**Holistic AI questions regarding the DCWP’s updates to Proposed Rules to implement Local Law 144 of 2021 (Automated Employment Decision Tools)**

January 20, 2023

Department of Consumer and Worker Protection
DCWP Commissioner Vilda Vera Mayuga

42 Broadway, New York

NY, 10004

**RE: Proposed Rules to implement Local Law 144, Automated Employment Decision Tools**

Dear Commissioner of the New York City Department of Consumer and Worker Protection (DCWP),

Thank you for the opportunity to provide questions on this important matter.

**1. About Holistic AI**

Holistic AI is an AI Governance Risk and Compliance (GRC) company, with a mission to empower enterprises to adopt and scale AI with confidence. Holistic AI has a multidisciplinary team of AI and machine learning engineers, data scientists, ethicists, business psychologists, and law and policy experts.

We have deep practical experience auditing AI systems, having assured over 100 enterprise AI projects covering more than 20,000 different algorithms. Our clients and partners include Fortune 500 corporations, SMEs, governments, and regulators. We work with several companies to conduct independent bias audits, including in preparation for Local Law 144.

**2. Key questions**

We would like to clarify some points regarding updates to the Proposed Rules and their implementation.

**2a. Calculating Impact Ratios with Small Sample Size**

The Proposed Rules specify that the ethnicity/race categories that should be examined are Hispanic or Latino, White, Black or African American, Native Hawaiian or Pacific Islander, Asian, Native American or Alaska Native, or two or more races. However, there are likely multiple categories with small samples, particularly for the Native Hawaiian or Pacific Islander, Native American or Alaska Native, and two or more races categories. However, the DCWP does not provide any clarification on what is considered an adequate sample size for analysis to be meaningful. Notably, examples provided by the updated Proposed Rules keep the categories separate, with one category listed representing less than 1.5% of the workforce, but the EEOC’s clarifications on the Uniform Guidelines[[1]](#footnote-2) specify that adverse impact analysis should only be carried out for groups who represent at least 2% of the labor force, meaning that the examples provided conflict with this guidance. Additionally, analyses based on samples representing less than 2% of the population are unlikely to be meaningful.

One approach to increase sample sizes is to combine the Native Hawaiian or Pacific Islander or Native American or Alaska Native categories into one broader “Other” category[[2]](#footnote-3). However, there is no guarantee that this will increase the sample to a sufficient size for a robust analysis and could mean that it is harder to identify and mitigate bias for particular subgroups if they do not have their own category.

We therefore request that the DCWP clarify what is considered an adequate sample size for analysis to be meaningful and to release guidance on calculating impact ratios when sample sizes are small or propose an alternative metric that is more suited to smaller samples. Such guidance would be particularly useful for intersectional analyses, where sample sizes are often small.

**2b**. **Calculating Impact Ratios for Regression Systems**

While the revised metric for calculating the impact ratio for regression systems using the median is a notable improvement over the initially proposed metric using the average score, this metric still has some concerning limitations. Most notably, the impact ratio metric is not always adequate for detecting bias in regression data.

*Example 1: Bimodal Distribution vs Unimodal Distribution*

Let’s consider the case where male and female candidates are scored from 0 to 100. Whereas male candidates consistently get scores around 50 (unimodal distribution), female candidates seem to be scored either approximately 25 or 75 (bimodal distribution), as seen in the figure below. The median of the full dataset (across males and females) is 50, because half the data falls below it and the other half falls above it.



If all candidates scoring above the median value of 50 are hired, then the data will be perfectly fair. However, if only the top 20% of candidates are hired, then almost all chosen candidates will be female. Given that the median is rarely used as a cutoff score, the system is likely to result in biased outcomes even if the audit does not find any evidence of bias using the median metric.

*Example 2: Two Bimodal distributions*

In this example, male candidates are scored either approximately 30 or 70, whereas female candidates are scored either approximately 30 or 80. In this example, seen in the figure below, there are two peaks for each, with the lower one being consistent for both male and females while the higher one is slightly different. Like in the above example, the median of the full dataset is approximately 50, because half the data falls below it and the other half falls above it.



If all candidates scoring above 50 are hired, then the data will be perfectly fair. This is because we are essentially reducing our data to a binary pass/fail classification, and the difference in the higher scores for male and females will be masked. However, as soon as our notion of success changes, big differences are revealed. To better observe this phenomenon, we can calculate how the impact ratio varies when the candidates hired are respectively the top 50% (which is equivalent to the median), 40%, 30%, 20% and 10%.



As is seen from the figure above, the computed binary disparate impact will greatly depend on the threshold we use. At the median value, we obtain perfect fairness. For any values above the median, the fairness rapidly decreases due to the distribution of the data.

Given these concerns, we encourage the DCWP to consider alternative metrics that would be better suited for measuring regression bias. For alternative metrics, see Holistic AI’s open-source library.[[3]](#footnote-4)

**3. Holistic AI resources**

In lieu of the fact that the field of algorithm audit and assessment is relatively new, below we link some resources and references to our open source and academic research.

* [**Holistic AI Open Source**](https://www.holisticai.com/open-source)
* [**The New York City Bias Audit Law: Regulating AI and automation in HR**](https://www.holisticai.com/blog/whitepaper-nyc-bias-audit)
* [**Towards Algorithm Auditing: A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms**](https://www.holisticai.com/papers/towards-algorithm-auditing)
* [**Systematizing Audit in Algorithmic Recruitment**](https://www.holisticai.com/papers/systematizing-audit-algorithmic-recruitment)
* [**Perceived Fairness of Algorithmic Recruitment Tools**](https://www.holisticai.com/papers/robots-judging-me-perceived-fairness-algorithmic-recruitment-tools)
* [**Overcoming Small Sample Sizes When Identifying Bias**](https://www.holisticai.com/blog/bias-audit-impact-ratio-small-sample)

**4. Concluding statement**

Holistic AI welcomes the opportunity to provide comments on this important matter. We appreciate the open, transparent and collaborative approach taken by the DCWP.

We support the important objectives of Local Law 144. We stand ready to support the DCWP, the New York City Council or other public authorities involved in the implementation and enforcement of this important law.

Please contact we@holisticai.com for any further information or follow-up on this submission.

Sincerely,

Holistic AI

<https://www.holisticai.com/>

1. https://www.eeoc.gov/laws/guidance/questions-and-answers-clarify-and-provide-common-interpretation-uniform-guidelines [↑](#footnote-ref-2)
2. https://www.eeoc.gov/data/introduction-race-and-ethnic-hispanic-origin-data-census-2000-special-eeo-file [↑](#footnote-ref-3)
3. https://www.holisticai.com/open-source [↑](#footnote-ref-4)