (AEDT) under NYC Local Law No. 144 of 2021, but we do not make any determination whether the model is, in fact, an AEDT under this law.

Company Responsibilities

It is the responsibility of Company representatives to support the direct engagement process. Company representatives are responsible for reviewing and approving the testing methodology proposed by BABL AI, providing all necessary resources, documentation, and access to personnel to ensure an effective engagement, and ensuring that all evidential material provided is accurate and free of misrepresentations.

BABL AI Responsibilities

It is the responsibility of BABL AI to perform bias testing in accordance with the criteria and proposed testing methodology. In light of the current absence of generally accepted standards for the auditing of algorithms and autonomous systems, our engagement was

conducted in accordance with the standards and normative references outlined in this report.

Within the scope of our engagement, we performed the following procedures:

- Interviewing Company employees to gain an understanding of the products
- Proposing bias testing methodology based on information about and risks associated with the Company's products
- Performing testing procedures in accordance with the proposed methodology to obtain evidence to provide a basis for our conclusion

We believe these procedures provide a reasonable basis for our assurance opinion.

Independence

Our role as an independent auditor conforms to ForHumanity and Sarbanes-Oxley definitions of Independence. Fees associated with this contract are for the provision of the service to assess compliance. The payment of fees is unrelated to the decision rendered. Our decision is grounded solely in the criteria presented below.

Opinion

Based on the procedures performed and evidence obtained, we provide limited assurance on the bias testing results of Company's products, as of 06/26/2026, as prepared in accordance with the criteria outlined below.

Emphasis of Matters

We emphasize a matter regarding the disparate impact testing dataset: (1) since self-reported demographic information was not available for the system, labels for gender and race/ethnicity were generated through the use of inference models, and (2) due to the use of such inferred labels, the data likely falls under "Test Data" per § 5-300. As a result, the findings derived from the disparate impact analysis should be interpreted in light of this constraint. To mitigate risks stemming from the uncertainty of these demographic inferences, we also performed a sensitivity analysis. Our formal opinion remains unchanged in light of this matter.

Sincerely,

BABL AI Inc. 2026-06-26

System Description

BABL AI was engaged to audit Candidate Vetting Product (the "System"). The system is built to conduct automated screening questionnaires via text conversations with candidates. After consenting to the conversational chatbot interaction, candidates are engaged in a structured vetting process featuring role-specific questions. A Large Language Model (LLM) evaluates these responses to determine if the candidate meets the necessary criteria. Candidates who successfully pass this evaluation are subsequently prompted to book an interview, coordinated with the hiring manager's current availability.

The system produces a pass/fail outcome for the vetting process. This outcome was used to calculate the "selection rate" for candidates that belong to different inferred demographic groups; this selection rate is displayed in the summary of the Disparate Impact results in the Findings section.

Audit Summary

Background

New York City Local Law No. 144 of 2021 requires yearly "bias audits" for automated employment decision tools (AEDTs) used to substantially assist or replace decisions in hiring or promotion. Specifically, the law states that (1) the bias audit must "assess the [AEDTs'] disparate impact" on certain persons, (2) the audit must be conducted by an "independent auditor … no more than one year prior to the use", and (3) a "summary of the results of the most recent bias audit … [must be] made publicly available on the website of the employer or employment agency." The audit outlined in this document has been conducted to satisfy the law's requirement for a bias audit only, and does not include other requirements such as candidate notifications. This report does not make any determination whether the model under this audit is, in fact, an automated employment decision tool as defined under NYC Local Law 144, or not.

Auditor Responsibilities

It is the responsibility of BABL AI auditors to:

1. Perform bias testing procedures of the auditee's products in accordance with the audit criteria and testing methodology set forth in this report,
2. Obtain limited assurance over the bias testing results, describing the procedures performed and the consistency of the findings with the audit criteria, and
3. Issue an auditor's report that includes an opinion.

As part of an audit in accordance with good auditing practice, BABL AI exercises professional judgment and maintains professional skepticism throughout the audit. Specifically, BABL AI auditors identify and assess the risks stemming from the tool architecture, user-interaction design, and/or use, perform analysis to mitigate those risks where possible, and obtain audit evidence to provide a basis for our conclusion. In addition, this audit report follows International Standard on Assurance Engagements (ISAE) 3000's guidelines on Assurance Report, where applicable.[1]

BABL AI is also responsible for maintaining auditors' independence and objectivity to ensure the integrity of the opinion and certification provided. BABL AI as an organization, and all employee and contract auditors, adhere to strict independence as codified by the Sarbanes–Oxley Act of 2002[2] and the ForHumanity's Code of Ethics.[3] In addition, BABL AI

[1] ISAE 3000: Assurance Engagements Other Than Audits or Reviews of Historical Financial Information
[2] Sarbanes–Oxley Act of 2002
[3] ForHumanity Certified Auditor Code of Ethics

Lead Auditors are ForHumanity Certified Auditors under NYC AEDT Bias Audit.[4] For more details about our methodology and process, see Appendix – Audit Methodology.

