Santa Clara High Technology Law Journal

Volume 39 | Issue 1 Article 1

8-1-2022

DARK SYSTEMS: REPROGRAMMING ARTIFICIAL INTELLIGENCE REGULATIONS TO PROMOTE FAIRNESS AND EMPLOYMENT NONDISCRIMINATION

Wennagel, Robert

Recommended Citation

Wennagel, Robert, *DARK SYSTEMS: REPROGRAMMING ARTIFICIAL INTELLIGENCE REGULATIONS TO PROMOTE FAIRNESS AND EMPLOYMENT NONDISCRIMINATION*, 39 SANTA CLARA HIGH TECH. L.J. 1 ().
Available at: https://digitalcommons.law.scu.edu/chtlj/vol39/iss1/1

DARK SYSTEMS: REPROGRAMMING ARTIFICIAL INTELLIGENCE REGULATIONS TO PROMOTE FAIRNESS AND EMPLOYMENT NONDISCRIMINATION

Robert Wennagel*

Automated decision-making ("ADM") systems, whether deploying artificial intelligence, machine learning, or other algorithmic processes, have become ubiquitous in modern life, but their use is often unnoticed or invisible to society at large. Currently no federal laws require notice or disclosure to individuals when an ADM is used to collect their data, evaluate them, or make determinations about their lives. This is particularly concerning for the employment relationship because notice and transparency are essential for personal privacy, and the surreptitious use of ADM systems deprives applicants and employees of the ability to understand employers' decision-making processes and to seek redress under applicable antidiscrimination laws. Some state and local governments have recognized this danger and have taken initial steps to protect applicants and employees, while the European Union has proposed a sweeping AI regulation that would govern all phases of development of such systems there. This article proposes a system of regulations based on notice and transparency that takes into consideration existing laws governing the employment relationship and complements those laws in order to produce a legal framework that promotes applicant and employee rights, while also allowing flexibility for the development of ADM systems that benefits employees, employers, and society.

[*] LL.M., Drexel School of Law (2022); J.D., UCLA School of Law (2005). I owe a debt of gratitude to Professor Neal Wiley at Drexel, who provided patient and wise guidance throughout the process of drafting this Article. Special thanks to Janelle Barbier and the staff of Santa Clara High Technology Law Journal for their consideration, editing acumen, and providing a home for this work.

CONTENTS

I. INTRODUCTION

Automated decision-making ("ADM") systems have become ubiquitous, often without our knowledge as consumers, employees, and citizens. For better or worse, they underpin processes such as the advertisements we see, the loan terms for our cars and houses, government benefits, college admission, and job applications.[1] The systems have been steadily permeating daily operations in both business and government, and this rise has only been hastened by the growth of machine learning ("ML") systems. From cradle to grave, mathematical algorithms measure, track, categorize, and score people based on often opaque formulas. Despite their promise, big data analytics may threaten long-standing civil rights protections, and few people understand such systems' ubiquity or impact on their lives.[2] According to recent research, less than half of people were familiar with the fact that a computer program may be solely responsible for reviewing their job application.[3]

Often classified by their marketers and proponents as artificial intelligence ("AI"), modern systems actually bear little resemblance to a "general artificial intelligence" that could act with common sense and understanding equivalent to a human, and these modern systems also suffer from several significant shortcomings. What they are good at - what they are very good at - is the analysis of large amounts of data.[4]

[1] *See* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV. 671, 673, 679 (2016); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 2 (2016); Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1252 (2008) (exploring the use of automated decision-making systems by governments and its due process implications); Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 2–3 (2014).

[2] EXEC. OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES, at III (May 2014), https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf (finding that "big data analytics have the potential to eclipse longstanding civil rights protections in how personal information is used in housing, credit, employment, health, education, and the marketplace").

[3] *See* Ifeoma Ajunwa, *An Auditing Imperative for Automated Hiring Systems*, 34 HARV. J.L. & TECH. 621, 628–29 (2021); Aaron Smith & Monica Anderson, *Automation in Everyday Life*, PEW RSCH. CTR. (Oct. 4, 2017), https://www.pewresearch.org/internet/2017/10/04/automation-in-everyday-life/.

[4] Jeremias Adams-Prassl, *What If Your Boss Was an Algorithm? Economic Incentives, Legal Challenges, and the Rise of Artificial Intelligence at Work*, 41 COMP. LAB. L. & POL'Y J. 123, 127–28 (2019); MELANIE

For example, ADM systems have been adopted by employers to evaluate large influxes of job applications and resumes. Many companies use ADM systems that automatically review resumes and produce scores correlating those job applicants to the required job description without any human intervention.[5] More advanced systems deploy machine learning systems to review available data and predict whether applicants may be a successful fit at the target company. These ADM systems have sometimes given rise to unfair hiring practices or replicated historical discrimination, necessitating abandonment of the projects altogether.[6] Faulty design or merely lack of proper oversight can give rise to data correlations that ultimately cause disparate impact discrimination even when the systems are blinded and do not receive direct inputs regarding protected information, such as race and gender.[7] For example, a system for judging applicants may discover a job correlation between gaps in an employee's work history and low potential tenure at their new job. If the ADM system weighed such gaps against an applicant, however, it would have a disproportionately negative impact on protected individuals who have a higher likelihood of taking leaves of absence or temporarily leaving the workforce, such as people suffering from disabilities, military service members, or women of child-bearing age.

MITCHELL, ARTIFICIAL INTELLIGENCE: A GUIDE FOR THINKING HUMANS 190–94 (2019) (discussing Google's Word2vec model, including the vast resources drawn from and correlations it made for purposes of natural language processing).

[5] Ajunwa, *supra* note 3, at 626, 685–97 (surveying current products and their automated features).

[6] Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. 857, 874 (2017); *see also* Ajunwa, *supra* note 3, at 635. Amazon designed recruiting software that mirrored past patterns of sex discrimination due to its workforce being primarily male. To its credit, after an investigation, Amazon's internal testing and validation discovered the problem and cancelled the project; Alex C. Engler, *Independent auditors are struggling to hold AI companies accountable*, FAST CO. (Jan. 26, 2021),

https://www.fastcompany.com/90597594/ai-algorithm-auditing-hirevue; Maya Oppenheim, *Amazon Scraps 'Sexist AI' Recruitment Tool*, INDEPENDENT (Oct. 11, 2018),

https://www.independent.co.uk/life-style/gadgets-and-tech/amazon-ai-sexist-recruitment-toolalgorithm-a8579161.html.

[7] Allan G. King & Marko Mrkonich, *"Big Data" and the Risk of Employment Discrimination*, 68 OKLA. L. REV. 555, 556 (2016); Woodrow Hartzog & Evan Selinger, *Big Data in Small Hands*, 66 STAN. L. REV. ONLINE 81, 83 (2013) (reviewing how protected class information can be determined from online data, such as social media).

Artificial intelligence also poses significant risks for employee privacy. Machine learning systems and deep neural networks depend on large amounts of training data to learn about the correlations between various inputs and desired outputs.[8] For example, some software developers and employers have exploited the explosion of workplace data to develop and market products that use AI systems to monitor employee behavior, even at those employees' homes, without providing notice or receiving consent for doing so.[9]

Despite significant potential problems, the deployment of artificial intelligence in the workplace has the potential to greatly benefit employees, employers, and customers. There are tremendous efficiency gains to be made. The management of large, heterogeneous workforces is challenging for even the most skilled human resources departments, but if artificial intelligence has proven anything, then it has shown its ability to efficiently collect, sort, and evaluate vast amounts of data.[10] Artificial intelligence systems may be deployed to aid employee recruitment, scheduling and workforce management, and employee performance management.[11]

ADM systems, when used thoughtfully, may also promote fairness and antidiscrimination. Employers may use algorithms to make hiring and promotion decisions based on machine learning criteria that carefully consider past historical injustices.[12] There are numerous different models that may be applied to assist systems and employers in evaluating whether AI practices promote fairness.[13]

ADM systems and AI therefore hold both great promise and the potential for significant, and often invisible, harm. When automated decision-making systems invisibly replicate historical

[8] Barocas & Selbst, *supra* note 1, at 680.

[9] *See* Zoë Corbyn, *Bossware is coming for almost every worker: the software you might not realize is watching you*, GUARDIAN (Apr. 27, 2022), https://www.theguardian.com/technology/2022/apr/27/remote-work-software-home-surveillance-computer-monitoring-pandemic.

[10] MITCHELL, *supra* note 4, at 214–21 (discussing IBM's Watson and its abilities related to natural language processing and performance on the quiz show Jeopardy!); Jeremy Nunn, *The Role Of AI Technologies In HR Data-Based Decision Making*, FORBES (July 3, 2019), https://www.forbes.com/sites/forbestechcouncil/2019/07/03/the-role-of-ai-technologies-in-hr-data-based-decision-making/?sh=222580a23f4e; Citron, *supra* note 1, at 1252.

[11] *See infra*, Section II.A.

[12] Jason R. Bent, *Is Algorithmic Affirmative Action Legal?*, 108 GEO. L.J. 803, 805 (2020).

[13] Doaa Abu-Elyounes, *Contextual Fairness: A Legal and Policy Analysis of Algorithmic Fairness*, 2020 U. ILL. J.L. TECH. & POL'Y 1, 7 (2020).

discrimination and bias, or when their inherent design limitations promulgate other unfair treatment of employees, I refer to them as "dark systems."[14]

At this time, the federal government in the United States has not engaged in any systematic effort to regulate such AI or "dark systems," and there are no federal regulations in place that govern employers' collection or use of personal data through ADM systems.[15] Some cities and states have taken initial steps to evaluate and regulate AI, but when doing so, they have not given adequate consideration to the existing legislative framework and the unique problems found in the employment context.[16] Instead, they mimic data privacy legislation while ignoring the specific issues arising in employment in the United States.[17] This privacy-centric approach emphasizes notice and transparency, generally following the pattern of the European Union's General Data Protection Regulation ("GDPR"), which requires data controllers deploying automated decision-making technology to provide data subjects with "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject."[18] California, the first state to pass a comprehensive data privacy regulation, plans to address ADM systems as part of its corresponding data privacy regulations.[19]

[14] I have adopted this term because it seems a fitting parallel to the "dark patterns" described in data privacy regulations and often deployed for online marketing and e-commerce purposes. Dark patterns were first defined in 2010 as "a user interface carefully crafted to trick users into doing things they might not otherwise do . . . with a solid understanding of human psychology, and [which] do not have the user's interests in mind." Harry Brignull, *Dark Patterns: inside the interfaces designed to trick you*, VERGE (Aug. 29, 2013),
http://www.theverge.com/2013/8/29/4640308/dark-patterns-insidethe-interfaces-designed-to-trick-you.

[15] Ajunwa, *supra* note 3, at 623 (noting that when these employees are deterred from even completing an application, such systems may "discreetly and disproportionately cull the applications of job seekers who are from legally protected classes").

[16] *See infra*, Section III.A.3.

[17] *See infra*, Sections III.A.3 and III.B.2.

[18] Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), art. 15(1)(h), 2016 O.J. (L 199) 1,
https://eur-lex.europa.eu/eli/reg/2016/679/oj [hereinafter "GDPR"].

[19] California Privacy Rights Act, 2018 Cal. Stat. 1807 (to be codified at CAL. CIV. CODE §§ 1798.100-199.100 (effective Jan. 1, 2023)) [hereinafter "CPRA"].

At this time, however, only Illinois and New York City have passed specific legislation addressing AI for employment purposes.[20] Illinois has sought to address the use of artificial intelligence when used for evaluating applicants during video interviews, and New York City has passed an ordinance that requires notice, the right to opt-out, and independent audits whenever an ADM system is used for hiring or promotional purposes.

Notice and transparency certainly benefit employee privacy, and, if properly implemented for the employer-employee relationship, could also significantly aid in the enforcement of existing antidiscrimination laws in the United States. Unfortunately, data privacy regulations do not fully address the unique problems of AI when used in the workplace.[21] Early opt-out systems, such as those contemplated by general data privacy laws and, in the case of the New York City AI Ordinance, undermine the capabilities of machine learning systems by depriving them of essential data necessary for functioning and learning, replacing data which those systems may use to improve their predictions with data biased due to self-selection.[22] Although such opt-out requirements may be intended to mitigate potential algorithmic errors by incorporating a human in the loop, they do so at an inappropriate time. Furthermore, without conveying any specifics about the ADM systems, the data that they utilize, or how they function, notification schemes also miss the opportunity to provide applicants and employees the information that would potentially aid them in identifying cases of disparate impact discrimination.

