International Review of Law, Computers & Technology

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/cirl20

Bias audit laws: how effective are they at preventing bias in automated employment decision tools?

Airlie Hilliard, Ayesha Gulley, Adriano Koshiyama & Emre Kazim

To cite this article: Airlie Hilliard, Ayesha Gulley, Adriano Koshiyama & Emre Kazim (12 Sep 2024): Bias audit laws: how effective are they at preventing bias in automated employment decision tools?, International Review of Law, Computers & Technology, DOI: 10.1080/13600869.2024.2403053

To link to this article: https://doi.org/10.1080/13600869.2024.2403053

INTERNATIONAL REVIEW OF LAW, COMPUTERS & TECHNOLOGY
https://doi.org/10.1080/13600869.2024.2403053

Bias audit laws: how effective are they at preventing bias in automated employment decision tools?

Airlie Hilliard[a,b], Ayesha Gulley[a], Adriano Koshiyama[a] and Emre Kazim[a]

[a]Holistic AI, London, UK; [b]Institute of Management Studies, Goldsmiths, University of London, London, UK

ABSTRACT

Automated employment decision tools use machine learning, artificial intelligence, predictive analytics, and other data-driven approaches to enhance candidate experiences and streamline employment related decision-making, allowing human resources to be concentrated where they are needed most. However, the use of these tools without appropriate safeguards has resulted in a number of high-profile scandals in recent years, particularly in regard to bias. Accordingly, lawmakers have started to propose laws that require bias audits of automated employment decision tools to examine their outputs for subgroup differences. The first of its kind was New York City Local Law 144, but other US states have since followed suit. In this paper, we examine the concerns about the effectiveness of this and other similar laws, including the suitability of metrics, the scope of the law, and low levels of compliance. We conclude that despite the law being a good initial first step towards greater transparency around automated employment decision tools and reducing bias, examining outcomes alone is not sufficient to prevent bias elsewhere in the tool. Moreover, effective bias prevention will require a multidisciplinary approach that combines expertise in IO psychology, law, and computer science to develop appropriate metrics and maximize the enforceability of such laws.

ARTICLE HISTORY
Received 19 March 2024
Accepted 8 September 2024

KEYWORDS
Bias audit; automated employment decision tool; adverse impact

1. Introduction

Automated tools are increasingly transforming human resource management processes, with a quarter of businesses reporting using intelligent automation to support their human resource functions (Maurer 2022). These tools can be used for a variety of human resource processes including attracting and sourcing candidates (Borisova et al. 2019; Said 2023), onboarding (Said 2023; Sambare et al. 2022), and even appraisals (Palshikar et al. 2019). Some of the most common applications of these tools, however, are for making hiring and promotion decisions. Termed automated employment decision tools (AEDTs), these solutions use technologies such as artificial intelligence, machine learning,

CONTACT Airlie Hilliard airlie.hilliard@holisticai.com Holistic AI, 18 Soho Square, W1D 3QH, London, UK; Institute of Management Studies, Goldsmiths, University of London, New Cross, SE14 6NW, London, UK

statistical modeling, and data analytics to make predictions and recommendations about a candidate's fit or performance based on input data.

AEDTs include automated video interviewing tools that make inferences about candidate fit or personality using verbal and non-verbal cues extracted from video interviews (Hickman et al. 2021), algorithmically scored image-based assessments (Hilliard et al. 2022a, 2022b; Leutner et al. 2017), and game-based assessments of constructs such as cognitive ability (Leutner et al. 2023). These alternative assessment formats are often used due to their potential to improve the candidate experience by increasing engagement and offering greater immersion (Kato and Kato 2017; Leutner et al. 2021; Lieberoth 2015), reducing test-taking anxiety (Kato and Kato 2017; Leutner et al. 2021; Leutner et al. 2023; Mavridis and Tsiatsos 2017; Smits and Charlier 2011) They can also offer increased flexibility since candidates typically interact with them using their own device at a convenient time.

The design, development, and deployment of these tools inherently involve interdisciplinary teams that combine expertise from multiple domains (Tippins, Oswald, and McPhail 2021). For example, industrial-organizational (I-O) psychologists are required to conduct a job analysis to identify the relevant traits and competencies to measure to predict performance in a particular job and ensure that assessments are grounded in psychological theory (Society for Industrial and Organizational Psychology 2018). On the other hand computer scientists are needed to create the algorithms to score the assessments. As such, the two disciplines must intersect to develop the associated documentation for the AEDT that details its design and development process and steps taken to examine its validity.

In this paper, we survey some of the ethical and legal concerns about AEDTs and the legal action that has been brought against them under existing laws. We then examine the emerging requirement to commission bias audits of AEDTs to ensure their legality and fairness. Specifically, we examine how the first law of its kind to require such audits, New York City Local Law 144, has set the precedent for bias auditing of AEDTs and how it has snowballed, influencing the proposal of similar laws across the East Coast of the United States. Within this, we examine the effectiveness of auditing the outcomes of AETs for bias and the proposal for a more comprehensive approach in New York State, discussing the implications this may have for compliance. We conclude that while the existing bias audit laws are certainly a step in the right direction towards the fairness and transparency of AEDTs, in their current form, they are not sufficient to completely prevent bias. As such, we recommend that interdisciplinary collaboration is needed to bring together expertise in law and compliance, I-O psychology, and policy to strengthen the impact of and compliance with similar laws in the future.

2. Ethical and legal concerns about AEDTs

Concerns about the novel risks that algorithms and automation can pose have been raised across sectors and applications, for example, about the explainability and transparency of algorithmic tools compared to human decision-making (Tippins, Oswald, and McPhail 2021). Moreover, when transparency efforts are made, in terms of providing explanations of the tools and the data they use, this can exacerbate concerns about the tools due to privacy concerns (Langer et al. 2021). This could question the extent to which *informed* consent could be granted if there is a lack of understanding about what the tool does and the data it collects. This is particularly a concern since AEDTs can use non-traditional data sources and make unintuitive connections between datapoints.