Scope & Objective

Audit Section	Audit Objective
Disparate Impact	To perform testing of their model to "assess the tool's disparate impact on persons of any component 1 category," – i.e., race and gender – as the minimal requirement for a bias audit under Local Law 144 of 2021.

Out of Scope

1. The scope of the analysis included only race/ethnicity and gender; no other protected classes were covered.
2. The audit did not provide a certification that the model is "bias-free".
3. The audit explicitly excludes any evaluation of compliance with any other legislation or regulation other than NYC Local Law No. 144

[4] ForHumanity NYC Bias Audit

Findings

Note: *The information disclosed under each criterion is not documentary evidence.*

Disparate Impact

Audit Criterion & Subcriteria		
Q.A.		**System Definition & Analysis Setup:** The auditee shall clearly define and comprehensively describe the system and the methodology used for disparate impact analysis, including a justification for selecting the setup and any relevant assumptions or limitations.
	1.	**System Description:** Evidence shall show: ● The scope, purpose, nature, context of the system; and ● How the system is used in the employment context.
	2.	**Settings or Parameters:** Evidence shall describe: ● The system settings or parameters available to users that may affect system output; ● Their extents of user configurability; ● Their default values, where applicable; and ● Justification for why the default values were appropriate.
	3.	**Analysis Setup:** Evidence shall show: ● A description of the setup used to measure disparate impact; and ● Justification for why the selected setup is appropriate for disparate impact analysis.
	4.	**Settings in Analysis:** Evidence shall specify the values of the user-configurable settings or parameters identified in Q.A.2 that were used for the disparate impact analysis of this audit.
	5.	**Date of Analysis:** Evidence shall show that the most recent analysis was performed within one year of this audit's start date.
	6.	**Improvements:** If an audit of the system has been previously conducted by BABL AI, evidence shall describe improvements made to the disparate impact analysis since the last audit.

Testing conducted by: BABL AI Inc.
Date of most recent testing: Jun 2026
User-configurable settings that can affect system output: Candidate questionnaire
Settings on which disparate impact was tested: Historical candidate questionnaires from the past year of the system's operation.

Audit Criterion & Subcriteria		
Q.B.		**Dataset for Disparate Impact Analysis:** The auditee shall clearly define and comprehensively describe the dataset used for disparate impact analysis, including the justification for the relevance and representativeness of the dataset and any relevant limitations.
	1.	**Dataset Description:** Evidence shall show a detailed description of the dataset used for disparate impact analysis, including: • Composition; • Timeframe of data collection; • Collection process; and • Any processing steps.
	2.	**Representativeness & Relevance:** Evidence shall show justification for why the dataset is representative and relevant for disparate impact analysis.
	3.	**Demographic Data Collection:** Evidence shall describe the method by which demographic data was collected or generated.
	3.1.	**Inference of Demographic Data:** If demographic data was generated by inference, evidence shall: • Describe the inference method, and • Show justification for why this inference method was appropriate.

Time span of data: May 2025 – May 2026

Justification for the use of test data: While the candidate data used for testing consists of real applicants and their outcomes from the past year, the auditee did not have access to self-reported demographic data. Therefore, demographic inference was used on this candidate data for testing.

Description of the generation of test data: A variant of BIFSG[5] without geographic data was used to infer candidates' race/ethnicity. The U.S. Social Security Popular Baby Names dataset[6] was used to infer gender. These methods were applied to the names of real candidates.

Audit Criterion & Subcriteria	
Q.C.	**Demographic Categories & Groups:** The auditee shall specify the demographic categories and groups that are included in disparate impact analysis.

[5] https://www.science.org/doi/10.1126/sciadv.adc9824
[6] https://www.ssa.gov/oact/babynames/limits.html

	1.	**Demographic Categories:** Evidence shall specify demographic categories that are included in the disparate impact analysis, and shall show that, at least, those categories include race/ethnicity and gender.
	2.	**Gender Groups:** Evidence shall show that the demographic groups for gender include at least: "Male," and "Female".
	3.	**Race/Ethnicity Groups:** Evidence shall show that the demographic groups for race/ethnicity include at least White, Black or African American, Hispanic or Latino, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, and Two or More Races.
	3.1.	**Substituted Groups:** If the demographic groups for race/ethnicity do not include all categories listed in criterion Q.C.3, evidence shall show justification for why such demographic groups were not included, and, if applicable, justification for any substituted groups.
	4.	**Intersectional Groups:** Evidence shall show that intersectional groups include all permutations of gender and race/ethnicity groups.

