

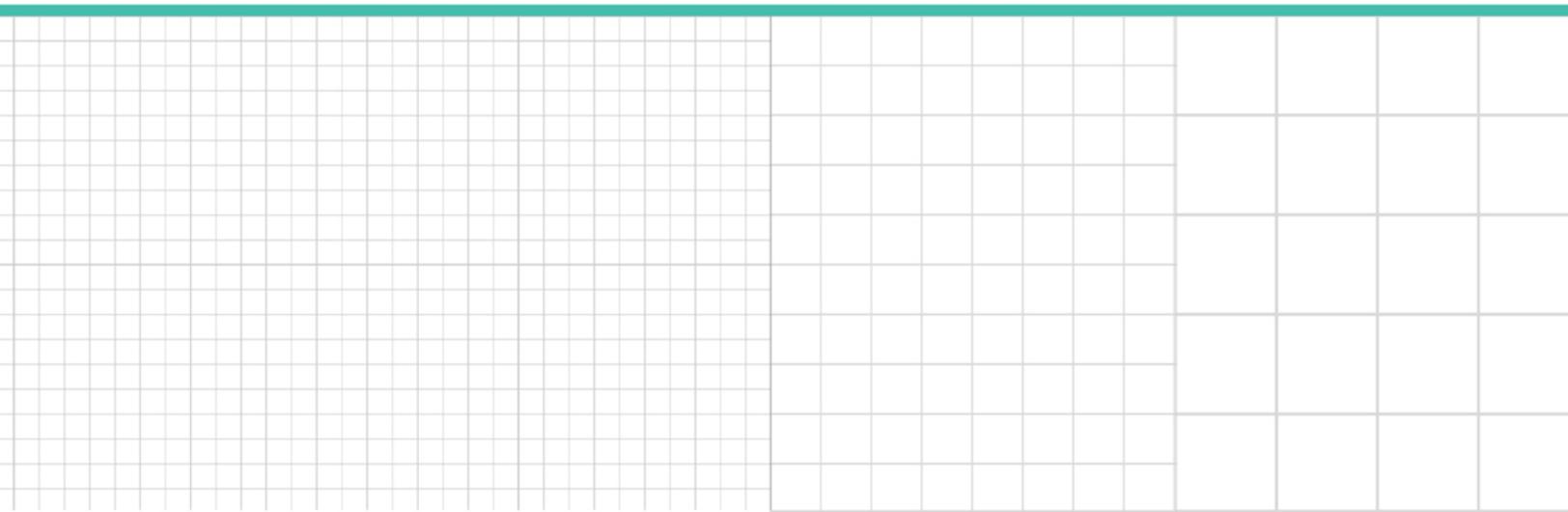


Professional Perspective

# AI Regulation and Risks to Employers

*Joseph O'Keefe, Danielle Moss, and Tony Martinez,  
Proskauer Rose*

Reproduced with permission. Published December 2019. Copyright © 2019 The Bureau of National Affairs, Inc. 800.372.1033. For further use, please visit: <http://bna.com/copyright-permission-request/>



# AI Regulation and Risks to Employers

Contributed by *Joseph O'Keefe, Danielle Moss, and Tony Martinez, Proskauer Rose*

The age of artificial intelligence is upon us as machine learning products and services are hitting the market with greater frequency. But these technologies are not new. Most of us already interact with AI on a daily basis when we type information into search engines, click suggested articles on news feeds, or peruse recommended products on shopping websites.

Notwithstanding the presence of AI in our everyday lives, some may be surprised to learn that algorithms are now being used in much more impactful ways, such as by companies in recruitment and performance management, as well as by governments to conduct predictive policing and impose sentencing. For example, police departments in New York and Los Angeles have started using predictive policing tools to identify where criminal activity may occur.

Given the gravity of the potential consequences that may arise when AI is applied in these circumstances, and in response to pressure from activists, governments at all levels have begun regulating algorithms and issuing guidance regarding their development and usage.

Before implementing AI in the workplace, employers should consider the following pros and cons, including but not limited to the potential risk of algorithmic bias and related legislative developments.

## Potential for Algorithmic Bias

Machine learning algorithms created by computer scientists are designed to “learn” based upon the algorithm's access to either a specific designated data set, or an algorithm-driven search for data sets residing on the internet or in a confined database. The algorithms then analyze the data and identify patterns within the training data set(s) to apply those patterns to assist in making future decisions.