Moreover, there are concerns about AEDTs resulting in biased outcomes since there are several possible sources of bias within the lifecycle of the tool(Society for Industrial and Organizational Psychology 2018; Tay et al. 2022; Tippins, Oswald, and McPhail 2021).. For example, if the models used to score the tools to provide predictions and recommendations are trained on prior human judgments and those judgments were biased, then the models are likely to perpetuate and even amplify these biases (Tay et al. 2022). Unrepresentative training data can also result in bias if algorithms are applied to populations they are not optimized for. Additionally, the interface of the tool could be problematic for those with disabilities if accessibility was not considered in the design process (Tippins, Oswald, and McPhail 2021).

These are not merely concerns, though; these issues have been realized several times over the past decade from the use of AEDTs without the appropriate safeguards. The most infamous example of this is arguably Amazon's scrapped resume screening tool, which was trained on the resumes of previous applicants for technical positions at the company over the past ten years. Given the fact that the training data was made up of mostly male applicants, the algorithm penalized the resumes of female applicants that had the word 'women's' in their resume since it was not optimized to screen female resumes and 'women's' was not a term that was used by the resumes of male applicants (Dastin 2018). However, it is important to note that this bias against female applicants was recognized before the tool was deployed, so it did not have an actual negative impact, only a potential one.

3. AEDTs and the law

There have also been a number of legal cases in recent years that highlight how existing laws can be applied to AI and automated systems. For example, a legal complaint against video interview and assessment provider HireVue alleged that the company violated Section 5 of the *FTC Act of 1914,* which prohibits unfair or deceptive acts or practices, through the use of facial recognition (Electronic Privacy Information Center 2019). The use of this technology also led to concerns about the implications it could have for disabled applicants. As a result of the complaint, HireVue removed facial analysis from their algorithms, since advances in natural language technology meant that algorithm performance was maintained even with the removal of facial recognition (Kahn 2021; Zuloaga 2021).

Moreover, there is currently an ongoing legal complaint against applicant tracking system provider Workday due to alleged racial, disability, and age discrimination (Mobley vs Workday, Inc. 2023). Plaintiff Derek Mobley, a Black man over 40 with anxiety and depression, claims he was repeatedly rejected for jobs using Workday's AI-driven screening tool despite having a Bachelor's and Associate's degree, although Workday asserts that the case is without merit and submitted a motion to dismiss the case at the end of 2023 ("Mobley v Workday, Inc. (4:23-Cv-00770)" 2023). This resulted in the dismissal of the case in January 2024 (Mobley v Workday, Inc. (3:23-cv-00770-RFL) 2024a), prompting Mobley to file an amended complaint in February 2024 (Mobley v Workday, Inc. (3:23-cv-00770-RFL) 2024b). The EEOC has also weighed in on the complaint with an Amicus brief (EEOC 2024) that confirms that Workday meets the criteria for an employment agency, as stated in Mobley's complaint. However, in a ruling on 12 July 2024, the court dismissed claims that Workday is an employment agency, while it did uphold Workday as an agent with liability and gave Mobley 21 days to file an amended complaint (Mobley v Workday Inc. (3:23-cv-00770-RFL) 2024c). At the time of writing, the lawsuit is still ongoing.

While these examples illustrate seemingly unintentional bias and discrimination, automated systems can also be configured to automatically reject candidates based on specific criteria, such as their age. Indeed, the EEOC in the US recently announced a $365,000 settlement with iTutorGroup as part of a discrimination lawsuit (EEOC 2023). The group, which represents three companies providing English tutoring services to students in China, was found to be using an automated tool to automatically reject applicants based on their age. Specifically, the application software configured to reject female applicants aged 55 or older and male applicants aged 60 or older. This violates the *Age Discrimination in Employment Act of 1967*, which protects those aged 40 and over against discrimination in employment decisions.

Moreover, the fact that existing laws apply to AI is something that has repeatedly been emphasized by Federal agencies and regulators in the US. This includes the EEOC, which released a 'technical assistance document' titled *Assessing Adverse Impact in Software, Algorithms, and Artificial Intelligence Used in Employment Selection Procedures Under Title VII of the Civil Rights Act of 1964* in May 2023. The document outlined the need for employers to regularly assess their algorithms for bias and discrimination, and to ensure they carry out due diligence when acquiring third-party tools from vendors to ensure there are robust mechanisms in place to monitor for, identify, and mitigate bias resulting from AEDTs, stressing the fact that Title VII of the Civil Rights Act of 1964 still applies to AI-driven tools. This Title is enforced by the *Uniform Guidelines on Employee Selection Procedures* (EEOC 1978), which requires employers to carry out adverse impact analyses to compare selection rates based on race, sex, or ethnic group. Specifically, the Uniform Guidelines endorses the four-fifths rule as a metric for determining adverse impact, where the selection rate of one subgroup should not be less than four-fifths of the selection rate of the group with the highest rate. Here, selection rates refer to the proportion of each group that is hired, promoted, or otherwise selected (EEOC 1978). However, the Uniform Guidelines note that violations of the four-fifths rule might not indicate an adverse impact for small sample sizes, so the metric is widely referred to as a rule of thumb (EEOC 1979).

4. Legislation is codifying the need to audit AEDTs

The pervasiveness of AEDT incidents highlights that due to the novel risks that AI and algorithms introduce, existing laws may not be sufficient to fully prevent AEDTS from resulting in biased employment decisions. Accordingly, a number of AI-specific laws have been introduced in recent years. For example, the European Union's *Artificial Intelligence Act*– which was first proposed by the European Commission in April 2021 (European Commission 2021) and published in the Official Journal of the EU on 12 July 2024 (European Parliament and Council of the European Union 2024) ahead of its entry into force on 1 August 2024 – takes a horizontal approach and imposes obligations across key high-risk applications of AI, including in employment decisions. There is also an increasing focus on AEDTs specifically, with bias audit laws being introduced and enforced in the US. Such laws require independent, impartial evaluations of the tools to examine whether they result in unequal outcomes. Therefore, bias audits can be thought of as third-party adverse impact analyses.

4.1. New York city's precedent for bias audit laws

The first law of its kind in the world was New York City Local Law 144 (The New York City Council 2021), which mandated bias audits of AEDTs used by employers and employment agencies to evaluate employees for promotion or candidates for employment in New York City. Here, an automated employment decision tool (AEDT) is defined as a computation process that is derived from machine learning, statistical modeling, data analytics, or artificial intelligence that is used to issue a simplified output, such as a score, classification, or recommendation. This simplified output is used to substantially assist or replace discretionary decision-making. Specifically, machine learning, statistical modeling, data analytics, or artificial intelligence refer to a group of mathematical, computer-based techniques used to generate a prediction or classification, where a computer is used at least in part to identify the inputs, relative importance of such inputs, and other parameters to improve the accuracy of the model. Further, for an AEDT to substantially assist or replace discretionary decision-making, it must be the only factor considered to make decisions, must be the most important factor, or must be able to override decisions made based on other factors, including human decision-making.