Likewise, current requirements for independent audits stand on faulty premises. Where bought and paid for by AI developers and employers, such audits have little chance of fairly evaluating or remediating disparate impact discrimination. Unlike more well-established procedures for financial audits, no guidelines exist for how AI audits should be performed. The contents and conclusions of any such audit are therefore easily open to manipulation for financial and commercial gain. Even if a regulator does provide proper guidelines, however, auditing ADM systems in the development stage will likely lead to significant gaps in understanding how they will behave once

[20] *See infra*, Section III.A.3.

[21] Automated decision-making systems, through both their use and their access to private data, also carry cybersecurity risks, the potential for data breaches, and the possibility of outright manipulation. As these concerns are not unique to the employment context, however, they can likely be addressed through more general regulations on data privacy and data breaches.

[22] *See infra*, note 190.

they are deployed to evaluate and make decisions about applicants and employees in real-time and with real data.

Given the shortcomings with current legislation and guidelines, this paper will propose regulatory protections that are appropriately tailored to address how future legislation can most effectively regulate ADM software in conjunction with existing data privacy and antidiscrimination laws.[23] Automated decision-making software, when used in the employment context, should primarily be transparent and subject to review by humans. With respect to transparency, any time that an ADM system has been utilized to make or assist in making an adverse employment decision, that use should be disclosed to the affected applicant or employee, along with information regarding what data was used for the system's inputs, what decision or output was produced by the ADM system, meaningful information about how the system arrived at its decision or output, and basic information about the ADM system such as its name and version. After an adverse decision has been made and the requisite notice has been provided, an applicant or employee should be provided the right to request human review of the decision. This system of transparency and human oversight will mitigate potential privacy issues, foster the responsible development and deployment of ADM systems by software developers and employers, encourage the timely review and correction of faulty decisions, and assist applicants and employees in enforcing their rights under existing antidiscrimination laws.

[23] Although many scholars have argued that current antidiscrimination laws, in particular the burden-shifting analysis under disparate impact discrimination, will provide ineffective in evaluating artificial intelligence decision-making, these claims are entirely speculative. No court has, as of this writing, actually utilized the disparate impact discrimination framework to evaluate an AI system in employment. Furthermore, courts have historically proven their capability of adapting antidiscrimination standards to meet the challenges of the times. *See Griggs v. Duke Power Co.*, 401 U.S. 424 (1971) (adopting the disparate impact discrimination theory). Many scholars have also proposed methods of how antidiscrimination law can effectively evaluate artificial intelligence systems. Kim, *supra* note 6, at 909–36 (arguing that Title VII directly prohibits "classification bias"); Stephanie Bornstein, *Antidiscriminatory Algorithms*, 70 ALA. L. REV. 519, 526 (2018) (proposing an anti-stereotyping approach to algorithmic discrimination, and stating that "stereotype theory allows [disparate treatment] to reach intentional actions that incorporate or are infected by even unrecognized bias").

II. AUTOMATED DECISION-MAKING, AI, AND MACHINE LEARNING SYSTEMS

A. *Current Capabilities and Technology*

ADM systems may be based on a variety of technologies, including those currently available and those which have been conceptualized but not realized by today's technical capabilities.[24] General artificial intelligence, also known as broad artificial intelligence, consists of yet-to-be realized systems that can emulate and even improve upon a broad class of human capabilities and intelligence. Despite many years of pursuit, such systems are likely still many years away.[25] Current artificial intelligence, or "narrow artificial intelligence," is capable of a great deal, from beating the best chess players in the world to high-frequency stock trading.[26] These systems, however, lack fundamental human capacities such as common

[24] This article will generally use the term automated decision-making system and artificial intelligence interchangeably, although artificial intelligence is only one method that an ADM system may incorporate. When regulators have endeavored to define such terms, they have tended to pursue technology-neutral definitions that are quite broad. The New York City Ordinance uses the term "automated employment decision tool," which it defines as: "[A]ny computational process, derived from machine learning, statistical modeling, data analytics, or artificial intelligence, that issues simplified output, including a score, classification, or recommendation, that is used to substantially assist or replace discretionary decision making for making employment decisions that impact natural persons." N.Y.C. MUN. CODE § 20-870 (2022). The E.U. AI Proposal utilizes a similar definition. *See* E.U. AI Proposal art. 3(1). ("'[A]rtificial intelligence system' (AI system) means software that is developed with one or more of the techniques [including machine learning, symbolic reasoning, or statistical approaches], for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.").

[25] MITCHELL, *supra* note 4, at 46.

[26] *Id.* at 245; Jack Kelly, *Artificial Intelligence Is Superseding Well-Paying Wall Street Jobs*, FORTUNE (Dec. 10, 2019), https://www.forbes.com/sites/jackkelly/2019/12/10/artificial-intelligence-is-superseding-well-paying-wall-street-jobs/?sh=7df3c57e524d.

sense, the ability to reason by analogy,[27] understanding of context,[28] perceptive abilities tying them to the outside world, and transfer learning.[29] The paradox recognized by AI researchers is that what is easy for humans is hard for computers.[30] Without these abilities, machine learning systems may be able to sort through and make deep correlations between mountains of data that would take humans lifetimes, but those same systems have persistent difficulties addressing dynamic real-world scenarios that humans regularly face and take for granted, such as driving cars in the rain or through heavily trafficked roads while under construction.[31]

Many of the current advancements and business applications of artificial intelligence have emerged from a field called machine learning.[32] Machine learning, or "ML," generally refers to computer programs, and in particular algorithms, that can improve through experience and the use of voluminous sets of data.[33] Such programs may be referred to as learning without being explicitly programmed to do so.[34]

Machine learning systems come in several varieties. Two of the main subtypes are supervised and unsupervised systems.[35] A

[27] GARY MARCUS & ERNEST DAVIS, REBOOTING AI: BUILDING ARTIFICIAL INTELLIGENCE WE CAN TRUST 62 (2019) ("In the end, deep learning just ain't that deep. It is important to recognize that, in the term deep learning, the word 'deep' refers to the number of layers in a neural network and nothing more. 'Deep,' in that context, doesn't mean that the system has learned anything particularly conceptually rich about the data it has seen.").

[28] MITCHELL, *supra* note 4, at 245.

[29] Adams-Prassl, *supra* note 4, at 128 (noting that this may be referred to as "Polanyi's paradox," after Michael Polanyi, who observed that "We know more than we can tell"); MITCHELL, *supra* note 4, at 235–65.

[30] MITCHELL, *supra* note 4, at 33 (quoting Marvin Minsky, *Easy things are hard*).

[31] *Id.* at 235–65, 267–70; MARCUS & DAVIS, *supra* note 27, at 149–79; SEAN GERRISH, HOW SMART MACHINES THINK 37–56 (2018); Adams-Prassl, *supra* note 4, at 128 (noting self-driving cars are distracted by environmental hazards such as ice, snow, and faded road markings).

[32] JERRY KAPLAN, ARTIFICIAL INTELLIGENCE: WHAT EVERYONE NEEDS TO KNOW 31–43 (2016).

[33] *Id.* at 27–28.

[34] *See* A.L. Samuel, *Some Studies in Machine Learning Using the Game of Checkers. II - Recent Progress*, 11 IBM J. 601, 601 (1967).

[35] MITCHELL, *supra* note 4, at 103. Unsupervised and supervised systems should be distinguished from reinforcement learning, which is often successfully used as a method for machine learning systems to learn and play games. These systems have enjoyed great success in mastering games with constrained borders and defined parameters. A chess set, for example, has a

supervised system typically learns through a process that is highly monitored by computer scientists and software engineers.[36] All of these real-world uses of machine learning systems require large amounts of training data.[37] If a machine learning system is to be used for predicting whether a job applicant will be a productive employee, a company may feed human resources data into the system, and the algorithms will then be able to make correlations between different data points (inputs) in order to issue scores or associations (outputs) on those candidates and their potential future job success. This same process could theoretically be used throughout the employee life cycle to predict employee turnover, potential for promotion, or when poor performance may suggest the need for termination.

An unfortunate limitation on AI and machine learning programs of this type, or of any current type, is that their training does not give them the ability to effectively address or solve new problems.[38] They lack what is commonly called transfer learning, or the ability to understand common principles based on similar situations or encounters. Whereas a child may play checkers and learn some basic ideas of how to move and take their opponent's pieces, which could then be applied when learning chess, AI and ML programs have not demonstrated this same innate human talent.[39] Therefore, in the employment context, an ADM system programmed to optimize the selection of job candidates from certain job descriptions and past performance data likely will need to be completely reprogrammed and retrained for the evaluation of current employees based on their day-to-day performance metrics.

B. *The Shortcomings Inherent in Current AI and ML Models*

1. Intentional Discrimination

Although it may be overlooked, artificial intelligence and machine learning deployed for the ostensible purposes of efficiency and fairness have the potential for outright abuse and intentional discrimination. Algorithms may be coded to search for or market job

finite number of pieces, each with known attributes, and delineated spaces on the board to which those pieces can move. The real world, on the other hand, is awash in a constant stream of objects and, for most purposes, is not bound or constrained in the same manner that a game is. *Id.* at 171–73.

[36] STUART RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 671–74, 840 (4th ed. 2022) (Global Edition).

[37] *Id.* at 100.

[38] *Id.* at 166.

[39] *Id.*

opportunities based on illegal criteria, such as age.[40] Others may be tempted to deploy an ADM because they know that it will have a disparate impact, or they set the system up with operating parameters that they know will result in discrimination.[41] The advantage to the perpetrator is that they can obfuscate their actions by placing any blame on the algorithm, while simultaneously claiming that any correlations the system found are defensible as they are "job-related and consistent with business necessity" under applicable antidiscrimination law.[42] When mining large amounts of data, ADM systems may also be able to determine sensitive information about an applicant or employee that was otherwise unascertainable. Access to this may either bias the decision-maker or allow them to take a discriminatory action they otherwise could not have without the system's input.[43]

In addition to these problems, an automated decision-maker may make the independent determination that discrimination benefits the employer or achieves whatever target goals that it is programmed to meet.[44] For example, it could discriminate against disabled workers, people predisposed to genetic disease, or older workers in order to lower healthcare costs, or it could discriminate against women of child-bearing age in order to reduce aggregate costs associated with leaves of absence.[45]

2. Disparate Impact Discrimination and the Replication of Bias

ADM systems may give rise to unintentional bias, sometimes referred to as "disparate impact discrimination," for a number of reasons. Current machine learning systems require a large amount of

[40] Ifeoma Ajunwa, *Age Discrimination by Platforms*, 40 BERKELEY J. EMP. & LAB. L. 1, 5 (2019); Julia Angwin, Noam Scheiber & Ariana Tobin, *Facebook Job Ads Raise Concerns About Age Discrimination*, N.Y. TIMES (Dec. 20, 2017), https://www.nytimes.com/2017/12/20/business/facebook-job-ads.html.

[41] Kim, *supra* note 6, at 884.

[42] *See infra*, Section III.A.1.

[43] Kim, *supra* note 6, at 885 (noting Target's ability to infer shoppers' pregnancies by mining data from their buying habits).

[44] Charles A. Sullivan, *Employing AI*, 63 VILL. L. REV. 395, 402 (2018).

[45] *Id.* Sullivan cites numerous sources, including the following: Sharona Hoffman, *Big Data and the Americans with Disabilities Act*, 68 HASTINGS L.J. 777, 793 (2017); Berhanu Alemayehu & Kenneth Warner, *The Lifetime Distribution of Health Care Costs*, 39 HEALTH SERV. RSCH. 627 (2004); Shannon Weeks McCormack, *Postpartum Taxation and the Squeezed Out Mom*, 105 GEO. L.J. 1323, 1333 (2017); and Ifeoma Ajunwa, *Genetic Data and Civil Rights*, 51 HARV. C.R.-C.L. L. REV. 75 (2016).

training data. An ML system used to evaluate candidates or employees cannot ever know how they will perform in the future, so it relies on proxies. Unfortunately, "proxies are bound to be inexact and often unfair."[46] Training data may be inaccurate, non-representative, or fundamentally biased due to social phenomenon.[47] If ML systems use historical data as inputs and models for future decisions, they likely will be biased by previous human decision-makers' propensity for racial and gender-based stereotypes and hiring decisions.[48] Studies have shown that even names which served as a proxy for the applicants' race can significantly affect callbacks, with white-sounding names receiving fifty percent more callbacks than black-sounding ones,[49] and such subtle human biases may be replicated by ML systems.

The inherent problem here may not be with the algorithms or ADMs themselves, but "in broader social processes."[50] Professor Sandra Mayson summarized the phenomenon as follows:

> All prediction functions like a mirror. . . . Algorithmic prediction produces a precise reflection of digital data. Subjective

[46] O'NEIL, *supra* note 1, at 108.