Take, for example, an algorithm programmed to screen resumes. To perform its task, the algorithm might begin by analyzing the training data to identify patterns and ultimately, pinpoint successful applicants who fit the “pattern.” In this case, that data might take the form of resumes previously submitted to the employer that were reviewed by humans (including potential implicit or explicit biases) and indicate which resumes led to job offers. In an attempt to replicate past successes, the algorithm would rely upon and amplify the discovered patterns in evaluating the resumes presented to it going forward.

One might assume that an algorithm is an ideal resume reviewer—after all, it is an impartial machine built specifically to identify the best available candidates without the apparent risk of human error or bias. However, recent research findings suggest that algorithms may not be neutral in circumstances where the training data provided contained flawed and/or biased information.

These research findings are not merely academic: this cautionary tale played out in 2017, when Amazon was forced to abandon its own algorithm because of such bias. There, the training data presented to the algorithm consisted of resumes submitted to Amazon by applicants over a 10-year period, most of whom were white males. Rather than sort the most qualified candidates based on qualifications or merit, the algorithm quickly taught itself to favor male candidates by prioritizing language more commonly used by males—going so far as to penalize the word “women's” in resumes and candidates who graduated from all-women's colleges. Immediately upon discovering this flaw, Amazon axed the algorithm and noted that it “was never used by Amazon recruiters to evaluate candidates.”

Likewise, employers have begun (or are considering) using algorithms to assist with performance management. Amazon, for example, uses an algorithm that determines how much time an employee in a fulfillment center spends unproductively. The algorithm refers to this as a “time off task” (TOT). When workers take significant breaks from fulfilling orders, including restroom breaks, the algorithm tracks this time and is capable of generating warnings and even terminations based on an employee's productivity levels.

Supervisors can, however, override the system's decision to warn or terminate an employee. As performance management becomes more data-driven, increasing numbers of employers will begin implementing similar algorithm-based systems. In using these systems, employers should be cautious to avoid the risk of penalizing employees for lawful breaks, such as those made pursuant to a reasonable accommodation.

The challenge with machine learning systems lies not in how neutrally they are coded, but with the quality of the training data provided to the program and the difficulty that lies in identifying bias within such training data. Additionally, risk exists in what the algorithm reviews; for example, if in reviewing resumes from a training data set, algorithms are able to determine a person's age, gender, race, or other protected characteristic(s), it may begin to impermissibly consider these traits. Simply put, the end result may be bias in, bias out.

For companies that are considering using algorithms at any stage of the employment process, such as in performance management (e.g., from bulk screening of resumes and applications to analyzing a candidate's online interviews), the following efforts by legislators and activists to AI may be of interest:

## Federal, State and Local Measures

In response to the above-outlined concerns, governments at all levels have begun creating task forces to study and regulate the use of algorithms. Some have further proposed and/or passed legislation aimed at the ethical development and use of AI technology.

### **AI Task Forces**

For example, New York City enacted the first algorithm accountability law in the United States in December 2017 when it passed [Local Law 49](#). The measure established a task force, appointed by the mayor and consisting of "persons with expertise in the areas of fairness, accountability and transparency relating to automated decision systems," to review the City's "automated decision systems."

The task force issued a final [report](#) in November 2019 containing several recommendations regarding the city's use of algorithms. The report calls for the creation of an "Organizational Structure within City government that would serve as a centralized resource for guiding agency management of [algorithmic decision systems] and carrying out city wide management functions." This includes developing citywide best practices, providing support to agencies in the development and implementation of algorithms, and receiving public input in the development of policies and protocols. In response to the task force's report, the mayor signed an executive order on November 19, 2019, establishing an Algorithms Management and Policy Officer within the Mayor's Office of Operations, who will serve as the centralized resource on the subject.

Vermont launched a similar task force charged with drafting recommendations for state regulation of AI, which are expected to be issued in a report by no later than January 15, 2020. Likewise, in May 2019, the Alabama legislature passed a [resolution](#) creating the Commission on Artificial Intelligence and Associated Technologies. The Commission is tasked with reviewing and advising the state "on all aspects of the growth of artificial intelligence" and its use in health care, education, transportation, and other industries, and is expected to submit a final report by May 1, 2020.