The law was first proposed in February 2020 and passed in December 2021, due to go into effect from 1 January 2023. However, in September 2022, the regulator responsible for enforcing Local Law 144, the Department of Consumer Protection, published rules for enforcement (hereafter referred to as the Rules) and held a public hearing in November 2022. The public response to the Rules raised a number of concerns, including the definition of an AEDT and the metric proposed for conducting bias audits.

4.1.1. Bias audit metrics

In its Rules, the DCWP proposed two metrics; one for categorical tools and one for continuous tools that result in an output such as a score or rating. While the categorical metric was the same as the four-fifths rule – although the Rules do not specify that the four-fifths threshold should be used to determine whether bias is occurring — the continuous metric took the average score for subgroups and divided it by the average score of the group with the greatest average, as can be seen below.

$$\frac{\text{Average score of a subgroup}}{\text{Average score of the highest scoring subgroup}}$$

This metric proved to be controversial for many reasons (for a summary see (Groves et al. 2024), including the fact that it was not a suitable metric for non-normal distributions, where bimodal distributions with two peaks could result in instances of bias being missed, depending on the cutoff score used (Zannone and Filippi 2023).

As such, the DCWP published a second version of its Rules in December, where it revised the metric for continuous systems such that scores would be binarised based on the median score for the data set being audited. Based on this new metric, those scoring above the median would be considered as passing and those below as failing.

Using this so-called scoring rate, the same calculation is carried out as the categorical system – the scoring rate of each group is then divided by the scoring rate of the group with the highest rate, as can be seen below.

$$\frac{\text{Scoring rate of a subgroup}}{\text{Scoring rate of the group with the highest scoring rate}}$$

Again, seemingly extrapolated from the four-fifths rule of thumb metric – except without the requirement to use the .80 threshold to determine if bias is occurring. However, research by Filippi et al. (2023) using real recruitment data indicated that using the median score as a threshold still may mask some subgroup differences. Indeed, by comparing the binarization of scores based on the dataset mean, dataset median, and the threshold used by the recruitment company, the study found that using the mean did not result in the identification of any subgroup differences using the four-fifths threshold, while the threshold used by the company resulted in the identification of lower pass rates for black and two or more ethnicity groups. The median threshold, proposed by the DCWP Rules, identified a lower passing rate for black candidates, but did not identify the lower rate for the two or more ethnicities group (Filippi et al. 2023). In short, identified subgroup differences can vary depending on the threshold used. While standardizing the use of the median score as a threshold does provide standardization for audits, which will aid interpretability, employers or employment agencies may not use the median score as a threshold in practice. Thiscould result in the bias audit identifying group differences that would not be present when used to make real-life decisions, or missing group differences that may be present in practice, because of the use of different evaluation thresholds.

Irrespective of these concerns this metric for regression was adopted in the DCWP's final Rules (DCWP 2023a), which were published in April 2023, when the final enforcement date of 5 July 2023 was announced.

4.1.2. Transparency and notification rules

The final Rules also confirmed requirements for the publication of a summary of the results of the audit and notifications required for candidates and employees. According to these Rules (DCWP 2023a), the summary of results must contain:

- The source and explanation of the data used to conduct the bias audit
- The number of applicants in each category and applicants not included in the analysis due to missing demographic data
- The distribution date of the tool and the date of the audit
- Whether any categories were excluded from the analysis due to a small sample size, representing less than 2% of the data
- The impact ratios for standalone and intersectional groups

This summary must be updated with the latest annual audit and is required to be kept online for at least six months after the tool is retired. The requirement can be met by hyperlinking to a report on the employment section of the employer or employment agency's website in a clear and conspicuous manner.

Moreover, notifications must be provided to employees or candidates at least ten working days before the tool is used to evaluate them. This notice can be provided to candidates through the employment section of the website in a clear and conspicuous manner, in a job posting, or through mail or e-mail. Similarly, notice can be given to employees in a written policy or procedure, in a job posting, or via mail or e-mail. Such notice must contain information on:

- How the AEDT will be used and the job qualifications and characteristics it will consider in generating a simplified output
- The type and source of data collected by the AEDT
- Instructions on how to request accommodations or an alternative selection procedure
- The AEDT data retention policy.

4.1.3. Final bias audit process

Per the final Rules (DCWP 2023a), the bias audit process can be summarized as follows:

1. **Identify an independent auditor** – auditors must be an impartial entity to the employer, employment agency, or vendor; they cannot have been involved in using, developing, or distributing an AEDT. Auditors also cannot have an employment relationship with the employer or employment agency or have a direct or material indirect financial interest in the employer or employment agency using the AEDT or the vendor that developed or distributed the AEDT.
2. **Provide the appropriate data** – employers or employment agencies must provide output data from the tool, including the corresponding sex and race/ethnicity of candidates. The required sex/gender categories are male and female (and optionally other), and the required race/ethnicity categories are Hispanic or Latino, White, Black or African American, Native Hawaiian or Pacific Islander, Asian, Native American or Alaska Native, and two or more races.
    - This data should be real-life (historical) data as far as possible. The historical data used to conduct a bias audit may be from one or more employers or employment agencies that use the AEDT. As such, vendors may commission a bias audit on behalf of their clients, who can all rely on the same report.
    - If an employer or employment agency has not yet used the tool to contribute data to be audited, they may rely on a bias audit of an AEDT that uses the historical data of other employers or employment agencies.
    - Alternatively, employers that do not have historical data can collect 'test data'. This is not defined in any detail in the final Rules, but may take the form of panel data for example, where the tool is used to evaluate participants, who also provide demographic information. This practice is common in I-O psychology in the validation of tools before they are deployed.
3. **Conduct the bias audit** – independent auditors use the DCWP-specified metrics to conduct a bias audit of the provided output data for both standalone categories (e.g. male vs female) and intersectional categories (e.g. white female vs black female). Cases that are not associated with the required demographic data can be

removed from the analysis, as can groups representing less than 2% of the data, providing this is indicated in the summary of results.