Demographic categories included by the analysis:

1. Gender
2. Race/ethnicity

Audit Criterion & Subcriteria		
Q.D.		**Metrics for Disparate Impact Analysis:** The auditee shall appropriately define the metrics used for disparate impact analysis and define and justify the chosen metric for the context of this analysis.
	1.	**Selection Rate or Scoring Rate:** Evidence shall: • Specify whether the analysis was performed using selection rate or scoring rate, and • Define the selection rate or scoring rate as applied in the analysis.
	2.1.	**Positive Outcome:** If selection rate is used, evidence shall show: • The definitions of the positive and negative outcomes in the employment context, and • A justification for why such definitions are appropriate based on the context of the use of the system.
	2.2.	**Thresholds for Positive Outcome:** One or more thresholds are used to determine positive/negative outcome for selection rate, evidence shall show justification for why

		the level (levels) of threshold was (were) appropriate given the intended use of the system.

Method of quantifying disparate impact: Selection rate, the proportion of candidates who completed the vetting questionnaire[7] and received a pass verdict
Positive outcome: Completing the vetting questionnaire and receiving a pass verdict

Audit Criterion & Subcriteria		
Q.E.		**Disparate Impact Calculations:** The auditee shall calculate selection rates or scoring rates, impact ratios, for all demographic categories and groups and provide a justification explaining potential contributing factors if any impact ratio falls below 0.8.
	1.	**Results and Calculations:** Evidence shall show, for all demographic groups listed in criteria Q.C.2, Q.C.3, and Q.C.4: • The number of applicants or candidates; • Selection rates or scoring rates; • Impact ratios; and • That the calculations for selection or scoring rates, and for impact ratios are accurate.
	2.	**Unknown Groups:** If a gender, race/ethnicity, or intersectional group is not known for a sample of candidates assessed by the system, evidence shall show the sample size of such a group.
	3.	**Exclusion of Groups:** If a demographic group accounts for less than two percent (2%) of the total sample size of its respective demographic category, such group may be excluded from impact ratio calculation, but evidence shall nonetheless show the sample size, and the selection rate or scoring rate for such group.
	4.	**Uncertainty Analysis:** Evidence shall show results of uncertainty analysis of selection rates or scoring rates and impact ratios.
	5.	**Fourth-Fifths Rule:** If the impact ratio of a demographic group is below 0.8, evidence shall provide a justification based on the potential sources of such outcome.
	6.	**Statistical Significance:** If selection rate is used, evidence shall show, for all demographic groups, calculations of statistical significance of the difference between the selection rates of two groups.

[7] Candidates who dropped off or otherwise did not complete the vetting questionnaire were not included in the analysis to avoid introducing additional noise to the test results.

Non-intersectional, Gender, sorted by Selection rate

	N applicants	Selection rate	Impact ratio
Male	343,292	0.673	1.000
Female	365,116	0.667	0.991

Non-intersectional, Race/ethnicity, sorted by Selection rate

	N applicants	Selection rate	Impact ratio
White	260,455	0.687	1.000
Asian	19,459	0.557	0.810
Black or African American	117,000	0.686	0.997
Hispanic or Latino	132,267	0.634	0.922

Intersectionals

			N applicants	Selection rate	Impact ratio[8]
Hispanic or Latino	Male		62,880	0.644	0.927
	Female		58,686	0.626	0.902
Non-Hispanic or Latino	Male	White	119,673	0.687	0.990
		Asian	7,769	0.561	0.808
		Black or African American	28,056	0.694	1.000
	Female	Asian	7,484	0.552	0.795
		White	121,019	0.691	0.996
		Black or African American	51,118	0.680	0.979

[8] N/A refers to the demographic group representing less than 2% of the total N applications in the table. Numbers in red indicate values below the four-fifths rule.

Note: Data on these applicants was not included in the calculations above:

1. 109,260 applicants with an unknown gender category
2. 288,487 applicants with an unknown race/ethnicity category, and
3. 348,637 applicants with at least an unknown gender or an unknown race/ethnicity

Response from LANDED on impact ratios:

"LANDED has reviewed the disparate impact results. All gender impact ratios and all non-intersectional race/ethnicity impact ratios meet the four-fifths (0.80) threshold. A single intersectional group, inferred Asian female, falls marginally below the threshold, at 0.795. Consistent with criterion Q.E.5, LANDED offers the following context on potential contributing factors for this result. Consistent with criterion Q.E.5, LANDED offers the following context on potential contributing factors:

First, these results are based on inferred rather than self-reported demographics. As noted in the Emphasis of Matters, LANDED does not collect candidates' race, ethnicity, sex, or gender. The labels used here were inferred from candidate names (a BIFSG variant for race/ethnicity and U.S. Social Security baby-name data for gender), which carries inherent uncertainty. The auditor's sensitivity analysis indicates the results are sensitive to the inference method, and the 0.795 value sits within approximately 0.005 of the threshold and well within the range of that uncertainty.

