PUBLICATION

Artificial Intelligence and Bias: Considerations to Prevent Bias and Mitigate Legal Risk of Employers

Authors: Aldo M. Leiva, Nakimuli O. Davis-Primer
Third Quarter 2020

Artificial intelligence (AI) tools have been implemented by employers in an ongoing effort to reduce recruitment and hiring costs by automating sourcing of applicants and screening candidates or by applying AI tools to the applicant assessment process, such as conversational chatbots and video interview assessment tools that monitor applicant facial expression, word choice, vocal tone and body language. While several AI industry advocates and lobbying organizations have emphasized the potential of AI to mitigate human bias in hiring decisions, others have identified unintended but potentially disparate and discriminatory effects that the use of AI may have on protected classes of individuals, which may result in discrimination claims.[1] AI tools may, for example, favor applicants who live within the same zip code as the employer's offices in order to minimize commutes and therefore increase the likelihood of longevity of employment but such efforts may have a disparate impact by, for example, screening out individuals from zip codes that are primarily composed of minority applicants.[2] AI tools may also sample successful profiles based on current employees who are disproportionately white males, thereby creating inherent disadvantages against women and minorities.[3] In order to mitigate against potential biases and related legal risks resulting from the use of AI in recruiting and hiring, employers should fully understand the science and technology comprising such AI tools and implement appropriate controls against unintended bias. Employers should also continue to monitor and assess legal developments in this emerging technology space.

AI and Human Decision-Making: Promise and Peril

AI has been recognized as a potential tool to improve human decision-making by implementing algorithms or machine learning systems that identify and reduce human bias. The algorithms generally disregard variables that do not accurately predict outcomes based on training data provided to them and provide a means of making decisions more objective. In contrast, human decisionmakers may ignore or be unaware of implicit factors, associations, assumptions, and value judgments that result in hiring decisions. In short, humans may not recognize or effectively control their own unconscious bias(es).[4]

Nevertheless, the issue of potential unintended bias in automated systems has been identified and studied for decades. These biases may arise in various forms including introducing bias into the computer program or relying on flawed training data. In 1988, the UK Commission for Racial Equality determined that a computer program used by a British medical school to screen applicants was biased against women and individuals with non-European names. Further investigation revealed that the program had been based on historical human admission decisions,[5] thereby effectively introducing human bias. More recently, bias against African-American defendants was identified by an investigative journalism team assessing the use of a criminal justice algorithm in Broward County, Florida to offer predictions of recidivism. The system was alleged to mislabel African-American defendants as "high risk" at a rate of almost double that of mislabeled white defendants.[6] Other research studies have found that training natural language processing models using news article content can lead to gender stereotypes.[7] The use of specific training data has also been found to result in AI systems that incorporate biased human decisions and historical/social inequities, even if such variables as race, gender, or sexual orientation are removed from the data.[8] For example, Amazon discontinued using a hiring algorithm after determining that it favored applicants who used certain words (such as "executed" or "captured") that

were more commonly found on men's resumes.[9] An additional source of potential bias is flawed data sampling at the training data phase, in which certain groups are over- or underrepresented, resulting in higher error rates in, for example, facial analysis scans for minorities (particularly minority women).[10]

As research into AI bias has progressed, there have been ongoing efforts to improve AI systems by preventing them from perpetuating human or societal biases or creating unintended biases. For example, researchers have developed complex tools to introduce, understand, and measure "fairness," by using models that either have equal predictive value across different groups or require that models generate equal false positive and false negative rates across groups.[11] However, as such measures are still being developed,[12] assessed, and implemented, AI tools that are currently available to employers may still have inherent biases that may lead to legal liability.

Legal Developments in AI Bias

Based on the potential issues of AI-based bias, the EEOC has reportedly started investigating alleged discrimination in AI-based recruitment and hiring.[13] These efforts are consistent with similar investigations conducted by the Federal Trade Commission (FTC) into automated decision-making processes by financial institutions, to enforce provisions of the Fair Credit Reporting Act (FCRA) and the Equal Credit Opportunity Act (ECOA), both of which, despite having been passed in the 1970s, address automated decision-making, and have been applied to machine-based credit underwriting models for decades.[14]

Congress has similarly taken notice of AI-based bias. If passed, the Algorithmic Accountability Act of 2019 would require the FTC to establish rules for companies to assess whether algorithms and their supported automated systems are biased or discriminatory.[15] State legislatures are also focusing on this issue. The Illinois State Legislature enacted the Artificial Intelligence Video Interview Act,[16] effective January 1, 2020, which imposes several requirements on employers that use AI tools to analyze video interviews of applicants for employment, including:

1. Employers must notify applicants that AI will be used in video interviews;
2. Employers must disclose to applicants how the AI tool works, and must identify the characteristics that will be assessed by the AI tool;
3. Applicants must consent to the use of the AI tool prior to the video interview;
4. Employers must maintain the confidentiality of the content of the video interview, and may only share the content with AI experts required to implement the tool or otherwise evaluate the applicant's fitness for an employment position; and
5. Employees are required to comply with any request by the applicant to destroy the video interview, within 30 days of any such request.