4. **Create the summary of results and implement notification procedures** – the required summary of results may be provided by auditors or may be created by vendors/employers/employment agencies themselves based on the outcome of the audit. Since notifications are not required to provide information on the results of the audit, notifications can be prepared while the audit is ongoing.

However, practitioners and experts report that the final Rules and required auditing methodology fall short of the spirit of the law and fail to protect job applicants from biased AEDTs in practice (Groves et al. 2024). Some of the points of contention include whether the notification will indeed help applicants to make more informed decisions about their application to roles (Groves et al. 2024), particularly because Local Law 144 bias audits do not require that AEDTs 'pass' the audit – just that they have one. Moreover, although the notification must provide instructions on how to request an alternative procedure or accommodation, Local Law 144 does not require this request to be met (DCWP 2023a), although this may have implications under other equal opportunity laws (Equal Employment Opportunity Commission 2022), such as the Americans with Disabilities Act (1990). Furthermore, since auditors purely are required to be independent and do not have to meet any other specific criteria or become qualified, this could see users of AEDTs receiving differing levels of expertise and services from auditor to auditor (Groves et al. 2024).

4.2. Other AEDT bias audit laws

Since being passed, Local Law 144 has seemingly inspired a number of laws to be proposed in the US at the state level, which we compare in Table 1. For example, New York has proposed two laws, AB567 and AB7859, which could be argued to be equivalent to Local Law 144 when combined. AB567 requires annual impartial 'disparate impact analyses' of AEDTs, along with the publication of the summary of results of the bias audit, while AB7859 requires notification for the use of AEDTs in the state, essentially having the same requirements as Local Law 144 when combined. Moreover, Pennsylvania has

Table 1. A comparison of automated employment decision tool bias audit laws in the US.

Law	Liability	Bias audit required	Notification required	Summary of results required
NYC Local Law 144	Employer	Independent, impartial bias audit of outcomes	At least 10 working days prior to use	Publicly available online
New Jersey S1588	Vendor	Impartial evaluation of outcome	Notification within 30 days of use	Available to purchaser
Pennsylvania HB1729	Employer	Impartial, independent bias audit	At least 10 days prior to use and obtain consent	Publicly available online
New York S7623	Employer	Impartial, independent bias audit that examines output, tool validity, proxy features, and training data	At least 10 days prior to use	Publicly available
New York AB7859	Employer	No	At least 10 working days prior to use	No
New York AB56	Employer	Impartial disparate impact analysis of outcomes	No	Publicly available online

introduced HB1729, which again requires annual, independent bias audits of AEDTS and the publication of a summary of the results of the bias audit and notification.

New Jersey, on the other hand, has introduced S1588. This law is a carry-over of AB 4909, which was first proposed in December 2022. In contrast to the previously discussed laws, the New Jersey law places liability on vendors of AEDTs instead of the employers and employment agencies that use them. Nevertheless, the requirements of S1588 are similar to those of Local Law 144 in that it requires annual disparate impact analysis of AEDTs, requiring vendors to offer this to clients at no additional cost.

With a perhaps more comprehensive approach than these laws, in August of 2023, New York introduced S7623 to restrict the use of electronic monitoring and AEDTs by employers and employment agencies, where an AEDT is defined similarly to Local Law 144 as: 'any computational process, automated system, or algorithm utilizing machine learning, statistical modeling, data analytics, artificial intelligence, or similar methods that issue a simplified output, including a score, classification, ranking, or recommendation, that is used to assist or replace decision making for employment decisions that impact natural persons'. Here, employment decisions cover wages, benefits, other compensation, hours, work schedule, performance evaluation, hiring, selecting for recruitment, discipline, promotion, termination, job content, assignment of work, access to work opportunities, and productivity requirements.

As well as setting out requirements reminiscent of the now-dead *California Workplace Technology Accountability Act* (Hilliard et al. 2023), S7623 makes it illegal to use an AEDT unless it has been subject to an independent, impartial bias audit within the past year. However, the bias audits required by the New York Law go beyond previously proposed requirements;the audit must identify and describe attributes and modeling techniques that the tool uses to produce outputs. Users must also note whether they are scientifically valid ways of evaluating performance or the ability to perform essential job functions.

Bias audits must also examine whether these attributes could function as a proxy for protected characteristics, as well as evaluate the training data for any disparities and how they might result in disparate impact. Moreover, the bias audits should recommend any necessary mitigations and evaluate how the tool may impact accessibility for those with disabilities. Employers and employment agencies must support these efforts by retaining pertinent documentation. As well as a comprehensive audit, S7623 requires notifications similar to those required by Local Law 144. Notifications must be given to employees and candidates at least ten business days before the use of the tool, in addition to meaningful human oversight of the use of AEDTs.

5. How effective are bias audits at preventing bias?

These bias audit laws make a notable contribution to greater transparency around the use of AEDTs and their capabilities through the required notifications and publications of summaries of results. They are,therefore, a step towards more informed consent for the use of AEDTs. However, whether or not they are effective at preventing bias from AEDTs is less clear.

5.1. Implications of assessing the outcomes of AEDTs

Indeed, there are longstanding concerns about the usefulness of the four-fifths rule, with legal concerns being raised in the same year as the publication of the uniform guidelines (Rubin 1978). Issues with the metric include a lack of reliability for small sample sizes, as noted in the guidelines, and the fact that the metric can contradict other adverse impact metrics such as the two standard deviations rule (Morris and Lobsenz 2000), which can be favored over the four-fifths rule by US courts (PSI 2018). Given the fact that the metric for calculating impact ratios under Local Law 144 seemingly extrapolates from the four-fifths rule, these concerns also transfer to bias audits of AEDTs. Indeed, despite being updated in the second and final versions of the Rules, the regression metric has raised concerns about how well it can handle data with certain distributions, such as in the case of bimodal distributions where there are two peaks in scores. Since the regression metric divides the dataset in half, where the lower half 'fails' and the upper half 'passes', the metric can fail to detect bias at certain thresholds of the distribution of scores is not normal (Filippi et al. 2023).