[47] Ajunwa, *supra* note 3, at 637. For example, machine learning systems trained for facial recognition have been often cited as performing poorly at recognizing the faces and emotions of people with darker skin. *See* Barocas & Selbst, *supra* note 1, at 680–81 ("[B]iased training data leads to discriminatory models."). Barocas and Selbst explore numerous potential problems with training data, including errors in data labeling, collection of incorrect, partial, or nonrepresentative data, bias in feature selection, proxies, and masking (intentional discrimination). *Id.* at 680–93.

[48] Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias*, PROPUBLICA (May 23, 2016) (examining the impact of race on recidivism algorithms used for criminal sentencing),

propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[49] O'NEIL, *supra* note 1, at 112. *See also* Latanya Sweeney, *Discrimination in Online Ad Delivery*, 56 COMM. ACM 44, 47 (MAY 2013) (discussing study findings that Google queries for black-sounding names were more likely to deliver advertisements for arrest records than searches for white-sounding names),

https://cacm.acm.org/magazines/2013/5/163753-discrimination-in-online-ad-delivery/fulltext.

[50] Pauline T. Kim, *Auditing Algorithms for Discrimination*, 166 U. PA. L. REV. ONLINE 189, 191 (2017). Algorithmic systems may derive correlations that predict both job-related goals, but which also have a disparate impact. James Grimmelmann & Daniel Westreich, *Incomprehensible Discrimination*, 7 CAL. L. REV. CIR. 164 (2016).

> prediction produces a cloudy reflection of anecdotal data. But the nature of the analysis is the same. To predict the future under status quo conditions is simply to project history forward.[51]

Machine learning systems tasked with mining through vast quantities of data may also make correlations with applicant or employee attributes or actions that do not have any clear connection to job performance.[52] In one instance, a data mining algorithm discovered a high correlation between computer programmers' abilities and their visits to a particular Japanese manga website.[53] In other cases, these correlations have proven more overtly problematic. When it set up an algorithm to search for information associated with higher employer turnover, Xerox found several surprising categories. One correlation was that job applicants who had longer commutes tended to have higher turnover, but this also had a socioeconomic correlation. Those applicants with longer commutes were coming from poor neighborhoods.[54] Allowing an algorithm to make an employment decision based on this information would have been a classic case of redlining.[55]

Bias may also arise because of fundamental problems in the configuration or setup of ADM systems. For example, an applicant intake system may be purposefully or inadvertently coded to require that applicants input a recent graduation date, thereby excluding older workers.[56] The initial problem that an employer wishes to have an AI examine may also be ambiguous and unsuitable to reduction to computer code. If an ADM is tasked with finding applicants who will perform well in a given position, how will performance be defined? In

[51] Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2224 (2019) (arguing that a fundamental problem with algorithmic systems lies in the nature of prediction itself).

[52] Kim, *supra* note 6, at 874.

[53] King & Mrkonich, *supra* note 7, at 560; Don Peck, *They're Watching You at Work*, ATLANTIC MONTHLY (Dec. 2013), https://www.theatlantic.com/magazine/archive/2013/12/theyre-watching-you-at-work/354681/.

[54] O'NEIL, *supra* note 1, at 118–19; MARCUS & DAVIS, *supra* note 27, at 36–37.

[55] Kim, *supra* note 6, at 863.

[56] Ajunwa, *supra* note 3, at 622–23 (noting that when these employees are deterred from even completing an application, such systems may "discreetly and disproportionately cull the applications of job seekers who are from legally protected classes").

setting the target variables for the system, the data miner may unintentionally program it in ways that disadvantage people based on a protected characteristic under the law.[57] This may stem from the intentional or unintentional bias of programmers, misinterpretations of system goals, or from programmers' preference for binary questions which can more easily be translated into code.[58] Lastly, if an ADM is set up to provide scores or feedback for human reviews, users may trust and therefore defer to the algorithm in a process called "automation bias."[59]

3. Technical Problems

Even assuming that a model is free of bias, however, other issues arise. Machine learning models are limited to learning from successive iterations of data to which they have access. However, when rejecting an employee, the models currently have no way of knowing if the decision was a false negative, or improper rejection, because they receive no feedback about that employee's future career trajectory or success.[60] Companies likely will do little to update their ADM systems as they strive to efficiently manage large applicant pools, and the systems may continue to grow outdated and reject more and more qualified candidates.[61]

[57] Barocas & Selbst, *supra* note 1, at 677–79; Bornstein, *supra* note 23, at 562–64; Citron, *supra* note 1, at 1261, 1267–71 ("Although all translations shade meaning, the translation of policy from human language into code is more likely to result in a significant alteration of meaning than would the translation of policy from English into another human language.").

[58] Citron, *supra* note 1, at 1261–62; Kim, *supra* note 50, at 192–93 ("Designing a system to be accountable for a substantive goal like nondiscrimination is difficult because it requires specifying the policy goals in terms precise enough to be reduced to code. What constitutes forbidden discrimination is highly contested in the legal and political spheres, and these debates pose a problem for computer programmers.").

[59] Ajunwa, *supra* note 3, at 636; Citron, *supra* note 1, at 1261, 1271 ("The cognitive system's engineering literature has found that human beings view automated systems as error-resistant. Operations of automated systems tend to trust a computer's answers. As a result, operators of government decision systems are less likely to search for information that would contradict a computer-generated solution. Studies show that human beings rely on automated decisions even when they suspect system malfunction . . . . Automation bias effectively turns a computer program's suggested answer into a trusted final decision.").

[60] O'NEIL, *supra* note 1, at 111; Kim, *supra* note 6, at 881–82.

[61] O'NEIL, *supra* note 1, at 111.

Machine learning systems and their programs also constantly struggle with the problem of overfitting. Overfitting occurs when a machine learning algorithm learns all of the particular examples from a given set of training data, but does not understand the general pattern that underlies them, or misunderstands these patterns by focusing on some other correlated factors in the data.[62] In one study, an image recognition program was tasked with differentiating pictures of Siberian Huskies and wolves.[63] Instead of focusing on any inherent characteristics of the Huskies or wolves, however, the machine learning system learned to identify wolves because researchers always presented them in photographs with snowy backgrounds.[64]

Another class of technical issues which has significant implications for discrimination and fairness is referred to as the "black box problem."[65] Machine learning and deep learning systems are often "opaque."[66] Their data, correlations, and functions make little sense to humans, and even experts struggle to understand and explain why particular systems make the decisions that they do.[67] Jerry Kaplan describes the problem as follows:

> In most cases, it's impossible for the creators of machine learning programs to peer into

[62] NICK POLSON & JAMES SCOTT, AIQ: HOW PEOPLE AND MACHINES ARE SMARTER TOGETHER 67 (2018) (stating that overfitting "happens when a model just memorizes the random noise in the training data rather than learns the underlying pattern. An overfit model may describe the past with perfect accuracy, yet still be bad at predicting the future").

[63] Marco T. Ribeiro, Samir Singh & Carlos Guestrin, *"Why Should I Trust You?": Explaining the Predictions of Any Classifier*, KDD '16: PROC. 22ND ACM SIGKDD INT'L CONF. ON KNOWLEDGE DISCOVERY & DATA MINING 1135, 1142–43 (2016), https://arxiv.org/pdf/1602.04938.pdf.

[64] *Id.*

[65] Kim, *supra* note 6, at 921–22.

[66] Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, 1 BIG DATA & SOC'Y., 1, 1 (2016) ("[R]arely does one have any concrete sense of how or why a particular classification has been arrived at from inputs.")

[67] MARCUS & DAVIS, *supra* note 27, at 66 ("The problem is particularly acute since neural networks can't give human-style explanations for their answers, correct or otherwise. Instead neural networks are 'black boxes'; they do what they do, and it is hard to understand what's inside."). Without legislative or regulatory intervention, software developers have little incentive to develop better frameworks for explainability. FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 2–10 (2015) (exploring the legal and economic background leading to the widespread intentional obfuscation of algorithmic systems).

> their intricate, evolving structure to understand or explain what they know or how they solve a problem, any more than I can look into your brain to understand what you are thinking about. These programs are no better able to articulate what they do and how they do it than human experts - they just know the answer. They are best understood as developing their own intuitions and acting on instinct: a far cry from the old canard that they "can only do what they are programmed to do.[68]

Given this, it is often difficult for human developers or researchers to explain why such systems may be right sometimes and wrong other times. This problem is compounded by the fact that the systems cannot issue human-style explanations themselves.[69] As a result, some scholars argue that AI poses unique problems for antidiscrimination, as it may replicate historical patterns of prejudice or reflect preexisting biases, but it does so in a manner that is difficult to detect or explain to juries and courts.[70]

[68] JERRY KAPLAN, HUMANS NEED NOT APPLY: A GUIDE TO WEALTH AND WORK IN THE AGE OF ARTIFICIAL INTELLIGENCE 30 (2015).

[69] MARCUS & DAVIS, *supra* note 27, at 66. "The problem of explainability may be exacerbated when the machine learning system's decision parameters change over time if audit records are not kept regarding individual decisions." Merle Temme, *Algorithms and Transparency in View of the New General Data Protection Regulation*, 3 EUR. DATA PROT. L. REV. 473, 479 (2017).

[70] Barocas & Selbst, *supra* note 1, at 672, 677 ("By definition, data mining is *always* a form of statistical (and therefore seemingly rational) discrimination. Indeed, the very point of data mining is to provide a rational basis upon which to distinguish between individuals and to reliably confer to the individual the qualities possessed by those who seem statistically similar. Nevertheless, data mining holds the potential to unduly discount members of legally protected classes and to place them at systematic relative disadvantage."). Taken from another point of view, however, this problem is not unique to AI or machine learning programs. Mitchell Kapor succinctly stated of the similar problem of humans, "Human intelligence is a marvelous, subtle, and poorly understood phenomenon. There is no danger of duplicating it anytime soon." Kurt Anderson, *Enthusiasts and Skeptics Debate Artificial Intelligence*, VANITY FAIR (Nov. 26, 2014),

https://www.vanityfair.com/news/tech/2014/11/artificial-intelligence-singularity-theory. Indeed, it is fundamentally difficult to program artificial intelligence to think like humans partly because there is so much that we do not understand of our own minds. Many hope for a so-called magic bullet to

Lastly, machine learning systems can be deliberately manipulated and fooled by programmers who wish to influence them and falsify their outputs. For instance, researchers from the University of Wisconsin were able to use genetic algorithms to "evolve" images that appear to be no more than static to a human observer, but which neural networks would confidently classify as recognizable images of everyday objects.[71] Others have taken recognizable pictures and made small but specific changes to the image pixels that caused such systems to make significant errors, such as mistaking Shih Tzu puppies and school buses for ostriches.[72]

4. Implications for the Data Privacy Rights of Applicants and Employees.

In addition to issues of bias and technical errors, ADM systems raise various privacy concerns depending on the underlying data used during training and deployment.[73] By their very nature, machine learning systems and deep neural networks process large amounts of training data during their development, and they continue to use large sets of data once deployed in order to function. This may be particularly concerning in the employment context, because employees have typically been permitted a lower reasonable expectation of privacy at work than they have in public or at home.[74] ADM systems

unlock artificial intelligence, but Marvin Minsky understood that there would be no such easy route forward. "What magical trick makes us intelligence? The trick is that there is no trick. The power of intelligence stems from our vast diversity, not from any single, perfect principle." MARVIN MINSKY, THE SOCIETY OF MIND 308 (1986).

[71] Anh Nguyen, Jason Yosinski & Jeff Clune, *Deep Neural Networks Are Easily Fooled: High Confidence Predictions for Unrecognizable Images*, PROC. IEEE CONF. ON COMPUT. VISION & PATTERN RECOGNITION, 427, 434 (2015), https://arxiv.org/abs/1412.1897.

[72] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow & Rob Fergus, *Intriguing Properties of Neural Networks*, PROC. INT'L CONF. ON LEARNING REPRESENTATIONS (2014), https://arxiv.org/abs/1312.6199.

[73] The right to privacy was first defined as "the right to be let alone." Charles S. Warren & Louis Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 193–205 (1890). Conceptions of the right to privacy steadily grew to incorporate four distinct areas: (1) informational privacy, (2) bodily privacy, (3) territorial privacy, and (4) communications privacy. PETER SWIRE & DEBRAE KENNEDY-MAYO, U.S. PRIVATE-SECTOR PRIVACY 13–14 (3d ed. 2020).

