

Testimony of Manish Raghavan before New York City Council Committee on Technology regarding Int. 1894

Manish Raghavan

November 9, 2020

Dear members of the committee,

My name is Manish Raghavan. I am a researcher at Cornell University studying the societal impacts of algorithmic decision-making, particularly in the context of hiring. I have extensively studied the types of automated employment decision tools being discussed today, and my testimony is largely based on this research.¹ In this testimony, I offer my recommendations regarding Int. 1894, which seeks to regulate algorithmic tools deployed for candidate evaluation.

I appreciate the Council’s attention on this important topic. Automated employment decision tools are increasing in prevalence, often with little to no public transparency into their inner workings. In my view, this bill is a step in the right direction. In its current form, it carries some vital provisions to ensure that automated hiring tools are carefully scrutinized for potential discrimination.

At the same time, it’s important to recognize the limitations of this bill (and indeed, any attempt to regulate these tools through prospective auditing). In this testimony, I will detail two such limitations:

1. Current interpretations of anti-discrimination law do not preclude all discriminatory behavior that algorithms can exhibit.
2. Audits have limited power to detect discrimination in terms of undisclosed attributes, such as sexual orientation or disability status.

Before diving deeper into these points, it’s important to note that hiring tools can perpetuate discrimination even in the absence of explicit bad actors. Due to historical patterns of inequity, algorithms can behave in discriminatory ways simply due to negligence or insufficient attentiveness to these issues. It’s crucial that we implement guardrails that protect us from these more subtle, insidious forms of discrimination.

¹Raghavan, Manish, et al. “Mitigating bias in algorithmic hiring: Evaluating claims and practices.” Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. 2020; Raghavan, M., and S. Barocas. “Challenges for mitigating bias in algorithmic hiring.” Brookings. 2019.

Current interpretations of anti-discrimination law do not preclude discriminatory behavior. Vendors of automated employment decision tools, to the extent that they consider issues of bias at all, typically think of anti-discrimination law in terms of the EEOC’s 4/5 rule. The 4/5 rule requires that applicants from different protected groups be selected at roughly the same rate—that is, if half of the candidates evaluated are women, then approximately half of the candidates selected by the tool should be women. A violation of the 4/5 rule do not necessarily constitute discrimination, but it can be the basis to open a discrimination suit.

In the absence of specific requirements, it is natural that bias audits will focus on ensuring that tool in question satisfies the 4/5 rule. In my view, this is insufficient, and inconsistent with standards in industrial-organizational psychology.²

A particularly important metric to consider is validity, which measures how good a tool is at correctly identifying high- vs. low-performing candidates. How is validity related to bias? One key way in which algorithmic tools can discriminate is via **differential validity**, which occurs when a tool is better at evaluating members of one group than another. For example, if the tool is very good at identifying the top-performing white candidates and not very good at identifying the top-performing African-American candidates, this would be an instance of differential validity.

Even if an assessment satisfies the 4/5 rule, meaning it recommends candidates from all racial groups at roughly equal rates, the top-performing African-American candidates would be more likely to be screened out by the assessment than their white counterparts. Differential validity has been repeatedly found in practical applications of data-driven decision-making,³ and it’s important to ensure that employment decision tools don’t perpetuate this form of discrimination.

Assessments that exhibit differential validity are not explicitly illegal, according to current interpretations of the law. However, simply **requiring an auditor to report on measures of differential validity** may induce vendors of automated employment decision tools to ensure that their products work well for everyone, not just those who have been well-represented in historical data. In my view, testing whether a tool performs well across the entire population should be an integral part of any bias audit, and to this end, I believe this bill should explicitly require differential validity testing.

Audits have limited power to detect discrimination in terms of attributes like sexual orientation or disability status. Audits can only be performed with respect to protected attributes on which vendors maintain data. If a vendor doesn’t collect data about, say, applicants’ sexual orientation, it is impossible for an auditor to know whether a tool produces disparities along these attributes. Nor is it necessarily desirable that vendors maintain this sort of sensitive data; applicants may not feel comfortable divulging this information.

²Society for Industrial, Organizational Psychology (US), and American Psychological Association. Division of Industrial-Organizational Psychology. “Principles for the validation and use of personnel selection procedures.” 2018.

³Buolamwini, Joy, and Timnit Gebru. “Gender shades: Intersectional accuracy disparities in commercial gender classification.” Conference on fairness, accountability and transparency. 2018; Koenecke, Allison, et al. “Racial disparities in automated speech recognition.” Proceedings of the National Academy of Sciences 117.14 (2020): 7684–7689.

Thus, an audit cannot identify all forms of illegal discrimination, and as such, it's important to be clear on the goals of such an audit. The current language of Int. 1894 refers to compliance with “any . . . applicable law relating to discrimination in employment.” In practice, this will not be possible. We should **acknowledge the narrow scope of what is possible through audits**, and what forms of discrimination cannot be detected through these means.

Recommendations. While the above challenges are in a sense inherent to the problem of auditing for bias, there are concrete steps we can take to begin to address them.

1. Set specific standards for what measures should be included in an audit.
2. Require auditors to report on metrics of differential validity.
3. Use caution in interpreting the results of audits. An audit can only test for specific discriminatory behaviors; it cannot certify that a tool is free of bias.

Thank you for your attention.
Manish Raghavan