Notwithstanding issues with the metric itself, bias audits solely focused on the outputs of AEDTs could fail to detect biased treatment. For example, several approaches to mitigating bias have been proposed in the machine learning field, typically categorized as pre-processing, in-processing, or post-processing (Raghavan and Barocas 2019):

- Pre-processing – Transforming the input data or training data to minimize its relationship to protected attributes while maximizing similarity to the original data
- In processing – Introducing constraints to penalize the use of protected attributes to make predictions
- Post-processing – Outputs are adjusted to weaken the relationship to subgroup membership

Some approaches to bias mitigation could be particularly problematic in social applications of AI and algorithms where predictions are made about humans instead of objects (Mullainathan and Obermeyer 2017). This is because post-processing approaches can require prediction labels to be changed for some individuals so that the distribution of outcomes is more equal across groups. For example, Calders, Kamiran, and Pechenizkiy (2009) proposed a post-processing approach to achieving independence of predictions and protected attributes where some data labels in the training are changed, or *massaged*, such that the model makes predictions that are no longer dependent on the protected attribute. When applied to recruitment, if in the training data, more males are recommended than females to progress to an in-person interview, the designation for some males is changed to the negative condition and some females changed to the positive condition so that the distribution is more equal. However, changing scores based on subgroup membership is illegal in the US under the Civil Rights Act of 1991, so, although only the training data is being altered and not the predictions, the legality of this approach when used for social applications could be questioned and could spark concerns about disparate treatment. As such, a bias audit solely focused on the outcome with limited details on how predictions are made or the model is trained would be unlikely to detect this.

Moreover, NYC Local Law 144 requires bias audits against sex and race/ethnicity at a minimum, as required for adverse impact testing under the Universal Guidelines. However, the use of technology could also pose bias based on age, with older adults reporting less technology self-efficacy compared to younger adults (Ellison et al. 2020) and taking a more formal approach to video interviews than younger adults (McColl and Michelotti 2019). Moreover, given that technology might result in disability discrimination, the EEOC has issued guidance on how to ensure compliance with the *Americans with Disabilities Act* while using AEDTs (EEOC 2022). However, bias audits focusing on only sex and race/ethnicity might overlook this. As such, there is a need for more comprehensive approaches that examine more elements of the design and development of AEDTs, not just the deployment and outcomes. Given that the approach of New York S7623 requires examination of a wider range of data and covers multiple points in the lifecycle of a tool, touching data collection, model features and training, and deployment, it is more likely to capture disparate treatment and issues with model optimization compared to the outcome audit required by Local Law 144.

5.2. Effectiveness requires compliance

Although the New York law seeks to require a more comprehensive bias audit of AEDTs in comparison to other laws, it is essential to remember that a law can only be effective if it is complied with. Indeed, a recent study investigating Local Law 144 compliance found that of a sample of 391 employers, only 8 posted audit reports and 13 posted transparency notices within the first week of the law being enforced (Wright et al. 2024). While a first instinct might be to justify this by pointing out the fact that Local Law 144 is a Local Law, it does have global implications given that it applies to employers and employment agencies using the tools in the City, and therefore can have global implications for any employer hiring or promoting using AEDTs in New York City.

One reason for the seeming lack of compliance with Local Law 144 could be the narrow definition of 'to substantially assist or replace discretionary decision making'. According to the Rules, (DCWP 2023a), substantial assistance means the tool is (i) the only factor considered, (ii) is weighted more heavily than any other factor, (iii) or is used to override decisions based on other factors. Indeed, in the law text, the definition of an AEDT is wide and likely to capture many tools. However, the way in which the Rules define key terms within this definition significantly narrows the scope and allows many employers and employment agencies to argue that they are not covered by the law since many AEDTs are designed to support human decision-making instead of making the decisions themselves. Further, while the law text stated that it covered AEDTs used to evaluate candidates or employees that *reside* in NYC, the FAQs on the law (DCWP 2023b) clarify that to use an AEDT in the city means that the employment agency is based in NYC or the job location is in NYC at least part-time or the job is remote but associated with a NYC office. As such, fully remote businesses hiring out of New York City, for example, may not be required to comply with the law if they do not have an office in the City. As such, although it cannot be determined whether the lack of published bias audits is due to non-compliance or whether the seemingly deliberate narrowing of the scope of the law has left only a very limited number of AEDTs in scope, the lack of bias audits indicates that Local Law 144's impact may be limited.

Furthermore, the penalties for non-compliance with Local Law 144 are not very large compared to other laws (e.g. up to €7 million for the *EU AI Act*; European Council 2023) at $500 for the first violation and $1500 for subsequent violations, where not having a bias audit and not providing notice are separate violations. It is also not clear exactly what constitutes a single violation – whether each day of using the AEDT without a bias audit or providing notice is a single violation, or whether this is multiples by the number of candidates or employees evaluated by the tool each day. This is not something we condone by any means, but take the (albeit unlikely) scenario of an employer that uses an AEDT solely for graduate positions and only hires once per year, scheduling assessments over a one-week (five business day) period. If the former definition of a violation is used, then the penalty will only be $13000 if they do not commission a bias audit or provide notice. On the other hand, a bias audit may be much more costly, so the legal team might take a calculated risk and make the unfavorable decision to pay any fines (if they are even investigated for a suspected violation) rather than commission a third-party bias audit– although this could lead to further financial damage through the reputational impact of non-compliance. Additionally, since the law has only recently come into effect, any precedent for enforcement against violations is yet to be set. Nevertheless, it is often the vendor that coordinates bias audits (Wright et al. 2024) since evaluations can be made on data aggregated across different employers that use the AEDT (DCWP 2023a). As such, it could in fact be the vendor that incurs the cost of the audit, while on the other hand, employers themselves would be liable for paying any penalties.

Moreover, given the seeming lack of compliance with Local Law 144, it begs the question, how effective would a law like NY S7623 be? Indeed, the New York State law would require employers and employment agencies to put themselves in a more vulnerable position than Local Law 144, where they not only have to provide the outputs of the tool, but also details on the model, features, and training data. Moreover, the text of S7623 does not outline any provisions for protecting trade secrets or intellectual property, and it has been suggested that certain elements of AI models do not meet the criteria for trade secret laws because they may not directly create commercial value by being secret (Mylly 2023; Sandeen and Aplin 2022). As such, willingness to comply may be impacted. However, the EU Digital Services Act, which requires designated Very Large Online Platforms and Very Large Online Search Engines to undergo independent audits of their transparency and accountability processes, binds both auditors and the European Commission – who audit reports must be sent to – to an obligation of professional secrecy in accordance with Article 84 (European Parliament and Council of the European Union 2022). As such, if passed, enforcement rules for S7623 may follow this example to provide greater reassurance that any exposed trade secrets will be considered privileged, which may support greater willingness to comply.