[74] Ifeoma Ajunwa, *Algorithms at Work: Productivity Monitoring Applications and Wearable Technology as the New Data-Centric Research Agenda for Employment and Labor Law*, 63 ST. LOUIS U. L.J. 21, 31 (2018);

tasked with monitoring employees' actions on their computers may theoretically engage in a wide variety of tracking, including counting employees' keystrokes, using video to monitor their movements, scanning for employees' use of personal email,[75] or even monitoring periods of inactivity which may reveal how often employees use the restroom, check their insulin, or engage in protected activity such as discussions of work activities or union organizing.[76]

The proliferation of ADM systems for workplace usage has also followed the expansion of data collected from three broad sources: digital information, sensors, and employee self-tracking through devices such as fitness trackers or smart phone apps.[77] The proliferation of wearable devices reveals location data that may be particularly sensitive,[78] but which has high utility for employers who wish to monitor employee productivity, interactions, break patterns, and safety.[79] For example, Amazon has patented technology that can be used to track employees' locations within their distribution center. In conjunction with the company's inventory tracking system, this technology could also direct those employees in real-time to the location of an item through haptic feedback, without the need for an employee to scan a computer screen.[80] Other systems have been designed to monitor employees' physical locations in relation to hazardous areas, and to alert their supervisors when the employees may stray out of a geofenced area within their worksite.[81] These systems help employees by physically protecting them, while also strengthening employers' overall safety plans and reducing workers' compensation claims.[82] Yet, any system for monitoring an employee's precise physical location also carries with it the potential for abuse. An

O'Connor v. Ortega, 480 U.S. 709, 717 (1987) ("Public employees' expectations of privacy in their offices, desks, and file cabinets, like similar expectations of employees in the private sector, may be reduced by virtue of actual office practices and procedures, or by legitimate regulation . . . An office is seldom a private enclave free from entry by supervisors, other employees, and business and personal invitees.").

[75] Corbyn, *supra* note 9.

[76] Protected activities under the NLRA can be found at National Labor Relations Act, 29 U.S.C. § 157.

[77] Adams-Prassl, *supra* note 4, at 134–35.

[78] The California Privacy Rights Act, for example, defines precise geolocation data as particularly sensitive personal information. CPRA § 1798.140(ae)(1)(C).

[79] Ajunwa, *supra* note 74, at 25–27, 46-47.

[80] *Id.* at 34–35.

[81] *Id.* at 39.

[82] *Id.* at 46–48.

employer may use a physical device on an employee's or an employer-owned automobile to track their driving for safety purposes or productivity during the work day, but such a device may also be used during non-working hours to track an employee's trips to the doctor, to their homes, or to other locations that reveal personal or protected information. They can also, in the hands of the wrong manager, be used for stalking and harassment.[83]

C. *Current Uses for ADMs and Artificial Intelligence in Employment*

Perhaps the first and largest current impact of ADM systems in employment is also the first step at which most individuals come into contact with a potential employer: at the time of submission of a resume. ADM systems often are built into applicant tracking systems (ATS) and have come to act as gatekeepers for employers both large and small. The justification for these systems is typically their efficiency.[84] As job posting information has spread online, the ease at which applicants can find openings has multiplied. Job mobility has also increased, and as a result, the number of applicants for an average job opening has skyrocketed. While the pool of available talent has increased, reviewing and scoring resumes consumes a significant amount of recruiters' time, costing employers money and slowing down the hiring process. Employers have turned to ADM systems to manage the increased flow of new applicants.[85] ATS allow recruiters to match keywords and filter applicants' resumes,[86] but more advanced

[83] *Id.* at 25–26.

[84] Bornstein, *supra* note 23, at 530–31. Automated systems can also mine information online from public sources about candidates. For example, systems can look at whether computer programmers have contributed to open-source projects and recommend them for jobs accordingly. Peck, *supra* note 53; Matt Richtel, *How Big Data Is Playing Recruiter for Specialized Workers*, N.Y. TIMES (Apr. 27, 2013), https://www.nytimes.com/2013/04/28/technology/how-big-data-is-playing-recruiter-for-specialized-workers.html.

[85] Studies have shown that nearly all of the Fortune 500 companies use applicant tracking systems in order to manage applications and resumes. Linda Qu, *99% of Fortune 500 Companies use Applicant Tracking Systems*, JOBSCAN (Nov. 7, 2019), https://www.jobscan.co/blog/99-percent-fortune-500-ats/.

[86] Gray Beltran, *The Pandemic Changed Everything About Work, Except the Humble Résumé*, N.Y. TIMES (Jan. 22, 2022), https://www.nytimes.com/2022/01/22/business/pandemic-work-resumes.html.

features of some ATS compare resumes to job descriptions and provide automated rankings.[87]

> As you might expect, human resources departments rely on automatic systems to winnow down piles of résumés. In fact, some 72 percent of résumés are never seen by human eyes. Computer programs flip through them, pulling out the skills and experiences that the employer is looking for. Then they score each résumé as a match for the job opening. It's up to the people in the human resources department to decide where the cutoff is, but the more candidates they can eliminate with this first screening, the fewer human-hours they'll have to spend processing the top matches.[88]

For example, at Unilever, the Human Resources Department set a goal of increasing diversity by focusing on entry-level talent and hiring. Existing processes were not able to evaluate recruits in sufficient numbers and give applicants individual attention. So, Unilever adapted its processes to include two rounds of tests and interviews that deployed AI. In the first round, candidates were asked to play online games that assessed traits such as risk aversion. These helped to determine which individuals may be well-suited for particular positions. In round two, applicants were asked to submit videos with answers to questions related to their specific job opening. Responses were analyzed by AI, including applicants' body language and tone. Finally, in the third round, in-person interviews were conducted at Unilever. The process change resulted in doubling applicants to 30,000 in a year, increasing applications from universities from 840 to 2,600 in that year, and increasing the socioeconomic diversity of new hires. The process also proved efficient for Unilever's recruiting department. The average time from application to hiring decision dropped from 4 months to 4 weeks, and the time recruiters spent reviewing applications dropped by seventy-five percent.[89]

[87] Qu, *supra* note 85.

[88] O'NEIL, *supra* note 1, at 113–14.

[89] H. James Wilson & Paul R. Daugherty, *Collaborative Intelligence: Humans and AI Are Joining Forces*, HARV. BUS. REV. (July–Aug. 2018), https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces.

Employers can also be motivated by the belief that automated systems are less biased than human reviewers.[90] Many software providers tout their potential benefits for diversity and inclusion. HireVue, for instance, advertises that "Structured interviews ensure consistency" and "AI-Driven insights [are] proven to reduce bias."[91] HireEZ goes a step further and advertises the ability of its software to proactively find diverse candidates:

> Identify More Underrepresented Talent. Search for talent with hireEZ's Diversity Sourcing to focus on minority groups for your open roles . . . hireEZ analyzes profiles for pronouns, schools, memberships with diversity organizations, and more to support you in building more inclusive talent pipelines.[92]

Aside from the claims to benefit employers, there may be benefits to such systems for applicants. Technology that automates search processes may be able to discover qualified individuals and match them with jobs from outside of their geographic region, or outside of the typical types of jobs that they may have searched.[93]

Another increasingly popular use for artificial intelligence in recruiting is video interview software. Such software can take two forms. It can provide a channel for a live interview between two people, or it can be used for an asynchronous interview in which an applicant's interview is conducted and recorded, often using structured interview questions. Many of these video interview software systems have begun to incorporate machine learning components used to automatically score applicants. "These capabilities are rapidly developing in ways that help interviewers to analyze facial expressions, nonverbal behavior, and voice intonation for signs of how well a job

[90] Ifeoma Ajunwa, *Beware of Automated Hiring*, N.Y. TIMES (Oct. 8, 2019),
https://www.nytimes.com/2019/10/08/opinion/ai-hiring-discrimination.html; Kim, *supra* note 6, at 869–71.

[91] *Increase diversity and mitigate bias*, HIREVUE, https://www.hirevue.com/employment-diversity-bias (last visited Aug. 10, 2022).

[92] *Your Diversity Recruiting Just Got Easier*, HIREEZ, https://hireez.com/solutions/diversity-inclusion/ (last visited Aug. 10, 2022).

[93] Adams-Prassl, *supra* note 4, at 129.

seeker will fit in with a company's needs and culture."[94]

Artificial intelligence may be incorporated into scheduling software to regularly arrange work schedules or to slot part-time workers into new or open shifts depending on changes in business demand or employee absences. Current scheduling software can be programmed with a variety of rules, including legal compliance, and arrange workers in overlapping shifts so that they all have adequate time to take allotted meal breaks and avoid overtime.[95] There are potentially significant benefits in this area, as ADM systems can theoretically be used to discover patterns related to employee management and staffing that may in some cases save money, and in other cases save lives.[96] For example, future scheduling software systems may use information such as average patient data, upcoming scheduling procedures, previous work schedules, absences, and late arrivals for employees, and it may notice that a particular nurse is consistently fifteen minutes late on the third Thursday of every month. Due to a critical surgery scheduled for the upcoming Thursday, an ADM system could refrain from scheduling that nurse for the shift in question.[97] In doing so, the ADM system may benefit the hospital,

[94] *Video Interviewing*, ICIMS, https://www.icims.com/glossary/video-interviewing/ (last visited Aug. 10, 2022).

[95] *See, e.g.*, *Employee Scheduling Software for Businesses*, ADP, https://www.adp.com/resources/articles-and-insights/articles/e/employee-scheduling.aspx (last visited Aug. 10, 2022); *Restaurant Scheduling Software*, 7SHIFTS, https://www.7shifts.com/restaurant-employee-scheduling-software (last visited Aug. 10, 2022).

[96] Some criticize algorithmic scheduling for the potential negative effect on employees' work-life balance. *See, e.g.*, O'NEIL, *supra* note 1, at 123–40; Jodi Kantor, *Working Anything But 9 to 5*, N.Y. TIMES (Aug. 13, 2015), https://www.nytimes.com/interactive/2014/08/13/us/starbucks-workers-scheduling-hours.html.

[97] The decision could be construed as an adverse employment action, especially if carried out more than a single time, and it raises novel issues with respect to data privacy, fairness, and potential discrimination. Perhaps, for example, the nurse has a standing doctor's appointment of their own, and the scheduling deviation has been discussed with their manager, who granted an accommodation. Currently, however, there are no laws in the United States that would require the employer to give the nurse notice of the use of algorithms to determine their schedule, and there have been no significant cases interpreting the use of AI under Title VII or applicable antidiscrimination laws. This article argues that the best approach will likely be to provide transparency to the nurse in question, informing him or her of the ADM system's involvement. If the nurse believes the decision was made unfairly, he or she can raise an internal complaint, and if that fails, they can

doctor, patient, and even the nurse, who is provided a more appropriate work schedule.

On the other hand, scheduling software may be used in ways detrimental to employees, such as to ensure that they do not reach an adequate number of hours to gain benefits. Scheduling algorithms also may not be programmed to prioritize predictability, and this gives rise to a number of problems. Unpredictable schedules may cause issues with employees' ability to engage in consistent attendance at school or college, care for children or family members, or make necessary health appointments. Low-wage workers who need to work multiple jobs may find juggling inconsistent schedules to be impossible and may not be able to secure the work they need. Finally, inconsistent schedules created by software may take a toll on the families of workers, imposing chaotic lives on children just as they do so on parents, and severely harming childhood development.[98]

Artificial intelligence and ADMs may augment and eventually replace many managerial duties.[99] An infamous example of a company that widely deploys productivity tracking and algorithms to manage its workforce is Amazon.[100] Because of this, California became the first state to place limitations on Amazon's practices of using employee quotas at its warehouses in 2021. AB 701 requires that:

> Each employer shall provide to each employee, upon hire, or within 30 days of the effective date of this part, a written description of each quota to which the employee is subject, including the quantified number of tasks to be performed or materials to be produced or handled, within the defined time period, and any potential adverse employment

pursue their rights under applicable antidiscrimination law. The review of the ADM's decision can be placed in the hands of humans for a more thorough review, but the initial decision not to schedule the nurse likely is one that is suitable and appropriate for the ADM system.

[98] O'NEIL, *supra* note 1, at 128–30.

[99] Adams-Prassl, *supra* note 4, at 124 ("Instead of taking away workers' jobs, I suggest, advances in AI-driven decision-making will first and foremost change their managers' daily routines, augmenting and eventually replacing human day-to-day control over the workplace: we are witnessing the rise of the "algorithmic boss.").

[100] Anabelle Williams, *5 ways Amazon monitors its employees, from AI cameras to hiring a spy agency*, INSIDER (Apr. 5, 2021),

https://www.businessinsider.com/how-amazon-monitors-employees-ai-cameras-union-surveillance-spy-agency-2021-4.

