BABL AI INC.
The Algorithmic Bias Lab
630 Fairchild Street
Iowa City, Iowa 52245
https://babl.ai

From: **Shea Brown, Ph.D.**
Chief Executive Officer
BABL AI Inc.
sheabrown@bablai.com

To: **NYC Department of Consumer and Worker Protection**
Rulecomments@dca.nyc.gov

Re: **Public Comments on the Proposed Rules for Local Law 144 of 2021**

Oct. 23, 2022

To Whom It May Concern:

On behalf of the team at BABL AI, I welcome the opportunity to provide public comments on the proposed rules for Local Law 144 requiring annual bias audits for automated employment decision tools (AEDTs).

We thank the Department for providing guidance and clarity on many of the terms provided in the original law. As a company that audits algorithms for ethical risk, effective governance, bias, and disparate impact, BABL AI believes that the spirit of this law furthers our mission to promote and protect human flourishing in the age of AI. While the new rules directly address many of our previous concerns, we would like to comment on several areas that remain unclear and that pose barriers for companies wishing to make good-faith efforts to comply.

AEDT Definition: We find that the clarified definition of an AEDT is overly stringent. In its proposed form it allows for serious loopholes for employers to seek exemption from bias audits. To illustrate, a user of an ML-based assessment of personality traits, cognitive ability, or cultural fit whose results are shown to recruiters as part of a candidate profile could claim exemption from the bias audit, arguing that this assessment does not overrule

or modify recruiter decision making given that there are other factors recruiters also consider. While this may be true for some use-cases, our experience has shown that this is only determined through a careful risk/impact assessment. Given that many such ML systems tend to contain high risk of algorithmic bias, we view their potential exclusion as risky. Therefore, **we strongly encourage the department to provide a recalibrated definition of an AEDT.**

Independence: We are concerned that this new definition of independence is not sufficiently strong to provide the level of impartiality as desired by the law. There are two reasons for our concern:

1. This definition allows for internal parties with conflict of interest for a vendor or employer (e.g., operating under the same level of management) to conduct the audit which obfuscates impartiality altogether, and
2. This definition shifts the discourse on independence as a legal dispute to a semantic debate on the terms "using" and "developing."

Moreover, the term "developing" carries a dual meaning in the AI fields:

1. A narrow scope as in "the development phase" of the AEDT where the AI or model is first built, trained, and tested prior to deployment, and
2. A broad scope where continuous design, procurement, training/testing, deployment, and monitoring are all parts of AEDT development.

As a result, the number of parties involved in *developing* depends on its meaning and scope. For example, with the emergence of companies whose primary focus is on providing monitoring services for AI solutions and whose monitoring data is looped back to AEDT vendors to improve the tool, it is ambiguous whether these companies qualify as "group involved in developing" the AEDT.

Due to these ambiguities, **we encourage the department to either consider a different approach to defining independence—such as using a precedent as in the Sarbanes–Oxley Act of 2002 or provide special clarification for the terms "using" and "developing".**

Intersectional Analysis: While conventional disparate impact analysis as carried out by the EEOC does not require analysis at the intersectional level, the example shown in the amendments disclose selection rates and impact ratios for intersections of groups (e.g., white male vs. Hispanic female). We applaud the department for taking intersectional analysis into consideration, as we believe it provides better transparency and stronger protection for protected groups.

However, the proposed rules say "[t]hese calculations are consistent with Section 1607.4 of the EEOC Uniform Guidelines on Employee Selection Procedures," which are typically not calculated on an intersectional basis. **We recommend the department to explicitly clarify whether an intersectional breakdown for selection rates and impact ratios is required**.

The explicit clarification would have huge ramifications, as many AEDTs are monitored or tested prior to deployment, using the fourth-fifths rule as a reference point, but many are not controlled for *fairness* at the intersectional level.

Source of Testing Data for AEDTs: It is unclear from the proposed rules how a newly developed AEDT would be audited prior to deployment without access to historical demographic data. The two examples in the amendments show only historical data from an employer requesting for the bias audit who either has used the AEDT or plans to use the AEDT, with the assumption that such historical data exists for the tool.

This assumption is challenged by the way new AEDTs are currently developed, where many tools using ML/AI require new sources of data from candidates which can be bespoke to an employer, or whose historical data does not exist from that specific employer or from any employer. Examples of such data include: performance data of candidates playing games, audio/video/written/texting samples of candidates answering interview questions, or facial photos of candidates (e.g., in IDs, cameras, or webcams).

For many AEDTs, employers and vendors are looking to demonstrate compliance via alternative sources of data, such as using synthetic data or simply using the testing set of their training data. **We urge the department to clarify whether the use of synthetic data or training/testing data would be sufficient for employers and vendors to comply with the law.**

1. If so, we encourage the department to also provide requirements to circumvent issues where this data is not representative of the data during use—i.e., out-of-distribution sample, data drift, or
2. If not, the department should clarify explicitly and provide guidance for newly developed AEDTs where historical data is not available.

Scope of Bias. The scope of bias audit currently only includes disparate impact or impact ratio analysis. We are concerned that this scope would only address the technical and model-centric view of bias. Our work in conducting risk and bias assessments for these automated systems have shown us that biased outcomes can often arise from a variety of sources outside of the development of the tool, such as in the way employers configure the settings of the tool or in the way direct users of the AEDTs interpret simplified outputs. Therefore, we fear that the current audit does not address these sources of bias and does little to prevent harms despite the tool itself being supposedly controlled for bias.

While it is difficult for current proposed rules to circumvent all sources of non-technical bias, we have found that risk assessment is a strong AEDT-agnostic method to uncover biases beyond the technical aspects, and allow employers and vendors to acknowledge and confront these biases head-on. Furthermore, emerging legal and industry frameworks have universally proposed risk and impact assessments as effective tools for detecting and mitigating risk, including the risk of disparate impact. Therefore, **we encourage the**

department to consider the inclusion of some minimal elements of risk or impact assessment in the audit to control for non-technical bias.

I would like to thank the NYC Department of Consumer and Worker Protection for providing us the opportunity to comment on the Proposed Amendments of Local Law 144 of 2021, and we would be happy to provide further clarification on any of the above comments.

Contact

Shea Brown, Ph.D., CEO & Founder
sheabrown@bablai.com