Second, the system does not use protected characteristics. The model is not provided with, and does not score on, race, ethnicity, sex, or gender.

Third, vetting criteria are defined by each client employer, not by LANDED. Clients may configure requirements such as U.S. work authorization or English-language fluency based on the legitimate needs of their specific roles. LANDED's system evaluates candidates against these employer-defined criteria. Where differences in selection rates reflect those job-specific requirements, they are attributable to the employer's hiring criteria rather than to any bias in the model's evaluation.

LANDED considers these results to be within an acceptable range and does not regard them as evidence of unfair bias in the system. LANDED remains attentive to fairness in its products and continues to evaluate potential safeguards, including approaches that would prevent the model from accessing candidate names during evaluation."

Appendix

Audit Methodology

The Criterion Audit

The BABL AI audit framework is the *Criterion Audit Framework*,[9] defined as "a criteria-based independent external evaluation of an algorithmic system conducted by an auditor to determine whether the given system meets the requirements set by a normative framework." A criterion audit is modeled after the financial auditing practice, and is distinguished from other commonly used forms of assessment of algorithms, such as internal audits, critical third-party audits, and risk or impact assessments. The audit framework as performed as a direct engagement contains three main phases:

1. **Scoping** – The auditor gathers comprehensive details about the tool, including components, decision contexts, and available data. The auditor proposes a robust methodology for bias testing of the system which is then formally reviewed and approved by the auditee.
2. **Testing** – The auditor conducts independent testing and calculations adhering to the agreed-upon methodology. The auditor retains audit evidence to demonstrate adherence with the audit criteria.
3. **Reporting** – The auditor issues a formal report containing the bias testing results and the audit opinion.

Terminologies & Definitions

Term	Abbrev	Definition
automated employment decision tool	AEDT	"any computational process, derived from machine learning, statistical modeling, data analytics, or artificial intelligence, that issues simplified output, including a score, classification, or recommendation, that is used to substantially assist or replace discretionary decision making for making employment decisions that impact natural persons." – see § 20-870 of the Code and § 5-300 of the adopted rule for full definition

[9] Lam, K., Lange, B., Blili-Hamelin, B., Davidovic, J., Brown, S. & Hasan, A. (2024). A Framework for Assurance Audits of Algorithmic Systems. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24. ACM, June 2024. doi: 10.1145/3442188.3445924.

Term	Abbrev	Definition
disparate impact or adverse impact		"a selection rate for any race, sex, or ethnic group which is less than four-fifths (⅘) (or 80%) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact" – see § 60-3.4.D of UGESP (1978) for full definition
impact ratio		"either (1) the selection rate for a category divided by the selection rate of the most selected category or (2) the scoring rate for a category divided by the scoring rate for the highest scoring category. " – see § 5-300 of the adopted rule for full definition
scoring rate		"the rate at which individuals in a category receive a score above the sample's median score, where the score has been calculated by an AEDT"
justification		a compelling reason that illuminates the issue and carries normative force, as opposed to solely explanatory power
positive outcome		the basis for selection rate, the favorable outcome for a candidate from the use of the model, such as being selected to move forward in the hiring process or assigned a classification by an model
selection rate		"the rate at which individuals in a category are either selected to move forward in the hiring process or assigned a classification by an AEDT" – see § 5-300 of the adopted rule for full definition
testing dataset		the dataset used to test for or quantify disparate impact
uncertainty analysis		calculation or computation to quantify the uncertainty of a variable, outputting errors or error bars

Audit trail

Details

FILE NAME	2026_Landed_Public Summary - 6/26/26, 2:21 AM
STATUS	Signed
STATUS TIMESTAMP	2026/06/25 19:21:50 UTC

Activity

SENT	khoalam@bablai.com **sent** a signature request to: • BABL AI Inc. (khoalam@bablai.com)	2026/06/25 19:21:19 UTC
SIGNED	**Signed** by BABL AI Inc. (khoalam@bablai.com)	2026/06/25 19:21:50 UTC
COMPLETED	This document has been signed by all signers and is **complete**	2026/06/25 19:21:50 UTC

The email address indicated above for each signer may be associated with a Google account, and may either be the primary email address or secondary email address associated with that account.