> action that could result from failure to meet
> the quota.[101]

The law further states that employees should not be required to meet quotas that would prevent their ability to take meal breaks, rest breaks, or to use the bathroom.[102] Although modest in both scope and its probable effect (California already requires employees be "provided" meal and rest breaks[103]), AB 701 is indicative of the growing concerns of members of the public and legislators that algorithms are being deployed at an increasing rate to manage employees, often in manners that are opaque and hidden from the employees themselves.[104]

In one of the first cases to weigh the propriety of machine learning algorithms, the Houston Independent School District was sued by a group of teachers as the result of using an algorithm that incorporated test data from teachers' classes.[105] In *Hous. Fed'n of Teachers*, public school teachers and the teachers' union sued under the Fourteenth Amendment's due process and equal protection clauses seeking to enjoin the use of an algorithmic scoring system that rated their performance based on student test scores.[106] The school district utilized a third-party software to calculate student progress on test scores using a value-added model called the Educational Value-Added

[101] CAL. LAB. CODE § 2101 (West 2022).

[102] *Id.* at § 2102.

[103] *See Brinker Rest. Corp. v. Superior Court*, 53 Cal. 4th 1004, 1040 (2012) (holding an employer fulfills its duty to provide a meal period where it "relieves its employees of all duty, relinquishes control over their activities, permits them a reasonable opportunity to take an uninterrupted 30-minute break," and "does not impede or discourage them from doing so").

[104] Algorithmic management is not limited to employees., however. Platform services participating in the gig economy, such as Uber, deploy systems to gather information about their independent contractors' performance, including customer ratings and productivity. Adams-Prassl, *supra* note 4, at 131; Alex Rosenblat & Luke Stark, *Algorithmic Labor and Information Asymmetries: A Case Study of Uber's Drivers*, 10 INT'L J. COMMC'N 3758, 3761–66 (2016). A United States District Court examining Uber's practices determined that the level of monitoring arguably gave Uber a "tremendous amount of control over the 'manner and means' of its drivers' performance" for purposes of determining whether or not they were properly classified as independent contractors. *O'Connor v. Uber Techs., Inc.*, 82 F. Supp. 3d 1133, 1151 (N.D. Cal. 2015).

[105] *Hous. Fed'n of Tchrs., Loc. 2415 v. Hous. Indep. Sch. Dist.*, 251 F. Supp. 3d 1168 (S.D. Tex. 2017).

[106] *Id.* at 1171.

Assessment System (EVAAS).[107] EVAAS measured teacher effectiveness by tracking teachers' impact on student test scores over time, based on comparing average test scores of the teachers' students versus statewide test scores for students in the same grade and course.[108] Teachers were then assigned a rating (well above, above, no detectable difference, below, or well below), and shortly after adoption of the software, the school district began a policy of terminating teachers based on their algorithmically-generated EVAAS score.[109]

While granting the school district its motion for summary judgment on plaintiffs' claims for substantive due process and equal protection, the district court allowed a procedural due process violation claim to proceed.[110] Pointing to Supreme Court precedent, the court wrote, "The core requirement of procedural due process is the opportunity to be heard at a meaningful time and in a meaningful manner."[111] However, the teachers had no such opportunity. The software provider claimed trade secret protection over the algorithms and the software used to calculate scores and denied the school district access to such programs.[112] The school district consequently could not provide teachers access (and did not have access itself) to the computer algorithms and data necessary to verify the accuracy of the scores, and it took no steps to verify or audit the scores.[113] The court noted that scores could be "erroneously calculated for any number of reasons, ranging from data-entry mistakes to glitches in the computer code itself."[114] This system ran counter to due process requirements, as described by the Supreme Court:

[107] *Id.* at 1172.

[108] *Id.*

[109] *Id.* at 1172–74.

[110] *Id.* at 1180. With respect to plaintiffs' claims for violation of substantive due process and the equal protection clause, the court deployed rational basis scrutiny. Plaintiffs alleged that EVAAS failed rational basis scrutiny "because it is sytematically [sic] biased against large categories of teachers on the basis of the type and size of classrooms they teach, is highly volatile, is highly variable on the basis of which models or tests are used, and is highly divergent from other measures of teacher effectiveness." *Id.* The court disagreed, holding that the "constitutional status of rationality allows governments to use blunt tools which may produce only marginal results." *Id.* at 1182.

[111] *Hous. Fed'n of Tchrs., Loc. 2415.*, 251 F. Supp. 3d at 1175 (citing *Mathews v. Eldridge*, 424 U.S. 319, 333 (1976)).

[112] *Id.* at 1177.

[113] *Id.* at 1176–77.

[114] *Id.* at 1177.

> The purpose of [the due process] requirement is not only to ensure abstract fair play to the individual. Its purpose, more particularly, is to protect his use and possession of property from arbitrary encroachment—to minimize substantively unfair or mistaken deprivations of property . . . . For when a person has an opportunity to speak up in his own defense, and when the State must listen to what he has to say, substantively unfair and simply mistaken deprivations of property interests can be prevented.[115]

The court ruled that such protections were impossible for the teachers because of the plain lack of transparency. Without access to the information involved in the decision-making process, "EVAAS scores will remain a mysterious 'black box,' impervious to challenge."[116]

III. Legal Frameworks for ADMs and Artificial Intelligence in Employment

A. *Current Regulations in the United States*

1. Equal Employment Opportunity (EEO) Regulations

As discussed previously, one of the most significant concerns for ADMs in the workplace is the potential for bias and discrimination. The United States, however, already has strong antidiscrimination laws that provide redress when hiring or termination decisions are made based on an applicant or employees protected characteristics. Title VII of the Civil Rights Act makes it unlawful to discriminate against people because of their protected class in the terms and conditions of employment, which includes hiring.[117] The law "proscribes not only

[115] *Fuentes v. Shevin*, 407 U.S. 67, 80–81 (1972).

[116] *Hous. Fed'n of Tchrs., Loc. 2415*, 251 F. Supp. 3d at 1179.

[117] *See* Civil Rights Act of 1964, 42 U.S.C. § 2000e-2(a) ("It shall be an unlawful employment practice for an employer – to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual's race, color, religion, sex, or national origin."). Title VII has also been interpreted to prohibit discrimination based on sexual orientation or transgender status. *Bostock v. Clayton Cnty., Georgia*, 140 S. Ct. 1731, 1742–43 (2020).

overt discrimination but also practices that are fair in form, but discriminatory in operation."[118] Other federal laws guard against discrimination based on age, pregnancy, disability, protected military status, and genetic information.[119] The use of policies or practices that may not overtly target a protected class, but nonetheless have an impact on hiring based on a protected characteristic, is known as "disparate impact discrimination." Discrimination can therefore be caused by facially neutral policies and does not require a showing of bad intent.[120]

> [G]ood intent or absence of discriminatory intent does not redeem employment procedures or testing mechanisms that operate as "built-in headwinds" for minority groups and are unrelated to measuring job capability.[121]

[118] *Griggs v. Duke Power Co.*, 401 U.S. 424, 431 (1971). The substance of disparate impact discrimination standards and procedures depend on whether we approach discrimination with the goal of anticlassification (formal equality) or antisubordination (substantive equality). *Bornstein*, *supra* note 23, at 525. Although this may ultimately impact how such algorithms are evaluated under existing laws, it is not within the scope of this article.

[119] *See* Age Discrimination in Employment Act of 1967, 29 U.S.C. §§ 621–634; Pregnancy Discrimination Act of 1978, Pub. L. No. 95-555, 92 Stat. 2076 (amending Title VII to include discrimination on the basis of pregnancy); Americans with Disabilities Act of 1990, 42 U.S.C. §§ 12101–12213; Uniformed Services Employment and Reemployment Rights Act of 1994, 38 U.S.C. §§ 4301–4335; Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233, 122 Stat. 881.

[120] *Raytheon Co. v. Hernandez*, 540 U.S. 44, 52–53 (2003) ("[D]isparate-impact claims 'involve employment practices that are facially neutral in their treatment of different groups but that in fact fall more harshly on one group than another and cannot be justified by business necessity.' . . . '[A] facially neutral employment practice may be deemed illegally discriminatory without evidence of the employer's subjective intent to discriminate that is required in a 'disparate-treatment' case.").

[121] *Griggs*, 401 U.S. at 432. Disparate impact rules can even adversely impact employers who attempt to proactively implement policies to improve workforce diversity. The Supreme Court has held that employers who disregard the result of a valid job selection process, such as performance-based job tests for hiring, because the tests did not yield a racially diverse group of candidates may be found to have intentionally discriminated against the initially successful candidates based on their race, "absent a strong basis in evidence that the test was deficient and that discarding the results is necessary to avoid violating the disparate-impact provision." *Ricci v. DeStefano*, 557 U.S. 557, 583–84 (2009).

The burden of proof for a disparate impact claim is initially on a plaintiff to show that a particular employment practice caused the exclusion of applicants for jobs or promotions because of their protected characteristic.[122] Plaintiffs typically demonstrate this through mathematical modeling of the impact of the practice or policy on applicants or employees using either the "four-fifths rule" or a statistical significance test.[123] "Briefly stated, under the four-fifths rule, a disparity is actionable when one group's pass rate is less than four-fifths (eighty percent) of another group's pass rate."[124] Statistical significance tests, on the other hand, attempt to distinguish whether deviations in the data are due to random chance or caused by a specific policy through a number of different mathematical models. "Researchers most commonly use the ninety-five percent confidence level, which is also termed the five percent (0.05) level of significance . . . . At the ninety-five percent level, we can be ninety-five percent certain that the observed disparity in the applicant pool reflects a real disparity in the relevant labor market with respect to the challenged practice. There is still, however, a one in twenty possibility that there is no disparity in the overall population."[125]

Following the plaintiff's initial showing, the defendant in a disparate impact case can then attempt to rebut the plaintiff's statistics, or it can show that the requirement has a relationship to employment. In other words, they must show the policy or practice is "job related"

[122] *See Watson v. Fort Worth Bank & Trust*, 487 U.S. 977, 994 (1988).

[123] "To establish a prima facie case of disparate impact, plaintiffs must show that a particular employment practice caused an adverse impact on the basis of a protected status, such as race. Plaintiffs generally prove such causation by comparing the selection rates of majority and minority applicants for a position and then showing that the disparity is statistically significant or that it violates the four-fifths rule. The Supreme Court has rejected a 'rigid mathematical formula' for disparate impact, providing instead the ambiguous guidance to lower courts that 'statistical disparities must be sufficiently substantial that they raise . . . an inference of causation.'" Jennifer L. Peresie, *Toward a Coherent Test for Disparate Impact Discrimination*, 84 IND. L.J. 773, 777–78 (2009).

[124] *Id.* at 774.

[125] *Id.* at 785–86. Each of these tests has significant shortcomings. Statistical significance tests are highly sensitive to sample size and tend to insulate smaller employers from legal claims. *Id.* at 787. The four-fifths rule tends to disproportionately burden smaller employers, as well as imply that there is an acceptable level (twenty percent) of discrimination which the law will condone. McKenzie Raub, *Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices*, 71 ARK. L. REV. 529, 546–47 (2018).

and "consistent with business necessity."[126] The burden then shifts back to plaintiff to show that an alternative employment practice would have served the employer's legitimate interests without a similar discriminatory effect, and that the employer "refuses to adopt such alternative employment practice."[127]

Because of the nature of machine learning systems, as previously discussed, they often give rise to black box problems because they make data correlations, associating various inputs in ways that are not intelligible to programmers, developers, or auditors, and then they issue determinations and outputs based on these opaque processes.[128] This problem has been cited as particularly acute in the employment context because of the burden shifting analysis performed in disparate impact cases.[129] Where the correlations and reasoning for the automated system's decision are opaque, a defendant company may plausibly argue any data correlation to job performance establishes the "job-related" defense.[130] It may then be exceedingly difficult for the plaintiff to demonstrate that an alternative and less burdensome option existed.[131] To date, however, no court in the United States has published any significant opinion addressing these problems under the disparate impact theory.

[126] *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 406 (1975); *Watson*, 487 U.S. at 998; 42 U.S.C. § 2000e-2(k)(1)(A)(i).

[127] 42 U.S.C. § 2000e-2(k)(1)(A)(ii); *Watson*, 487 U.S. at 998.

[128] *See infra*, Section II.A.3.

[129] Grimmelmann & Westreich, *supra* note 50, at 177 ("*Incomprehensible discrimination will not stand*. Applicants who are judged and found wanting deserve a better explanation than, 'The computer said so.' Sometimes computers say so for the wrong reasons–and it is employers' duty to ensure that they do not.").

[130] Kim, *supra* note 6, at 920 ("If a statistical correlation were sufficient to satisfy the defense of job-relatedness, the standard would be a tautology rather than a meaningful legal test.").