Despite such requirements, however, the law includes no explicit enforcement mechanism and does not include any statutory cause of action by the applicant, should the employer fail to comply with the law.

Employer Considerations to Address AI Bias and Mitigate Legal Risk

Rather than rely on marketing communications or representations contained in AI vendor contracts, employers should thoroughly understand AI tools and ongoing research and developments in this rapidly-growing field, by consulting various organizations that monitor such issues.[17] They should conduct careful and competent assessment of AI tools prior to implementation, perhaps through pilot programs to assess bias issues. Employers should also assess their own decision-making criteria and objectives to ensure that the essential requirements of the positions at issue are clear, accurate and objective. Having objective and diverse criteria will lessen the likelihood of feeding algorithms and computer models with biased information. Similarly, employers should evaluate the make-up of their workforces and make adjustments as necessary to ensure that the training sampling data is not flawed. The computer algorithm is a tool to assist in streamlining the process,

but employers must not lose sight of their own obligations to ensure that their employment decisions do not violate federal, state and local laws by disparately impacting protected groups or otherwise resulting in discriminatory recruiting and hiring practices. Indeed, while employers' legal counsel may attempt to seek indemnification or other remedies from AI vendors who provide such tools, vendor contracts typically include non-liability or limited liability provisions that may shift most or all of the liability to the employer. Therefore, it is critical that employers actively engage in the process and assess the results to ensure there are no unintended consequences in order to reduce their legal risk.

Further, employers should consider deploying additional processes or tools to mitigate against bias, such as using internal "red teams" or third-party auditors to assess bias.[18] Employers may also factor in the human element at later stages in the process. For example, if the algorithm is used to screen applicants and maybe the first stage of interviews, a group of managers and/or human resource personnel should be used in later stages of the interview process to continue to engage in controlled processing to reduce the potential impact of bias. If the applicants progressing through the ranks are consistently nondiverse then someone should be appointed to address this issue and serve as a "bias interrupter" of sorts to lessen the likelihood of bias throughout the process. Because bias may be either introduced into AI tools by human users or, alternatively, because recruitment/hiring processes continue to rely on human assessment of AI-produced reports and recommendations, employers should continue to control for such issues by running algorithms and comparing such results to human decision makers' results, to help account for any differing outcomes between the two processes.[19] As indicated above, these actions would allow for "human in the loop" systems to incorporate and assess human decisions at critical phases of an automated process and to gauge any bias inherent in automated systems or within human users of such systems.[20]

The issues of bias, whether human or incorporated into AI tools, remain important considerations for any employer or counsel seeking to navigate this terrain for a client, given the potential legal liability and reputational risk that could impact a company found to employ AI tools or processes (human or otherwise) that have discriminatory effects. Please contact the authors – Al Leiva or Nakimuli Davis-Primer – or any member of Baker Donelson's Employment Law Group with any questions regarding effective use of this technology.

[1] See "The Legal Risks of Using Artificial Intelligence in Hiring and Recruiting," February 20, 2020, Allison Sues, Chamber Dispatch.

[2] *Id.*

[3] *Id.*

[4] See "What Do We Do About the Biases In AI," October 25, 2019, Harvard Business Review .

[5] *Id.*

[6] *Id.*

[7] *Id.*

[8] *Id.*

[9] *Id*.

[10] *Id.*

[11] *See* Note 4.

[12] *See* "Tackling Bias in Artificial Intelligence (and in Humans)," June 6, 2019, McKinsey Global Institute for a comprehensive discussion of such measures.

[13] See "AI Hiring Could Mean Robot Discrimination Will Head to Courts," November 12, 2019, Bloomberg Law News (Subscription Required).

[14] See "Using Artificial Intelligence and Algorithms," April 8, 2020, Federal Trade Commission.

[15] See Algorithmic Accountability Act of 2019.

[16] See Artificial Intelligence Video Interview Act.

[17] Potential resources include the AI Now Institute's annual reports, the Partnership on AI, and the Alan Turing Institute's Fairness, Transparency, and Privacy group.

[18] See Note 4, referencing such tools/services as Google AI's technical services, and IBM's Fairness 360 framework.

[19] Id.

[20] Id.