Written Comments from Merve Hickok
Regarding Proposed Rules on NYC Local Law 144 of 2021 in relation to Automated Employment Decision Tools

TO: New York City Department of Consumer and Worker Protection (DCWP)
Rulecomments@dca.nyc.gov 10/23/2022

Dear Chair and Members of DCWP,

As the Founder of AIethicist.org and Lighthouse Career Consulting LLC, I welcome the public comment opportunity for the rules proposed by DCWP on implementation of Local Law 144 of 2021, regulating automated employment decision tools (AEDT).[1] Accordingly, I would like to 1) provide feedback on the proposal, and 2) request further clarification regarding the proposed rules for the members of DCWP to consider. My work is focused on Artificial Intelligence (AI) ethics and AI policy and regulation globally. I am also a certified human resource professional with almost two decades of experience across Fortune 100 companies. As the founder of AIethicist.org, I provide research, training, and consulting on how to develop, use and govern algorithmic systems in a responsible way. I conduct research and training on AI policy and regulation at Center for AI & Digital Policy and teach data science ethics at University of Michigan. I submitted written comments to DCWP's previous rulemaking.[2]

I again congratulate New York City Council on the passing of this very important and impactful law. The following questions and recommendations build upon my previous written comments to DCWP and would help protect individual's rights and make the expectations from the employers and vendors clearer. In return, it would make it easier to detect the violations, and prevent different interpretations of the requirements. I also recently published two policy briefs on AI & Recruitment systems relevant to DCWP's work.

- First brief provides an in-depth analysis of NYC Law 144, including the benefits, obligations, limitations, and impact on employers, vendors, and candidates.[3]
- The second brief provides a landscape analysis of AI policy and regulations in employment decisions globally.[4] The brief reviews the city, state, federal and international across multiple jurisdictions and provides comparisons.

Definition of 'independence': The Proposal suggest Independent Auditor means "a person or group that is not involved in using or developing an AEDT that is responsible for conducting a bias audit of such AEDT." This proposal suggests the auditor could be internal to the vendor or the employer. However, the wording opens some other questions.

[1] Local Law 144 of year 2021: http://nyc.legistar1.com/nyc/attachments/c5b7616e-2b3d-41e0-a723-cc25bca3c653.pdf
[2] Merve Hickok (June 6, 2022). *Written Comments regarding NYC legislation on Automated Employment Decision Tools* (Local Law 144), submitted to New York City Department of Consumer & Worker Protection. https://rules.cityofnewyork.us/rule/force-fed-products-open-captioning-in-motion-picture-theaters-and-automated-employment-decision-tools/
[3] Center for AI and Digital Policy (August 11, 2022). *Policy Brief – NYC Bias Audit Law*. https://www.caidp.org/app/download/8407232163/AIethicist-NYC-Bias-Audit-Law-08112022.pdf
[4] Center for AI and Digital Policy (September 2, 2022). *Policy Brief - State of AI Policy and Regulations in Employment Decisions.* https://www.caidp.org/app/download/8410870963/AIethicist-HumanResources-RecentDevelopments-09022022.pdf

1) What is meant by 'involved in using or developing an AEDT'? For example, if the vendor/employer's internal audit teams use the tool to hire for their own department, can they be independent? If the vendor's internal audit department conducts regular audits of the development teams, can they be independent? If an employer's internal auditor was involved in procurement due diligence, can they be independent?

A clearer criterion is needed to determine independence if the auditor is to be internal to either the vendor or the client.
It is equally important that a lawyer-client privileged relationship is not used for audit purposes. Such a relationship cannot be independent. It can also have negative consequences for DCWP and impacted parties and their access to adequate transparency on AEDTs when needed.

2) In the absence of any AEDT bias audit accreditation scheme, it is crucial that vendors do not decide by themselves which historical dataset to audit, conduct the audit and then decide on what to report – all in a closed circuit.

Therefore, the scope and rules of audit must be developed independent of the internal auditor conducting the audit to prevent conflict of interest and ensure integrity of audit.