[131] Many scholars have noted the potential problem with the burden shifting and the "job-related" defense, while others have proposed new interpretations of Title VII that would support disparate impact claims against AI systems. A full examination of this lies beyond the scope of this article. Barocas & Selbst, *supra* note 1, at 701–712; Kim, *supra* note 6 (arguing Title VII incorporates an anti-classification theory applicable to AI systems); Sullivan, *supra* note 44, at 398; Bornstein, *supra* note 23, at 527 (proposing an anti-stereotyping approach to algorithmic discrimination); Joshua A. Kroll, Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson & Harlan Yu, *Accountable Algorithms*, 165 U. PA. L. REV. 633, 637 (2017) (proposing technical solutions to the issues of discriminatory algorithms).

The dearth of court authority on the issue of AI and disparate impact, despite the ubiquity of such systems in recruitment and hiring, likely stems directly from their relative invisibility. Most employees and applicants simply lack the knowledge that they were ever subject to a flawed policy or practice to begin with. No federal laws in the United States require disclosure to applicants or employees when an ADM system is used as part of their hiring or termination. Two laws have recently been passed that would offer a degree of transparency in Illinois and New York City, but those are of limited application.[132] Furthermore, the data privacy laws being passed and contemplated by various state legislatures typically carve out employees from their coverage.[133] This is a marked difference with the European Union, which has extended transparency both with respect to employee data privacy under the GDPR, and which has proposed coverage of AI systems affecting employees under the E.U. AI Proposal introduced in April 2022.[134]

The issue of transparency is not unique to ADM systems with respect to employment decisions, and in particular hiring. Employees often are not told—and do not ask—about the reasons for an employer denying their job application. Even if an applicant does ask why they did not get the job, employers generally have no obligation to provide them such information. Without the barest of information, employees are unlikely to file suit. This leads to opaque hiring practices, and employers who may use this to their advantage. Take, for example, the use of personality tests as a prerequisite for hiring. These tests may superficially grade applicants for personality types such as extraversion, agreeableness, conscientiousness, neuroticism, and openness to ideas, but they also may have a disparate impact on individuals with histories of mental illness and create a significant burden for them to obtain even entry level jobs for which they are otherwise qualified.[135] These individuals, however, rarely learn that the personality test formed the basis for their rejection, and they are unlikely to seek or retain an attorney to challenge those decisions.[136]

[132] *See infra*, Section III.A.3.

[133] *See infra*, Section III.A.2.ii.

[134] GDPR art. 88; Proposal for a Regulation Of The European Parliament And Of The Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts, COM (2015) 452 final (Apr. 1, 2021) [hereinafter "E.U. AI Proposal"].

[135] O'NEIL, *supra* note 1, at 105–22.

[136] *Id.* at 105–06. *Compare id. with Griggs*, 401 U.S. at 436 (holding that intelligence tests are illegal unless an employer can demonstrate they are a reasonable measure of job performance).

Thus, we have strong antidiscrimination laws in the United States, but often applicants and employees have no understanding that they may have been wronged, and have no reason to pursue remedies under the existing laws. The ADM systems are operating as "dark systems," and there is the significant possibility these systems are perpetuating historical discrimination or simply operating pursuant to biased and unfair parameters.

2. Data Privacy Regulations in the United States

i. The Fair Credit Reporting Act ("FCRA")

One significant deviation from the general rule that employees are not entitled to transparency regarding the reasons for an adverse employment decision comes from the area of background checks. In 1970, Congress passed the first legislation in the United States governing the privacy of consumer information, the Fair Credit Reporting Act ("FCRA").[137] In so doing, Congress found that "[a]n elaborate mechanism has been developed for investigating and evaluating the credit worthiness, credit standing, credit capacity, character, and general reputation of consumers," and "[t]here is need to insure that consumer reporting agencies exercise their grave responsibilities with fairness, impartiality, and a respect for the consumer's right to privacy."[138] Congress therefore enacted the FCRA with the goal of establishing "reasonable procedures" that are "fair and equitable to the consumer, with regard to confidentiality, accuracy, relevancy, and proper utilization of such information."[139]

As a key component of this legislation, the government required that when employers perform background checks on employees or applicants, they must engage in an "adverse action process" whenever they uncover negative information that they intend to use as the basis for employment terminations or when declining an applicant for employment.[140] That adverse action process consists of

[137] The Fair Credit Reporting Act, Pub. L. No. 91-508, 84 Stat. 1114 (1970) (codified as amended at 15 U.S.C. § 1681); PRISCILLA M. REGAN, LEGISLATING PRIVACY: TECHNOLOGY, SOCIAL VALUES, AND PUBLIC POLICY 101 (1995) (noting the FCRA was the first information or data privacy legislation in the United States).

[138] 15 U.S.C. § 1681(a)(1)(4).

[139] *Id.* at § 1681(b).

[140] *Id.* at §§ 1681b(b)(3)(A), 1681m(a). The FCRA uses the term "consumer report," but this article will refer to such reports by the more commonplace term, "background check."

two steps. First, the employer must provide a pre-adverse action notice to the employee or applicant containing a copy of the background check and a description of that person's rights under the FCRA.[141] The employee or applicant then has the opportunity to provide evidence that the background check may contain errors. Although the FCRA does not specify how much time an employer should provide to an employee, the Federal Trade Commission has said, "[s]ome reasonable period of time must elapse, but the minimum length will vary depending on the particular circumstances involved."[142] After this reasonable time period has passed, the employer may take action based on the information contained in the background check, but it is required to provide a final adverse action notice containing information regarding the employee's credit score (if applicable), information regarding the credit reporting agency ("CRA") which provided the background check, a statement that the CRA did not make the adverse action decision, information regarding how an employee may request a copy of the report, and details regarding how the employee or applicant can contact the CRA to dispute the accuracy or completeness of the report.[143]

ii. State General Data Protection Laws

As discussed previously, several states have passed general data protection legislation, including California, Colorado, and Virginia, and each of these states grapples with issues related to automated decision-making. However, none of these laws adequately address the problem of automated decision-making with respect to its use in the employment context.

The first general data protection legislation passed in California, the California Consumer Privacy Act of 2018 ("CCPA") did not call for any specific requirements regarding automated

[141] *Id.* at § 1681b(b)(3)(A) ("Except as provided in subparagraph (B), in using a consumer report for employment purposes, before taking any adverse action based in whole or in part on the report, the person intending to take such adverse action shall provide to the consumer to whom the report relates - (i) a copy of the report; and (ii) a description in writing of the rights of the consumer under this subchapter, as described by the Bureau under Section 1681g(c)(3) of this title.")

[142] FED. TRADE COMM'N, 40 YEARS OF EXPERIENCE WITH THE FAIR CREDIT REPORTING ACT: AN FTC STAFF REPORT WITH SUMMARY OF INTERPRETATIONS 52 (2011).

[143] 15 U.S.C. § 1681m(a).

decision-making.[144] That changed, however, in 2020, when California voters opted in favor of the California Privacy Rights Act ("CPRA").[145] The CPRA updated and significantly expanded many aspects of the CCPA, and it added several provisions related to automated decision-making, including a requirement that the newly-formed California Privacy Protection Agency ("CPPA") release regulations on the issue.[146] The CPRA defines "profiling" as broadly including the "automated processing of personal information" which would include an individual's performance at work:

> (z) 'Profiling' means ***any form of automated processing of personal information***, . . . ***in particular to analyze or predict aspects concerning that natural person's performance at work***, economic situation, health, personal preferences, interests, reliability, behavior, location, or movements.[147]

Under Section 1798.185, the CPRA requires agency rulemaking on the issue of automated decision-making, and in particular, requires the regulations to address the extent to which businesses may engage in automated decision-making, and how they may respond to requests for information about the software's internal logic. Specifically, the CPPA is tasked to issue:

> regulations governing access and opt-out rights with respect to businesses' use of automated decisionmaking technology, including profiling and requiring businesses' response to access requests ***to include meaningful information about the logic involved in those decisionmaking processes***,

[144] California Consumer Privacy Act, CAL. CIV. CODE §§ 1798.100-199.100 (West 2022).

[145] California Privacy Rights Act, 2018 Cal. Stat. 1807 (to be codified at CAL. CIV. CODE §§ 1798.100-199.100 (effective Jan. 1, 2023)) [hereinafter "CPRA"]; Cameron F. Kerry & Caitlin Chin, *By passing Proposition 24, California voters up the ante on federal privacy law*, BROOKINGS (Nov. 17, 2020), https://www.brookings.edu/blog/techtank/2020/11/17/by-passing-proposition-24-california-voters-up-the-ante-on-federal-privacy-law/.

[146] *Id.* at §§ 1798.140, 1798.185(a)(16).

[147] *Id.* at § 1798.140 (emphasis added).

> as well as a description of the likely outcome of the process with respect to the consumer.[148]

The CPRA becomes effective on January 1, 2023, but at the time of this writing, the California Private Protection Agency has not yet released proposed regulations related to automated decision-making.

While the full scope of California's protections has yet to be developed, Colorado and Virginia's data protection laws ultimately provide little more than an ability for consumers to opt out of automated decision-making with respect to recruiting emails. Unlike the law in California, the data protection laws passed in Virginia and Colorado specifically exclude employees from their general provisions.[149] However, both laws do provide consumers the ability to opt out of the automated processing of their data. For example, the Virginia Consumer Data Protection Act ("VCDPA") permits a consumer to opt out of data processing for the purposes of "(i) targeted advertising, (ii) the sale of personal data, or (iii) profiling in furtherance of decisions that produce legal or similarly significant effects concerning the consumer."[150]

Even though the VCDPA and the Colorado Privacy Act ("CPA") broadly carve out coverage of employees, they make a distinction for automated decision-making for recruiting purposes. In defining the phrase "profiling in further of decisions that produce legal or similarly significant effects concerning the consumer," the VCDPA specifically references decisions that may affect "employment opportunities":

[148] *Id.* at § 1798.185(a)(16) (emphasis added).

[149] Virginia Consumer Data Protection Act, S. 1392, 2021 Va. Acts (2021) (to be codified at VA. CODE. ANN. § 59.1-571 (West 2023) (effective Jan. 1, 2023)) ("'Consumer' means a natural person who is a resident of the Commonwealth acting only in an individual or household context. It does not include a natural person acting in a commercial or employment context.") [hereinafter "VCDPA"]; Colorado Privacy Act, 2021 Colo. Sess. Laws 3445 (to be codified at COL. REV. STAT. ANN. § 6-1-1303(6) (West 2023) (effective Jan. 1, 2023)) ("Consumer" means "an individual who is a Colorado resident acting only in an individual or household context; and (b) does not include an individual acting in a commercial or employment context, as a job applicant, or as a beneficiary of someone acting in an employment context.") [hereinafter "CPA"].

[150] VCDPA § 59.1-573. The VCDPA further defines "profiling" in a manner that would capture automated decision-making: "any form of automated processing performed on personal data to evaluate, analyze, or predict personal aspects related to an identified or identifiable natural person's economic situation, health, personal preferences, interests, reliability, behavior, location, or movements." *Id.* at § 59.1-571.

> 'Decisions that produce legal or similarly significant effects concerning a consumer' means a decision made by the controller that results in the provision or denial by the controller of . . . employment opportunities.[151]

In language that is nearly verbatim to that passed by Virginia, the Colorado Privacy Act also provides consumers the ability to opt out of "profiling," including "decisions that produce legal or similarly significant effects concerning a consumer," and defines such decisions to include those that affect employment opportunities.[152] As they contain general exemptions for employees and applicants, the CPA and VCDPA appear to provide little more than the ability of consumers to opt out of unsolicited emails which recruiters have sent as the result of using automated recruiting software, and therefore they do little to mitigate potential negative consequences and unfairness associated generally with ADM systems for employment purposes.

Protections under current data privacy laws for automated decision-making in the employment context are therefore quite limited in the United States. Although regulations in California have yet to be released, if they follow in the same general direction as those in Virginia and Colorado, they may provide little more than the right to opt out of automated decision-making for non-applicants. In other words, once an individual actively engages in the application process, their rights to opt out may disappear. Even if the right to opt out were extended to employees and applicants, these laws would provide no requirement for employers to provide notice that they have been subject to an adverse decision as the result of automated decision-making software. Employees therefore will likely not have knowledge of even the basic involvement of ADM software, and they will be unlikely to challenge the decisions on that basis, if at all.

3. Direct Regulations of AI Products in the United States

i. Illinois Artificial Intelligence Video Interview Act

In 2020, Illinois became the first state to pass a standalone law directly addressing the use of artificial intelligence systems for

[151] *Id.* at § 59.1-571.