### **AI Regulation**

While some states have created task forces aimed at examining their relationships with AI, other states have taken more sweeping measures. For example, on March 28, 2019, Idaho Governor Brad Little signed into law a bill requiring public disclosure of all documents, data, and other information used to create a pretrial risk assessment tool. The purpose of the law is to provide public access to this information for purposes of public inspection, auditing, and testing. The law also provides criminal defendants the right to review "all calculations and data used to calculate the defendant's ... risk score" where a court or expert witness have used or relied upon a pretrial risk assessment tool in any respect, including for sentencing purposes.

In a similar vein, Illinois' Artificial Intelligence Video Interview Act was signed into law on August 9, 2019, and becomes effective on January 1, 2020. The [Act](#) imposes strict limitations on employers who use videotaped interviews for recruiting job candidates, and provides that employers can only use videotaped interviews where the applicant is notified in advance that their interview may be analyzed by AI (to evaluate the applicant's facial expressions and other relevant characteristics), is given information regarding how the AI works and what characteristics it will use to evaluate them, and consents (orally or in writing) to being evaluated by such AI.

The Act also provides that employers may only share the videos with individuals who have the requisite expertise to evaluate an applicant's fitness for a position, and affords applicants the right to request that their videotaped interview be destroyed (including copies) within 30 days of the receipt of such request.

On May 20, 2019, New Jersey legislators introduced a similar bill, A5430, into the New Jersey General Assembly. Dubbed the New Jersey Algorithmic Accountability Act, the [measure](#) would require large businesses to conduct impact assessments of their automated decision systems with the help of independent auditors and technology experts. The NJAAA would also require these business to make a record of any indication of racial or other bias or threats to any consumer's personally identifiable information discovered by the assessment. Of course, the discoverability of such assessments would be a prime area of concern for any employer using such AI in New Jersey. Finally, the NJAAA would permit the state to bring a civil action against businesses whose practices adversely affect residents. The bill has thus far stalled since its introduction.

Washington State legislators introduced companion bills in the Washington House and Senate on January 23, 2019, aimed at addressing the impact of algorithms, "including potential bias, inaccuracy, or disparate impacts." The measure would require state agencies to make their automated systems and data sets "freely available" to the public before, during, and after deployment to permit independent third parties to test, audit, and research such systems.

The measure also prohibits public agencies from using automated systems that discriminate against individuals or that treat individuals less favorably than others. Finally, the measure permits individuals who are discriminated against by an agency's automated system to bring suit against that agency. Notably, both bills have stalled in their respective committees since their introduction.

## Federal Action

On February 11, 2019, President Trump issued an executive order, "Maintaining American Leadership in Artificial Intelligence." The order tasks federal agencies with prioritizing AI in terms of research, development, and issuing grants and identifying federal datasets suitable for public access that will facilitate AI research and development by the private sector. The order also requires the National Institute of Standards and Technology ("NIST") and the National Science and Technology Council to provide guidance regarding AI regulation and development in the United States.

On April 10, 2019, in apparent tension with President Trump's executive order, congressional legislators introduced the Algorithmic Accountability Act into the House and Senate. The [Act](#) was introduced by Representative Yvette Clarke (D-NY) in the House of Representatives and by Senator Cory Booker (D-NJ) and Senator Ron Wyden (D-OR) in the Senate. Significantly, if passed, the Act would be the first federal law aimed at regulating the use of algorithms by private companies and would empower the Federal Trade Commission to create regulations requiring individuals and companies to conduct impact assessments evaluating their use of "automated decision systems" (as defined by the Act) and to timely address any identified biases. As currently drafted, the Act would apply to companies with revenues in excess of \$50 million per year, that possess information relating to at least one million people or devices, or that act as data brokers who buy and sell consumer data.

Thereafter, on August 9, 2019 and consistent with President Trump's executive order, NIST released its plan for "Federal Engagement in Developing Technical Standards and Related Tools." The plan recommends that the federal government "commit to deeper, consistent, long-term engagement in AI standards development activities to help the United States to speed the pace of reliable, robust, and trustworthy AI technology development," and recommends that the federal government "[b]olster AI standards-related knowledge, leadership, and coordination among [f]ederal agencies to maximize effectiveness and efficiency," "[p]romote focused research to advance and accelerate broader exploration and understanding of how aspects of trustworthiness can be practically incorporated within standards and standards-related tools," "[s]upport and expand public-private partnerships to develop and use AI standards and related tools to advance reliable, robust, and trustworthy AI," and "[s]trategically engage with international parties to advance AI standards for U.S. economic and national security needs." To maintain leadership in the realm of AI, NIST officials noted in the plan that "[a]ctive involvement and leadership by the private sector, as well as academia, is required."