5.3. The need for interdisciplinary collaboration

Effective bias audit laws require collaboration from psychologists, computer scientists, and policymakers to ensure that metrics are appropriate and effective for the tools being evaluated. Some progress has already been made towards this, particularly among computer scientists and psychologists. Indeed, IO psychology already has the concept of predictive bias, referring to when the same regression line cannot be applied to all subgroups, resulting in different subgroups with the same underlying ability being predicted different scores (Berry and Zhao 2015; Society for Industrial and Organizational Psychology 2018). Similarly, computer science has several definitions of fairness centered around ensuring regressions fit different subgroups equally, such as equalized odds – which says that the true and false positive rates of subgroups should be equal – and accuracy equality, where the probability of true positives and true negatives is equal for different subgroups (Verma and Rubin 2018).

Combining insights from both fields, Tay et al. (2022) proposed the machine learning measurement bias framework for investigating the differential functioning of machine learning models in psychometric assessments for different subgroups. While a specific metric for measuring machine learning measurement bias was not proposed, the framework outlined several ways to measure bias in both the training data and model features, providing for a more holistic approach to identifying bias rather than solely focusing on the outputs. Moreover, Kazim et al. (2021) provide a framework for systematizing holistic algorithm audits of recruitment tools that combines insights from computer science and IO psychology, suggesting a range of metrics that can be used from each field to test for bias and other key algorithm risks. However, given the nascency of algorithm auditing and the disparity between how AEDTs are examined in an audit and how they are used in practice, additional research is needed in this field to continue to build on this groundwork and provide additional metrics for the evaluation of AEDTs.

6. Conclusion

While it can be acknowledged that New York City Local Law 144 has made important initial progress towards the fairer and more transparent use of AEDTs and has paved the way for future legislation, already resulting in similar laws being proposed in other states, it is not sufficient to completely prevent bias from AEDTs alone. A law is only effective if it is complied with. In the case of Local Law 144, a low number of bias audits have been completed. While this is likely due to employers and employment agencies working within legal loopholes created by definitions provided in the final version of the enforcement rules instead of non-compliance, it highlights the limited impact of the law. Indeed, to solve bias in automated employment decision tools requires as many employers and employment agencies as possible to take action, with Local Law 144 only impacting a small proportion of employers and employment agencies in a constricted region. As such, while Local Law 144 has perhaps set the precedent, truly solving the problem of bias in outputs requires extra-territorial laws with a wide scope.

Moreover, solely examining the outputs of systems is not guaranteed to be effective for preventing upstream disparate treatment, and the metric for examining outputs itself is widely acknowledged to be flawed. As such, we call for increased interdisciplinary efforts that combine expertise in law and policy, I-O psychology, computer science, and data science in order to maximize the utility of bias audits of AEDTs, ensure that metrics are appropriate for different distributions of data, and ensure that different types of bias can be identified, as well as to maximize compliance. Nevertheless, we must start somewhere, and Local Law 144 is certainly a good starting point that has set the precedent for future action.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Americans with Disabilities Act. 1990. https://www.govinfo.gov/content/pkg/STATUTE-104/pdf/STATUTE-104-Pg327.pdf.

Berry, Christopher M., and Peng Zhao. 2015. "Addressing Criticisms of Existing Predictive Bias Research: Cognitive Ability Test Scores Still Overpredict African Americans' Job Performance." *Journal of Applied Psychology* 100 (1): American Psychological Association Inc.: 162–179. https://doi.org/10.1037/a0037615.

Borisova, Alena, Madina Rakhimberdinova, Elvira Madiyarova, Inna Riazantseva, Natalia Mikidenko, and D Serikbayev East Kazakhstan. 2019. "Staffing Search and Recruitment of Personnel on the Basis of Artificial Intelligence Technologies." *Entrepreneurship and Sustainability Issues* 6 (4): 2456–2469. https://doi.org/10.9770/jesi.2019.6.4(66).

Calders, Toon, Faisal Kamiran, and Mykola Pechenizkiy. 2009. "Building Classifiers with Independency Constraints." In *ICDM Workshops 2009 - IEEE International Conference on Data Mining*, 13–18. https://doi.org/10.1109/ICDMW.2009.83.

Dastin, Jeffrey. 2018. "Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women." https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

DCWP. 2023a. "Notice of Adoption of Final Rules." https://rules.cityofnewyork.us/wp-content/uploads/2023/04/DCWP-NOA-for-Use-of-Automated-Employment-Decisionmaking-Tools-2.pdf.

DCWP. 2023b. "Automated Employment Decision Tools: Frequently Asked Questions." https://codelibrary.amlegal.com/codes/newyorkcity/latest/NYCrules/0-0-0-138530.

Derek L. Mobley vs Workday, Inc. 2023. *3:23-Cv-00770-TSH*.

EEOC. 1979. "Questions and Answers to Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures | U.S. Equal Employment Opportunity Commission." https://www.eeoc.gov/laws/guidance/questions-and-answers-clarify-and-provide-common-interpretation-uniform-guidelines.

EEOC. 2023. "ITutorGroup to Pay $365,000 to Settle EEOC Discriminatory Hiring Suit | U.S. Equal Employment Opportunity Commission." https://www.eeoc.gov/newsroom/itutorgroup-pay-365000-settle-eeoc-discriminatory-hiring-suit.

EEOC. 2024. Brief of the Equal Employment Opportunity Commission as Amicus Curiae in Support of Plaintiff and in Opposition to Defendant's Motion to Dismiss. https://www.eeoc.gov/sites/default/files/2024-04/Mobley%20v%20Workday%20NDCal%20am-brf%2004-24%20sjw.pdf.

Electronic Privacy Information Center. 2019. *Complaint and Request for Investigation, Injunction, and Other Relief*. https://epic.org/wp-content/uploads/privacy/ftc/hirevue/EPIC_FTC_HireVue_Complaint.pdf.