[152] CPA §§ 6-1-1303(10)(20), 6-1-1306.

employment purposes.[153] The Illinois law contains a consent requirement and an additional feature related to accountability.

First, the Illinois Act requires a specific set of disclosures and consent from the applicant prior to an employer deploying an AI system during an applicant interview. The Act requires that the employer: (a) "[n]otify each applicant before the interview that artificial intelligence may be used to analyze the applicant's video interview"; (b) "[p]rovide each applicant with information before the interview explaining how the artificial intelligence works and what general types of characteristics it uses to evaluate applicants";[154] and (c) obtain consent before the interview for use of the AI program.[155]

Second, Illinois requires a significant degree of accountability through ongoing evaluation and reporting related to the use of such programs. It requires that companies that "rely solely" upon an AI analysis of a video interview to then gather data on the race and ethnicity of both those applicants who are offered an in-person interview and those applicants who are subsequently hired. The companies must subsequently report that data to the Illinois Department of Commerce and Economic Opportunity on a yearly basis.[156]

ii. New York City Artificial Intelligence Ordinance

Following the Illinois Artificial Intelligence Video Interview Act, New York City passed its own municipal ordinance regarding artificial intelligence in 2021. Like the Illinois act, New York City specifically targets the use of ADM systems in the recruiting process, but the New York Ordinance goes further.

First, the New York City AI Ordinance requires independent bias audits. These audits, whose requirements are not described in detail by the statute, must be conducted once a year prior to the use of

[153] Illinois Artificial Intelligence Video Interview Act, 820 ILL. COMP. STAT. 42/1 to 42/20 (2022) [hereinafter "Illinois AI Video Interview Act"].

[154] Professor Ifeoma Ajunwa criticizes the lack of specificity with respect to what information must be provided to applicants. Ajunwa, *supra* note 3, at 645. She does not, however, offer any specific alternative language. Given the nascent nature of the technology and the frequent preference of legislatures for technology neutral language, legal flexibility may be a strength of the law allowing courts or regulatory agencies to provide further guidance.

[155] Illinois AI Video Interview Act, 820 ILCS 42/5.

[156] *Id.* at 42/20.

such a tool, and the most recent bias audit must be published on the employer's website.[157]

Second, the Ordinance provides a general notice and right to opt-out. In particular, the employer must provide any employee or candidate screened: (a) notice no less than 10 business days before the use of the ADM system, and (b) the ability to request an alternative selection procedure.[158]

Lastly, the Ordinance requires disclosure of certain basic system information. An employer in the city utilizing an ADM system must provide information about the type of data collected and the source of data to the employee or candidate either (a) on the company website, or (b) within 30 days of a written request.[159]

B. *Global Frameworks*

1. E.U. General Data Protection Regulation

Although the United States has limited legislation regarding data privacy and AI systems, the same is not true in Europe. In 2016, the European Union passed the General Data Protection Regulation ("GDPR") in an attempt to both harmonize the data privacy practices of Member Nations for furtherance of trade and to protect the fundamental rights of privacy and nondiscrimination of member citizens.[160] Following its passage and implementation, the GDPR soon became the gold standard for data privacy, emulated by numerous foreign countries and used as a template for state legislatures in the United States hoping to pass their own data privacy laws.[161]

The GDPR bestows numerous rights and responsibilities on data subjects and data controllers, but it also includes a specific right of access that allows data subjects to request from a data controller what information is being processed about them. This right contains a specific requirement with respect to ADM systems:

> The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and where that is the case, access to the personal data and the following

[157] N.Y.C. ADMIN. CODE § 20-871(a)(1)(2) (2022).

[158] *See id.*

[159] *Id.* at § 20-871(b)(3).

[160] GDPR rec. 1, 10, 71.

[161] Elizabeth L. Feld, *United States Data Privacy Law: The Domino Effect after the GDPR*, 24 N.C. BANKING INST. 481, 489–96 (2020).

> information: . . . (h) the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, **meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject**.[162]

In addition to this "right to know," the GDPR also provides a right to opt-out of automated decision-making.[163] The GDPR gives further reasoning under Recital 71 and highlights employment as a particular area in which automated decision-making may create significant legal effects and where consumers' rights must be protected through transparency, consent, and the right to obtain human intervention or review of an ADM's processes.[164]

2. E.U. Artificial Intelligence Proposal

The European Union's Artificial Intelligence Proposal ("AI Proposal"), introduced in April 2021, builds on the rights provided to consumers in the GDPR. The AI Proposal extends governing mechanisms and accountability to all stages of production of AI systems: testing, bringing AI systems to market, post-market monitoring, and remediation of potential malfunctions and negatives impacts.[165]

The AI Proposal uses a risk classification scheme and defines certain types or uses of AI systems as "high risk," and therefore subject to stricter regulation. Annex III of the AI Proposal classifies most types of ADM systems that would be used both to support and to make employment-related decisions as "high risk":

[162] GDPR art. 15(1)(h) (emphasis added).

[163] *Id.* at art. 22(1) ("The data subject 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.").

[164] *Id.* at rec. 71 ("In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision.").

[165] E.U. AI Proposal art. 8–15.

> Employment, workers management and access to self-employment:
> (a) AI systems intended to be used for recruitment or selection of natural persons, notably for advertising vacancies, screening or filtering applications, evaluating candidates in the course of interviews or tests;
> (b) AI intended to be used for making decisions on promotion and termination of work-related contractual relationships, for task allocation and for monitoring and evaluating performance and behavior of persons in such relationships.[166]

The Proposal justifies this classification by referencing the significant impacts that such systems may have on people's livelihoods, their potential for the perpetuation of discrimination, and the possibility that they may be used in a way that invades employee privacy. It states:

> AI systems used in employment, workers management and access to self-employment, notably for the recruitment and selection of persons, for making decisions on promotion and termination and for task allocation, monitoring or evaluation of persons in work-related contractual relationships, should also be classified as high-risk, since those systems may appreciably impact future career prospects and livelihoods of these persons . . . . Throughout the recruitment process and in the evaluation, promotion, or retention of persons in work-related contractual relationships, such systems may perpetuate historical patterns of discrimination, for example against women, certain age groups, persons with disabilities, or persons of certain racial or ethnic origins or sexual orientation. AI systems used to monitor the performance and behaviour of these persons may also

[166] *Id.* at annex 3(4).

> impact their rights to data protection and privacy.[167]

As a result of this classification as "high-risk AI systems," the Proposal imposes restrictions on the development, marketing, and management of the products. Software developers are held to requirements including: (a) the establishment of a risk management system,[168] (b) a data governance system to ensure that "[t]raining, validation, and testing data shall be relevant, representative, free of errors and complete,"[169] (c) technical documentation including general characteristics of the system, and detailed descriptions of the systems specifications and functioning,[170] (d) record-keeping systems that include automated logs of decisions that the AI system makes,[171] (e) transparency and provision of information to AI system users in order to ensure that "operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately,"[172] and (f) building measures to ensure that the AI system can be "effectively overseen by natural persons" while in use, including giving humans the ability to override system outputs or stop system processes altogether.[173]

While the GDPR and the EU AI Proposal both provide extensive protections and guidelines for consumers and employees who may be affected by AI systems, they do not offer tailored solutions that take into consideration the business considerations or existing laws of the industries or practices they may be regulating. In the next Section, this Article will look at ways to specifically tailor an approach to AI and ADMs that is more appropriate for employees.

IV. APPLYING DATA PRIVACY PRINCIPLES IN CONJUNCTION WITH ANTIDISCRIMINATION LAW

Regulations and potential legislation aimed at monitoring, ensuring fairness, and remediating the negative implications of ADM systems for employment can take many forms. Primary among those are tools to promote transparency, including various types of privacy notices; principles that emphasize safe and fair systems during the design stage; ongoing risk management measures; impact assessments;

[167] *Id.* at rec. 36.

[168] *Id.* at art. 9.

[169] *Id.* at art. 10.

[170] *Id.* at art. 11.

[171] E.U. AI Proposal art. 12.

[172] *Id.* at art. 13.

[173] *Id.* at art. 14.

accountability such as reports to government agencies, technical documentation, and third-party audits; individual rights that allow opt-outs; human oversight of ADM systems; and requirements regarding accuracy and fairness of systems with respect to potential disparate impact discrimination. Frameworks built specifically for privacy, which have been imported into proposed schemes for artificial intelligence and ADM systems, however, do not necessarily fit, especially when applied to the employment relationship. Some work well for ADM systems, such as the necessity for transparency. Others, such as opt-out schemes, undermine the very validity and benefits of the AI or machine learning systems themselves.

As discussed previously, the United States has in place a significant legal framework aimed at preventing unfair discrimination based on characteristics such as race, sex, national origin, and age. This framework, unfortunately, has lay dormant due to the invisibility of these ADM "dark systems." In order to effectively leverage the existing anti-discrimination laws, we should first look to strengthen requirements regarding notice and transparency. In addition, because of the unique nature of ADM systems, it would also likely benefit employees and applicants to require a threshold level of human oversight and interaction, which would allow for better long-term benefits from ADM systems and may mitigate their negative effects by providing a faster and more efficient remedy as compared to litigation.

A. *Notice and Transparency Requirements for ADM Systems*

1. Disclosure of the Use of ADM System to Applicants and Employees

Systems for disclosure and notice may take many forms. However, at this point in time, employees and applicants are currently not entitled to any notice under federal law that they are being monitored, evaluated, or scored based on ADM systems. If an applicant applies for a job, and an employer utilizes an ATS to review and score their resume, the applicant may be sorted to the lowest category or assigned a specific score such that no human ever bothers to review their application. Although the ADM may not be tasked with making a final decision on a job application, it may make a de facto determination simply through scoring, sorting, or issuing a recommendation.[174]

Without notice and disclosure of their use, employees and their attorneys do not have adequate information to make an inference or

[174] Citron, *supra* note 1, at 1271.

form an opinion that they may have been subject to an adverse employment action based on unfair or illegal criteria.[175] Thus, antidiscrimination protections, such as those under Title VII or the Age Discrimination in Employment Act ("ADEA"), will likely not be invoked. This is, unfortunately, consistent with the state of the system that we have observed in the past. Although ADM systems have proliferated in use, there has been little litigation, and thus little or no chance for the systems to be reviewed and vetted by the public or in the courts.[176]

Disclosure of the use of such systems either at the time of their use, or in conjunction with notification regarding an adverse employment action, would have significant benefits.[177] It would, as noted above, give applicants, employees, and their attorneys fair notice that a policy or practice was utilized that may have been the root cause of disparate impact. If, for example, an applicant knows that they met all of the requirements with respect to knowledge, skills, and abilities associated with a job posting, possessing things such as the requisite education and experience, then the disclosure that their application was rejected by an ADM system may give rise to the inference that the system is not functioning as intended, or that it has begun to make inferences or correlations that reject candidates for unfair or discriminatory reasons. Likewise, because many issues of discrimination and harassment have garnered media attention, and because many companies often respond to public and media attention, these types of disclosures may spur both more responsible use of ADM systems by companies and increase the compliance efforts of companies which design and market ADM systems for use in human resources.

[175] Hannah Bloch-Wehba, *Access to Algorithms*, 88 FORDHAM L. REV. 1265, 1271 (2020) ("Disclosure is the core mechanism of U.S. transparency law, which enshrines values of public access to government decision-making."); Citron, *supra* note 1, at 1249 ("Due process requires agencies to provide individual notice and an opportunity to be heard . . . . This century's automated decision making systems combine individual adjudications with rulemaking while adhering to the procedural safeguards of neither.").

[176] Successful challenges to algorithmic decision-making systems have been made in the public sector, but those legal challenges have not been based on discrimination or unfairness. Instead, they have focused on statutory and constitutional procedural protections. Bloch-Wehba, *supra* note 175, at 1294.

[177] Citron & Pasquale, *supra* note 1, at 20 ("[T]he underlying values of due process - transparency, accuracy, accountability, participate, and fairness - should animate the oversight of scoring systems given their profound impact on people's lives.").

Transparency also reinforces individual privacy rights. Notice and consent typically support a defense against claims of common law violations of invasion of privacy.[178] Transparency also forms a key principle in data privacy regulations.[179] Requiring notice to applicants and employees when their data is gathered and used as inputs by an ADM to make an employment decision, while not providing the full panoply of rights under legislation such as the GDPR,[180] would be a significant step forward in enhancing privacy protections in the United States.