The United States is not alone in these efforts to regulate AI, which have also been introduced abroad.

## AI Legislation Abroad

In addition to introducing sweeping privacy protections, the European Union's General Data Protection Rules also provides protection from fully automated decision-making. Specifically, Article 22 of the GDPR provides that an individual "shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."

This provision effectively prevents companies from engaging in automated decision-making without human intervention that affects an individual's legal rights or that significantly affects an individual, such as completely automated recruiting. Under Article 22, companies can only engage in fully automated decision-making affecting legal or significant rights if required for entering or performing a contract between an organization and an individual, if authorized by law, or if the individual provides explicit consent.

Additionally, on April 8, 2019, the European Union published guidelines for the development of trustworthy and ethical AI. The guidelines include the following non-exhaustive list of requirements developers should keep in mind when creating and/or employing AI: human agency and oversight; technical robustness and safety; privacy and data governance, transparency; diversity, non-discrimination, and fairness; societal and environmental wellbeing; and accountability. While developers are not legally required to adhere to them, the guidelines set the stage for further legislative developments by the EU.

The United Kingdom's Centre for Data Ethics and Innovation ("CDEI") published its interim [reports](#) on algorithmic bias and online targeting on July 19, 2019, which focused on two public sector uses of algorithms (policing and local government) and two private sector uses of algorithms (financial services and recruitment).

In the employment context, and specifically with respect to financial services, the report warned against the introduction of bias through proxy characteristics (i.e., characteristics that an algorithm could evaluate in place of a protected characteristic to produce the same discriminatory result) when using algorithms to assist in credit and insurance pricing decisions. Noting the challenges Amazon faced in 2017, the report contrasted the benefits of using algorithms to assist in hiring decisions against the potential risk of discrimination and noted that the CDEI will work with companies to fully explore bias-mitigation options to shape industry standards. A final report with recommendations will be published in March 2020.

The EU and the UK are not alone in monitoring AI in the international arena. The Organisation for Economic Co-operation and Development recently released its [Recommendation on Artificial Intelligence](#) containing principles for the responsible development of AI. The OECD summarizes its five principles as follows:

AI should benefit people and the planet by driving inclusive growth, sustainable development and well-being [and] should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and [i]nclude appropriate safeguards—for example, enabling human intervention where necessary—to ensure a fair and just society. There should be transparency and responsible disclosure around AI systems to ensure that people understand when they are engaging with them and can challenge outcomes. AI systems must function in a robust, secure and safe way throughout their lifetimes, and potential risks should be continually assessed and managed. Organisations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning in line with the above principles.

The OECD also calls on governments to, among other things, facilitate public and private investment into the research and development of trustworthy AI. Although these principles are not legally binding, every member nation of the OECD, including the U.S. and six non-member nations (Argentina, Brazil, Colombia, Costa Rica, Peru, and Romania) have endorsed them.

## Activism and Calls for Reform

Many organizations, academics, and institutions have called on legislatures to take additional proactive measures to regulate AI out of concern for, among other risks, algorithmic bias. The American Civil Liberties Union is one such organization, and has published several articles and podcast episodes discussing algorithmic bias. Its publications warn against the use of algorithms in recruiting tools, by governments in predictive policing, by judges to assist with sentencing, in education, and for financing and lending decisions.

Likewise, other organizations have given considerable thought to the development and increasing prevalence of AI. Developed in 2017 by the Future of Life Institute, the Asilomar AI Principles are a group of 23 principles designed to guide developers in creating responsible and beneficial AI. The Principles are subdivided into three categories: research, ethics and values, and longer-term issues.

With respect to research, the principles call for AI research to prioritize beneficial research that addresses “thorny questions in computer science, economics, law, ethics, and social studies.” The research principles also call for a constructive dialogue between researchers and policy-makers, transparency between researchers and developers of AI, and caution against “corner-cutting on safety standards.”

The ethics and values principles call for the safety and transparency of algorithms and charge developers of AI systems as “stakeholders in the moral implication of their use, misuse, and actions, with a responsibility and opportunity to shape those implications,” as well as seek to align AI's goals with human values such as dignity, rights, freedoms, and diversity.