Ellison, Leah Joyce, Tara McClure Johnson, David Tomczak, Alina Siemsen, and Manuel Francisco Gonzalez. 2020. "Game on! Exploring Reactions to Game-Based Selection Assessments." *Journal of Managerial Psychology* 35 (4): Emerald Group Holdings Ltd.: 241–254. https://doi.org/10.1108/JMP-09-2018-0414.

Equal Employment Opportunity Commission. 1978. "Uniform Guidelines on Employee Selection Procedures." *Federal Register* 43 (166): 38290–38315.

Equal Employment Opportunity Commission. 1991. *Civil Rights Act of 1991*. https://www.eeoc.gov/civil-rights-act-1991-original-text.

Equal Employment Opportunity Commission. 2022. "The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees | U.S. Equal Employment Opportunity Commission." https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence.

European Commission. 2021. *Proposal for a Regulation Laying down Harmonised Rules on Artificial Intelligence*. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence.

European Council. 2023. "Artificial Intelligence Act: Council and Parliament Strike a Deal on the First Rules for AI in the World - Consilium." https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/.

European Parliament, and Council of the European Union. 2022. "REGULATION (EU) 2022/2065." https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R2065.

European Parliament, and Council of the European Union. 2024. *Regulation (EU) 2024/1689*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401689.

Filippi, Giulio, Sara Zannone, Airlie Hilliard, and Adriano Koshiyama. 2023. "Local Law 144: A Critical Analysis of Regression Metrics," February. http://arxiv.org/abs/2302.04119.

Groves, Lara, Jacob Metcalf, Alayna Kennedy, Briana Vecchione, and Andrew Strait. 2024. "Auditing Work: Exploring the New York City Algorithmic Bias Audit Regime," February. http://arxiv.org/abs/2402.08101.

Hickman, Louis, Nigel Bosch, Vincent Ng, Rachel Saef, Louis Tay, and Sang Eun Woo. 2021. "Automated Video Interview Personality Assessments: Reliability, Validity, and Generalizability Investigations." *Journal of Applied Psychology* 11 (1): 11–22. https://doi.org/10.1037/apl0000695.

Hilliard, Airlie, Emre Kazim, Theodoros Bitsakis, and Franziska Leutner. 2022a. "Measuring Personality through Images: Validating a Forced-Choice Image-Based Assessment of the Big Five Personality Traits." *Journal of Intelligence* 10 (1): Multidisciplinary Digital Publishing Institute: 12. https://doi.org/10.3390/jintelligence10010012.

Hilliard, Airlie, Emre Kazim, Theodoros Bitsakis, and Franziska Leutner. 2022b. "Scoring a Forced-Choice Image-Based Assessment of Personality: A Comparison of Machine Learning, Regression, and Summative Approaches." *Acta Psychologica* 228 (August): North-Holland: 103659. https://doi.org/10.1016/J.ACTPSY.2022.103659.

Hilliard, Airlie, Emre Kazim, Tom Kemp, and Kelvin Bageire. 2023. "Overview and Commentary of the California Workplace Technology Accountability Act." *International Review of Law, Computers & Technology* 37 (1): Routledge: 91–109. https://doi.org/10.1080/13600869.2022.2115749.

Kahn, Jeremy. 2021. "HireVue Stops Using Facial Expressions to Assess Job Candidates amid Audit of Its AI Algorithms." https://fortune.com/2021/01/19/hirevue-drops-facial-monitoring-amid-a-i-algorithm-audit/.

Kato, Pamela, and P. M. Kato. 2017. "Serious Games for Assessment: Welcome to the Jungle Sebastiaan de Klerk Cito Arnhem." *Journal of Applied Testing Technology* 18 (S1): 1–6. https://www.researchgate.net/publication/321492367.

Kazim, Emre, Adriano Soares Koshiyama, Airlie Hilliard, and Roseline Polle. 2021. "Systematizing Audit in Algorithmic Recruitment." *Journal of Intelligence* 9 (3): Multidisciplinary Digital Publishing Institute: 46. https://doi.org/10.3390/jintelligence9030046.

Langer, Markus, Kevin Baum, Cornelius J. König, Viviane Hähne, Daniel Oster, and Timo Speith. 2021. "Spare Me the Details: How the Type of Information about Automated Interviews Influences Applicant Reactions." *International Journal of Selection and Assessment* 29 (2): John Wiley & Sons, Ltd: 154–169. https://doi.org/10.1111/ijsa.12325.

Leutner, Franziska, Sonia-Cristina Codreanu, Suzanne Brink, and Theodoros Bitsakis. 2023. "Game Based Assessments of Cognitive Ability in Recruitment: Validity, Fairness and Test-Taking Experience." *Frontiers in Psychology* 13 (January), https://doi.org/10.3389/fpsyg.2022.942662.

Leutner, Franziska, Sonia-Cristina Codreanu, Josh Liff, and Nathan Mondragon. 2021. "The Potential of Game- and Video-Based Assessments for Social Attributes: Examples from Practice." *Journal of Managerial Psychology* 36 (7): 533–547. https://doi.org/10.1108/JMP-01-2020-0023.

Leutner, Franziska, Adam Yearsley, Sonia Cristina Codreanu, Yossi Borenstein, and Gorkan Ahmetoglu. 2017. "From Likert Scales to Images: Validating a Novel Creativity Measure with Image Based Response Scales." *Personality and Individual Differences* 106 (February): Elsevier Ltd: 36–40. https://doi.org/10.1016/j.paid.2016.10.007.

Lieberoth, Andreas. 2015. "Shallow Gamification: Testing Psychological Effects of Framing an Activity as a Game." *Games and Culture* 10 (3): 229–248. https://doi.org/10.1177/1555412014559978.

Maurer, Roy. 2022. "SHRM Research: AI Use on the Rise, Ethics Questions Remain." *Society for Human Resource Management*. 20/01/24. https://www.shrm.org/topics-tools/news/technology/shrm-research-ai-use-rise-ethics-questions-remain.

Mavridis, A., and T. Tsiatsos. 2017. "Game-Based Assessment: Investigating the Impact on Test Anxiety and Exam Performance." *Journal of Computer Assisted Learning* 33 (2): 137–150. https://doi.org/10.1111/jcal.12170.

