Jiahao Chen, PhD
Owner, Responsible Artificial Intelligence LLC
450 Lexington Ave, Unit 609
Manhattan, New York, 10163-9619
jiahao@responsibleai.tech

January 23, 2023

Department of Consumer and Worker Protection
New York City
42 Broadway
Manhattan, New York 10004-1617
Email: rulecomments@dcwp.nyc.gov

Dear committee members:

I am pleased to submit comments for *Requirement for Use of Automated Employment Decisionmaking Tools* (Ref. No. DCWP-21, dated December 15, 2022; "The Updated Rules"). Having previously submitted comments on October 23, 2022 (many of which are still relevant to the current version of the Rules), I would like to focus my comments here on just the new aspects of the Updated Rules.

1. Independent Auditor. The current definition of Independent Auditor could still use some clarification.

On one hand, the current definition of "financial interest" may be too restrictive - an auditor hired by a solutions vendor or employer and paid a fee for the audit could ostensibly be considered having a financial interest in the auditee, since the auditor has a vested interest in sustaining a business relationship in anticipation of future audit opportunities, and hence sustained revenue from audit fees. If such an interpretation were to be enforced so strictly, there would be no business opportunity for for-profit algorithmic auditors. Therefore, clarification would be useful on what constitutes "financial interest" when it comes to compensation for performing audit services, or for a sustaining agreement for such future services.

Conversely, an independent auditor meeting the current definitions in the Updated Rules may still be prevented from producing an objective and impartial audit. For example, a contract for auditing may restrict the tests the auditor can perform and report, and may require pre-clearance for publication to assess reputational risks or other disadvantageous disclosures, or even that the audit be conducted partially or fully under attorney-client privilege, which would undermine the goal of transparency in such audits. If auditees are allowed to exercise control over what is reported from an audit, they may cherry-pick the reported tests beyond the minimum requirements to show only those that portray the auditee positively, and suppress other test results that may be derogatory. There is therefore a need to clarify what restrictions may or may not be placed on auditors and their auditing contracts to prevent the gaming of the audit requirements, and the potential for ethics-washing by hiring auditors that are independent in name only.

2. Risks from reporting the statistics of small numbers. The example provided in the Updated Rules highlights clear privacy and statistical validity risks that could result from the audit, which I had described as a potential problem in my previous comments. The intersectional category of Female / Native American or

Jiahao Chen, PhD
Owner, Responsible Artificial Intelligence LLC
450 Lexington Ave, Unit 609
Manhattan, New York, 10163-9619
jiahao@responsibleai.tech

Alaska Native shows that 7 of 17 candidates were selected, with a corresponding impact ratio of 0.789. A statistically naive auditor may be tempted to report this ratio as falling below the commonly-used threshold of 0.8 and hence be a potential concern for discriminatory risk. However, it could be argued that with just 17 people in this category, the computed ratio is not statistically robust.

There is also a privacy risk inherent in reporting results from small numbers. In the extreme limit, an audit may report that 0 out of 1 applicants in some demographic category were hired. This means that if we happen to know this applicant, we can immediately know from the audit that the application was not hired. In this way, the audit may represent an unintended way to leak information about an employment decision. Such issues have been well studied, particularly in the compilation of census data, where such concerns are of high priority. Current best practices for mitigating such privacy risks include differential privacy and omitting the reporting of very small categories.

In summary, tests of statistical validity, such as reporting the level of significance in hypothesis tests, should be included in the minimum requirements for reporting; otherwise, there should be provisions to omit reporting for very rare categories that lack sufficient representation to provide meaningful statistics.

3. Construct validity of the AEDT. As described in my previous comments, it is vital to test that systems to make employment decisions are made on reasonable criteria, as already required under U.S. Federal employment laws like the Equal Employment Opportunity Act. An employment examination that does not test skills necessary to perform a particular job would not be considered valid for use in hiring decisions for that job. Such requirements ought to extend to data-driven decisions too. Under current regulations for fair lending compliance, a bank would not be permitted to approve or deny credit based on data that have no bearing on creditworthiness, such as the color of the applicant's car or the applicant's taste in music. By analogy, a hypothetical employer may simply perform a coin toss to decide who to hire. Such a decision-making process is trivially unbiased since there is no risk of preferential discrimination on a prohibited basis, but is arguably unfair since no qualifying characteristics of the applicant are considered. Such risks are endemic to AI systems trained on bad data, even in the presence of cross-validation or other validation techniques. Therefore, all the discrimination testing in the world can still obscure the fact that an AEDT may simply be executing a very expensive coin flip, disguised by good performance on cherry-picked test data. Therefore, there is ample precedent for my recommendations, which are:

A. to broaden the definition to AEDT to any data-driven decision-making system (whether or not it is executed on a computer or by hand; the medium of computation should be irrelevant); and
B. to have algorithmic audits of AEDTs include tests for whether the data used for its decision making has sufficient predictive signal to enable useful decision making.

4. Data collection and retention requirements for discrimination testing. As described in my previous comments, collecting accurate demographic labels of applicants and employees will be vital for proper

Jiahao Chen, PhD
Owner, Responsible Artificial Intelligence LLC
450 Lexington Ave, Unit 609
Manhattan, New York, 10163-9619
jiahao@responsibleai.tech

testing. However, the Updated Rules do not define minimum requirements for collecting and keeping such records on prohibited bases. Without such specifications, an employer has no incentive to improve their collection of demographic information, which may have downstream privacy and security risks on collected data, and obscuring the extent of discrimination on candidates and employees whose racial/sexual/etc. identities are unknown to the employer. Furthermore, if such records were not properly kept, it would be difficult to validate an audit if its results were significant enough to trigger further enforcement action from a regulator. Therefore, the Rules ought to specify:

A. the minimum efforts employers are required to spend on collecting information on race, sex, age, etc.;
B. acceptable usage of such records, such as restricting their use solely to auditing; and
C. how records should be kept and retained for any further investigation.

Once again, I would like to congratulate the City for its progressive innovation toward eradicating employment discrimination in the age of data-driven decision-making, and encourage the City to consider these comments (in addition to those previously submitted) to maximize the intended effect of the Law toward building a more equitable employment market in the City. Please do not hesitate to reach out if I may be able to provide further clarifications on these comments.

Yours sincerely,

Jiahao Chen, Ph.D.
Owner
Responsible Artificial Intelligence LLC