Finally, the longer-term principles caution against the profound impact that AI and its upper limits could have, and call for the implementation of mitigation efforts to ameliorate risks posed by AI systems.

The Principles are the most widely supported principles released regarding AI to date, as they have been signed by 1,273 robotics researchers and over 2,500 individuals, including Elon Musk and the late Stephen Hawking. Notably, California unanimously adopted a measure expressing the legislature's support for the Principles.

The Brookings Institution also issued [guidance](#) with respect to algorithms on May 22, 2019, when it published a report highlighting various areas where algorithms can cause bias, including in hiring, online advertisements, facial recognition technology, and criminal justice. The extensive and in-depth report calls for bias detection strategies, self-regulatory practices, and for changes to antidiscrimination laws to include protections against algorithmic bias. In addition, the report encouraged organizations that rely on algorithms to conduct a bias impact statement to help identify where bias may be infecting their algorithms and provided guidance on how to conduct such an impact statement.

## Federal Law and Algorithmic Bias

Title VII prohibits discrimination on the basis of any protected characteristic (e.g., sex, race, color, disability, age, military or veteran status, etc.). Generally speaking, plaintiffs filing a lawsuit claiming that they were discriminated against in violation of Title VII must allege disparate treatment, disparate impact, or both. Most states have analogues to Title VII, many with even broader protections (including California and New York).

In order to plead a cognizable disparate treatment claim, a plaintiff must demonstrate that the employer treated them differently than other employees on the basis of their status as a member of a protected class and that such treatment was motivated by discriminatory intent. Proof of discriminatory motive is critical, and failure to establish that an employer intended to discriminate is fatal to the claim. For example, a plaintiff alleging race discrimination must demonstrate that their employer treated them less favorably than other similarly situated employees at work, and that the employer intended to treat them less favorably because of their race.

Some commentators have argued that because algorithms, as machines, do not have discriminatory motives or intent, plaintiffs cannot succeed on a disparate treatment claim on that basis. Specifically, they note that because algorithms merely identify and replicate patterns, any discrimination that may result is by its very nature incidental, and thus, not intentional.

However, some courts around the country have permitted disparate treatment claims to proceed based on allegations of unconscious or implicit bias. That is, a plaintiff may advance a claim of disparate treatment where they allege to have been treated less favorably because of a supervisor or employer's implicit or unconscious bias against them (because of their protected characteristic).

In addition to bringing a claim alleging disparate treatment, a plaintiff alleging a violation of Title VII might also bring a disparate impact claim. To prevail on such a claim, a plaintiff must show that a facially neutral employment practice disproportionately impacts or burdens a protected group. Significantly, no showing of intent is required. Courts analyzing disparate impact claims have often relied on statistical significance to demonstrate where an employment practice violates Title VII. Under a statistical significance approach, most researchers rely on a 95% confidence level, that is, the calculation is 95% percent certain the disparity is reflective of an existing disparity in the labor market and is not a result of discrimination. As some commentators have noted, because AI is data-driven, it is theoretically possible for a plaintiff to argue that an AI-driven employment practice has a disparate impact on members of a protected class.

If a plaintiff were to succeed in making an initial showing that an employment practice causes a disparate impact, one hurdle would remain: the business necessity defense. Under the defense, the employer would have an opportunity to demonstrate that the challenged practice has some relationship to the job itself and that the algorithm's predictions are accurate. Given the difficulty of examining an algorithm to determine precisely how it functions and the algorithm's predictive purpose, it is unclear whether a plaintiff could overcome the business necessity defense. However, even where an employer satisfies its burden under the business necessity defense, a plaintiff may still succeed on a disparate impact claim by demonstrating that a less discriminatory alternative practice exists that could serve the same job performance-related business interest.

As cases alleging algorithmic bias in violation of Title VII or state analogues are brought, courts will weigh in regarding how to analyze these lawsuits. In the meantime, it is foreseeable that advocates for perceived or alleged victims may attempt to argue that algorithmic bias opens the door to relief under current law, and/or that new laws and regulations should be passed to permit same.

## Conclusion

There is no question that AI offers organizations an opportunity to save resources by streamlining systems at a rate never before seen. Notwithstanding this alluring benefit, employers should proceed with caution before incorporating AI into their pre-and-post hire processes to avoid the above-outlined risks (such as algorithmic bias) as well as to ensure compliance with the growing list of laws and regulations.