McColl, Rod, and Marco Michelotti. 2019. "Sorry, Could You Repeat the Question? Exploring Video-Interview Recruitment Practice in HRM." *Human Resource Management Journal* 29 (4): John Wiley & Sons, Ltd: 637–656. https://doi.org/10.1111/1748-8583.12249.

Mobley v Workday Inc. (3:23-cv-00770-RFL). 2024a. "Order Granting Motion to Dismiss with Leave to Amend." https://storage.courtlistener.com/recap/gov.uscourts.cand.408645/gov.uscourts.cand.408645.45.0.pdf.

Mobley v Workday Inc. (3:23-cv-00770-RFL). 2024b. "First Amended Class Action Complaint." https://storage.courtlistener.com/recap/gov.uscourts.cand.408645/gov.uscourts.cand.408645.47.0.pdf.

Mobley v Workday Inc. (3:23-cv-00770-RFL). 2024c. "Order Granting in Part and Denying in Part Motion to Dismiss." https://blogs.duanemorris.com/classactiondefense/wp-content/uploads/sites/56/2024/07/Mobley-v.-Workday-Order.pdf.

Mobley v Workday Inc. (4:23-cv-00770). 2023. "Complaint." Northern District of California. https://www.courtlistener.com/docket/66831340/1/mobley-v-workday-inc/.

Morris, Scott B., and Russell B. Lobsenz. 2000. "Significance Tests and Confidence Intervals for the Adverse Impact Ratio." *Personnel Psychology* 53 (1): Personnel Psychology, Inc.: 89–111. https://doi.org/10.1111/j.1744-6570.2000.tb00195.x.

Mullainathan, Sendhil, and Ziad Obermeyer. 2017. "Does Machine Learning Automate Moral Hazard and Error?" *American Economic Review* 107:476–480. NIH Public Access. https://doi.org/10.1257/aer.p20171084.

Mylly, Ulla-Maija. 2023. "Transparent AI? Navigating Between Rules on Trade Secrets and Access to Information." *IIC - International Review of Intellectual Property and Competition Law* 54 (7): 1013–1043. https://doi.org/10.1007/s40319-023-01328-5.

The New York City Council. 2021. *Int 1894-2020*. https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9&Options=Advanced&Search.

Palshikar, Girish Keshav, Manoj Apte, Sachin Pawar, and Nitin Ramrakhiyani. 2019. "HiSPEED: A System for Mining Performance Appraisal Data and Text." *International Journal of Data Science and Analytics* 8 (1): Springer Science and Business Media Deutschland GmbH: 95–111. https://doi.org/10.1007/S41060-018-0142-X/FIGURES/5.

PSI. 2018. *Understanding Adverse Impact in the Hiring Process*. https://content.psionline.com/hubfs/Talent Management White Papers/WP_PSI Understanding Adverse Impact.pdf.

Raghavan, Manish, and Solon Barocas. 2019. "Challenges for Mitigating Bias in Algorithmic Hiring." https://www.brookings.edu/research/challenges-for-mitigating-bias-in-algorithmic-hiring/.

Rubin, Ronald B. 1978. "The Uniform Guidelines on Employee Selection Procedures: Compromises and Controversies." *Catholic University Law Review* 28:605–634. https://scholarship.law.edu/lawreview/vol28/iss3/7.

Said, Jenn. 2023. "Using HR Automation to Attract Stronger Construction Candidates and Simplify Remote Onboarding Challenges." *Strategic HR Review* 22 (2): Emerald Publishing Limited: 58–60. https://doi.org/10.1108/SHR-01-2023-0009.

Sambare, S. S., Akriti Singh, Chirag Kriplani, Shweta Kale, and Tanuj Balkhande. 2022. "Automated Platform for Onboarding Employee." *2022 6th International Conference on Computing, Communication, Control and Automation, ICCUBEA 2022*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICCUBEA54992.2022.10010715.

Sandeen, Sharon K., and Tanya Aplin. 2022. "Trade Secrecy, Factual Secrecy and the Hype Surrounding AI." In *The Human Cause*, edited by Ryan Abbott, 443–460. Cheltenham, UK: Edward Elgar Publishing. https://doi.org/10.4337/9781800881907.00032.

Smits, Jarka, and Nathalie Charlier. 2011. "Game-Based Assessment and the Effect on Test Anxiety: A Case Study." In *Proceedings of the European Conference on Games-Based Learning*, 562–566.

Society for Industrial and Organizational Psychology. 2018. *Principles for the Validation and Use of Personnel Selection Procedures*. 5th ed. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/iop.2018.195.

Tay, Louis, Sang Eun Woo, Louis Hickman, Brandon M. Booth, and Sidney D'Mello. 2022. "A Conceptual Framework for Investigating and Mitigating Machine-Learning Measurement Bias (MLMB) in Psychological Assessment." *Advances in Methods and Practices in Psychological Science* 5 (1): 1–30. https://doi.org/10.1177/25152459211061337.

Tippins, Nancy T., Frederick Oswald, and S. Morton McPhail. 2021. "Scientific, Legal, and Ethical Concerns About AI-Based Personnel Selection Tools: A Call to Action." *Personnel Assessment and Decisions* 7 (2): 1–22. https://doi.org/10.25035/pad.2021.02.001.

Verma, Sahil, and Julia Rubin. 2018. "Fairness Definitions Explained." *IEEE/ACM International Workshop on Software Fairness* 18. ACM. https://doi.org/10.1145/3194770.3194776.

Wright, Lucas, Roxana Mika Muenster, Briana Vecchione, Tianyao Qu, Senhuang (Pika) Cai, Alan Smith, Jake Metcalf, and J. Nathan Matias. 2024. "Null Compliance: NYC Local Law 144 and the Challenges of Algorithm Accountability." OSF. https://doi.org/10.17605/OSF.IO/UPFDK.

Zannone, Sara, and Giulio Filippi. 2023. "Disparate Impact in Bias Audits: Evaluating the DCWP's Impact Ratio Metrics for Regression Systems." https://www.holisticai.com/blog/nyc-bias-audit-impact-ratio-regression.

Zuloaga, Lindsey. 2021. "Industry Leadership: New Audit Results and Decision on Visual Analysis | HireVue." https://www.hirevue.com/blog/hiring/industry-leadership-new-audit-results-and-decision-on-visual-analysis.