Definition of "Machine learning, statistical modelling, data analytics, or artificial intelligence": Proposal defines three conditions for a mathematical/computer-based technique to be considered under this definition and proposes "and" between each of these conditions. Is it the intent of DCWP that a technique should meet all three criteria to be considered as subject to Law 144? If yes, this change significantly compromises the scope and intent.

Definition of "Simplified Output": Initial definition of AEDT in 20-870 of the Code was reflective of the intended scope of NYC Council's decision. However, narrowing the definition by using the qualifiers like "rely solely on a simplified output" or "with no other factors considered" significantly compromises the scope and intent. I welcome the "examples" provided in Proposal such as prediction, classification, score, tag, categorization, recommendation, or ranking. However, recommend DCWP not include any qualifiers and narrow the scope.

Definition of 'bias': Responsible AI practices and consideration of bias must be across practical and statistical tests, and across design decisions. The assumptions, design/model decisions and trade-offs made by vendor and employers must be included as audit criteria. For example, different cut-off scores used for outputs may change the results significantly across different groups.
Alternatively, a 'culture fit' criteria embedded in the system can discriminate against certain groups behind the veil of technological objectivity.

Proposal suggests that a simple Impact Ratio analysis can be enough to satisfy the requirement of this law. IR is just a rule of thumb and only one of the methods used to calculate disparate impact. If this is the only method suggested by DCWP, there should also be a guidance for those job categories audited where the sample size is small enough for the proposed IR calculation to be irrelevant.

Data used for analysis of Impact Ratio: Proposal, in one of the examples, suggests "vendor uses historical data it has collected from employers on applicants selected for each category to conduct a bias audit." Like above, there is need for further clarification on the dataset used for audit.

1) Vendors need to be equally responsible and transparent to ensure the data and model(s) do not result in disparate impact.

- If a vendor uses a general model and does not tailor / train models with client's own data, an aggregate dataset made up of all clients' historical data could suffice for a bias audit & report.
- However, if the vendor tailors its model and/or trains model with client employer's data, there is a possibility for the outcomes to vary across different employers. If an employer is liable for the outcomes of an AEDT model trained with its own data, then running an audit on the aggregate data of many employers may not be reflective of what each employer is using. Aggregation may not correctly reflect impact ratios for different employers. This necessitates separate audits for each model in use.

For (b), solution would be for the vendor to conduct 2 separate analyses: a) Impact Ratio for all candidates in the system, regardless of clients, AND b) a separate client-based analysis. Since the Proposal defines Bias Audit as a simple Impact Ratio analysis, and since the vendors should already be monitoring client models for quality and liability purposes, this should not be an extra burden. This approach would also give more confidence to employers regarding the safeguards.

2) For each audit, the data should also be analyzed for different 'job categories." For example, AEDT outcomes might be biased for females in administrative jobs with traditionally female hires, or vice versa for executive or technical jobs with traditionally male hires. If multiple job categories (which require different skills, traits) are combined for a single analysis, the aggregate dataset may not correctly reflect the possible biases in hiring decisions.

Notice to Candidates: The law requires employers to provide candidates minimum 10-day notice about the future use of AEDT and include the specific job qualifications and characteristics the tool will use in the assessment of candidate. This is to allow a candidate to request an alternative selection process or accommodation. To be able to request an accommodation or alternative, a candidate will need to know if the tool will require the use of and/or assess, for example, any physical, cognitive, or motor skills, or mental, emotional, character capabilities or competencies. DCWP Proposal omits this last part in the notice requirement.

Published results of the most recent bias audit: I strongly support the proposal that public AEDT audit summary results to "include the selection rates and impact ratios for all categories." This is crucial for the success and impact of Local Law 144.
As mentioned in 'Notice' section, I also propose the public audit summary to also include "characteristics the tool will use in determining its outcome." This could be the same as the notice provided to candidates during their application process. Such transparency would magnify the intent of the Law.

Thank you for your consideration of my views. I would welcome the opportunity to discuss further about these recommendations.

Merve Hickok, SHRM-SCP
Founder, AIethicist.org, and Lighthouse Career Consulting LLC
merve@lighthousecareerconsulting.com