2. Disclosure of System Specifics

In addition to notice of their use, further disclosures regarding ADM systems and their functionality would also benefit applicants and employees.[181] Disclosures should first include the types of data used, in other words, the inputs. As a baseline, notice of the specific inputs used can be reviewed to ensure that the ADM systems are not specifically taking into account legally-protected characteristics, but they can also be used to determine the possible use of data that may act as commonly-known proxies for protected characteristics, such as gaps in tenure which may be a proxy for gender, or location of employees' homes, which may be a proxy for race.

The specifics of the decisions or recommendations made by the ADM systems, and whether those decisions are reviewed by humans, should also be disclosed to applicants and employees. Whether a system makes the autonomous decision to hire or fire an employee, or merely creates a score or recommendation that is subsequently reviewed or used as consideration when making the final decision, is a critical piece of information. Sometimes, reviewers may fall subject to automation bias, in which they reflexively agree with a computer recommendation, or they assign too much weight to that

[178] Matthew E. Swaya & Stacey R. Eisenstein, *Emerging Technology in the Workplace*, 21 LAB. LAW. 1, 13 (2005).

[179] Frank Hendrickx, *Privacy 4.0 at Work: Regulating Employment, Technology and Automation*, 41 COMP. LAB. L. & POL'Y J. 147, 161 (2019).

[180] GDPR art. 13–15.

[181] Some scholars argue that even robust disclosures, such as opening up inspection of algorithmic code, is insufficient because it does not allow the recipients to determine exactly how an individual decision was reached in their case. *See, e.g.*, Bloch-Wehba, *supra* note 175, at 1270. However, even if initial disclosures about the use of ADMs are not determinative, disclosure may be sufficient to provide applicants and employees with enough information to form an inference of discrimination, which is often all that many aggrieved parties have in cases of disparate impact discrimination prior to litigation and the formal discovery process.

recommendation due to their high degree of trust in the program.[182] In other scenarios, however, ADM systems may be fully entrusted by companies and employers to make the final decision on an employee's status. In both cases, the applicant or employee should be provided this information, as it is a key component in understanding what the potential impacts of the ADM system were on their particular case, and if those impacts were potentially mitigated by human oversight.

Notices should also include a reasonable and appropriate description of how the ADM system functions or arrives at its outputs based on a given set of inputs.[183] The specifics and extent of such disclosure merit some examination.[184] Although this may be the most challenging aspect of disclosure for employers or developers, it will also likely be the most informative and useful information for the applicants and employees. In the case of ADM systems that merely conduct a statistical analysis or a comparison of straightforward data sets, the requirement will not be difficult to meet. For example, a company could state no more than, "The ADM system compares key words found in your resume . . . to the words in the employer['s] job description and issues a percentage-based score reflecting the relative match." For more advanced ADM systems deploying machine learning, however, the underlying difficulty is that such systems are opaque, and human programmers may not be able to accurately describe how a given system arrived at an output because they themselves cannot adequately track the system's correlations and logic.

This uncertainty is reflected in the current legislation and

[182] Margot Kaminski & Jennifer M. Urban, *The Right to Contest AI*, 121 COLUM. L. REV. 1957, 1960–61 (2021).

[183] *See* Kroll et al., *supra* note 131, at 634. Kroll et al. argues in favor of technical solutions to the underlying issues of AI and discrimination. In so doing, however, they place specific emphasis on transparency and procedural regularity. Key to their proposal is that "these techniques can assure that decisions are made with the key governance attribute of procedural regularity, meaning that decisions are made under an announced set of rules consistently applied in each case." *Id.*

[184] Such notice may depend on many factors, including the specific employment decision being made. It is therefore difficult to conceptualize any uniform or precise requirement. Even in areas that have addressed notice issues for decades, such standards are flexible. *See, e.g.*, Citron & Pasquale, *supra* note 1, at 27 ("Under the Due Process Clause, notice must be 'reasonably calculated' to inform individuals of the government's claims against them. The sufficiency of notice depends upon its ability to inform affected individuals about the issues to be decided, the evidence supporting the government's decision, and the agency's decisional process.") (internal citations omitted).

proposals. New York City's AI Ordinance, for example, requires disclosure of the inputs, but does not require disclosure of the system functionality. The GDPR requires disclosure of "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject."[185] Legislators appear trapped between the desire to provide specific and meaningful information to data subjects, and the inability to know what those pieces of information will be due to evolving technology and machine learning black box problems. Given uncertainty both with respect to how these systems will evolve in the future, as well as the wide degree of approaches that they may currently use, the most prudent current approach would be to craft a requirement that takes into consideration the reasonableness of the disclosure and its relative appropriateness for the use in question.[186] While this flexible standard may not provide certainty to employers, it should encourage a high degree of disclosure while also providing room for creative development of different ADM systems in the future.

Lastly, companies should be required to disclose the specific ADM system used, such as its name, model, and version number. This information can be used as an easy reference point for applicants, employees, litigants, and even the public to cross-reference and determine whether certain systems are disproportionately identified as being involved in unfair or illegal adverse employment actions.[187]

3. A Hypothetical Application and Comparison with the FCRA

The risks of unfairness inherent in an employee background check are analogous to those risks arising from ADM systems used for employment purposes. Take for example a hypothetical applicant, Leeloo, who applies for a job as a Senior Sales Representative at Zorg Industries. Leeloo has years of experience as a sales representative, and she meets all of the minimum and preferred qualifications for the

[185] GDPR art. 15(1)(h).

[186] One viable approach to the specific legislative language may be found in the Fair Credit Reporting Act. When an individual is denied credit based on a background or credit check, the FCRA requires the consumer be provided with notice that includes "all relevant elements or reasons adversely affecting the credit score for the particular individual, listed in the order of their importance." 15 U.S.C. §1681g(f)(2)(B).

[187] Hannah Block-Wehba explores a number of algorithmic bias cases in the public sector, and argues that the cases demonstrate an underlying implication that the public has a right to know of the methodology which is being deployed to rate and score them. Bloch-Wehba, *supra* note 175, at 1296, 1306–14.

job with Zorg Industries. But the company's Human Resources Department is undermanned and has been ordered by its tyrannical owner to terminate a million employees and slash hiring. To increase the efficiency of its applicant screening process, it has instituted two policies. First, it prescreens all job applicants based on a criminal background check. Leeloo, unfortunately, shares a common name with a different and unrelated Leeloo living in her county who has a recent conviction for armed robbery and murder. Second, Zorg Industries uses a new recruiting software product which uses automated decision-making to compare applicants' resumes based on similarity of language and keywords as compared to the Senior Sales Representative job description. The written job description, however, uses many words that contain gender-specific connotations, including such descriptors as "proven leader," "aggressive salesperson," and "doesn't take no for an answer."

Zorg's background check process and ADM software, unencumbered by legal restriction, would likely screen out Leeloo from employment prior to an interview or a human review of her resume. But, under current law, Leeloo is protected by the FCRA. Zorg Industries would have to provide a notice and disclosure of its intent to use a background check, and it would be required to provide a pre-adverse action notice containing a copy of the background check with the erroneous reports for aggravated robbery and murder. The innocent Leeloo would then have a "reasonable time" to correct the record by demonstrating that she has no pending criminal charges against her.[188]

With respect to the use of the new recruiting software, however, Zorg Industries would have no current legal obligation under federal law to inform Leeloo of the reasons for declining her application or the fact that an ADM system was even used to evaluate her resume and ultimately decide her fate. If she were provided such information and an opportunity to object to the decision, Leeloo could establish simply by referral to her resume that she met the qualifications of the job. Under existing antidiscrimination laws, such as Title VII, Leeloo could establish a prima facie case of disparate

[188] Presented with such evidence, a credit reporting agency that failed to correct an erroneous background check report could be held liable under the FCRA. 15 U.S.C. § 1681e(b) ("Whenever a consumer reporting agency prepares a consumer report it shall follow reasonable procedures to assure maximum possible accuracy of the information concerning the individual about whom the report relates.").

treatment discrimination for failure to hire.[189] Zorg Industries, however, would likely be able to defeat a claim for disparate treatment by demonstrating that it had a legitimate, nondiscriminatory reason for the rejection of Leeloo's application through demonstration that the ADM software made the decision without reference to her protected class.[190]

Instead, the stronger legal claim by Leeloo would be for disparate impact discrimination. A disparate impact claim would require that she establish the particular employment practice, in this case the ADM system, declined her resume because of her membership in a protected class.[191] Usually, such a claim could be supported by statistical evidence.[192] If Leeloo is provided with notice and disclosures regarding the ADM systems involvement in declining her application, she may draw an inference of discrimination and raise a complaint with Zorg's Human Resources Department. Even if she does not have statistical evidence, Zorg Industries will have access to the underlying data for job applicants in the recruiting process, and it could likely determine even at this early stage of the complaint process that the use of the ADM software was having a disparate impact on protected classes of applicants. But if even the company did not have the early advantage of a statistical analysis, the Human Resources Department or hiring managers could review Leeloo's resume and determine she possessed the necessary qualifications. This would raise an inference that the ADM software made a decision based on something that may not qualify as a bona fide factor other than sex or

[189] A plaintiff may establish a prima facie case of discrimination in a failure to hire case by offering evidence that: (a) the plaintiff belongs to a protected class, (b) the plaintiff applied for and was qualified for a job for which the employer was seeking applicants, (c) plaintiff was rejected despite his or her qualifications, and (d) the position remained open and the employer continued to seek applicants from persons of plaintiff's qualifications. *McDonnell Douglas Corp. v. Green*, 411 U.S. 798, 802 (1973).

[190] If an employee presents evidence to establish a prima facie case of discrimination, the burden shifts to the employer to articulate a "legitimate nondiscriminatory reason" for the adverse employment action. *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 802 (1973); *Texas Dept. of Comm. Affairs v. Burdine*, 450 U.S. 248, 252–53 (1981).

[191] *Watson v. Fort Worth Bank and Trust*, 487 U.S. 977, 994 (1988).

[192] "A prima facie case of disparate impact is usually accomplished by statistical evidence showing that an employment practice selects members of a protected class in a proportion smaller than their percentage in the pool of actual applicants." *Stout v. Potter*, 276 F.3d 1118, 1122 (9th Cir. 2002).

race.[193] Zorg Industries would then have several options, including passing Leeloo on to the next stage of the selection process and beginning an internal audit of the ADM software.

The FCRA adverse action process provides a model for what should be adopted when automated decision-making systems are employed for purposes of adverse action decisions in the employment context. Like third-party background checks, ADM systems have quickly become part of an elaborate and often opaque system of employee and applicant evaluations, and much like the historical credit reporting practices in place prior to the passage of the FCRA, employees and applicants who are subject to these systems currently have no legal right to receive notice of the basis of those decisions or to dispute their accuracy or fairness. Providing such rights would increase transparency into ADM systems and their potential misuse, and the ability to object to inaccurate and unfair preliminary decisions would raise the likelihood that employers would correct them before they can be made final.

4. Benefits of Transparency

There are numerous benefits to transparency for ADM systems used for employment purposes. First, transparency provides the requisite information for individuals to know whether they might have a legal claim for disparate impact discrimination, which is currently lacking.[194] Although basic notifications may not provide information such as relative selection rate or statistical effects on various protected categories, and such a requirement would likely prove infeasible, notice may still provide an ability for potential plaintiffs to understand that an ADM system may have influenced their hiring or firing. From

[193] The California Equal Pay Act bars employers from disparate wages rates except where the employer demonstrates the wage differential is based upon a seniority system, a merit system, a system that measures earnings by quantity or quality of production, or a "bona fide factor other than sex, such as education, training, or experience." CAL. LAB. CODE § 1197.5(a) (West 2022).

[194] Citron, *supra* note 1, at 1281–82 (stating that, with respect to governmental due process, "Automated decision systems endanger the basic right to be given notice of an agency's intended actions. This right requires that notice be 'reasonably calculated' to inform individuals of the government's claims. The sufficiency of notice depends upon its ability to inform affected individuals about the issues to be decided, the evidence supporting the government's position, and the agency's decisional process. Clear notice decreases the likelihood that agency action will rest upon 'incorrect or misleading factual premises or on the misapplication of rules.'") (internal citations omitted).

that point, the individual may have enough information to form an inference of discrimination if, for example, they know that they were qualified for the job in question, or if they know they were performing their job commensurately with other employees that were not terminated in the case of an ADM system that judges work performance.[195] Gathering further information and discovery on the ADM system's impacts on applicants or employees can then be accomplished through litigation, which is best suited to such a purpose.

Second, even when not leading to litigation, a system of disclosure will allow applicants and employees the right to contest a decision early and efficiently. Take, for example, the earlier example of Leeloo's application to Zorg Industries, and the legal model of the Fair Credit Reporting Act. By providing notice to applicants like Leeloo that an ADM system has made or contributed to an adverse employment decision, and by granting them a limited window to request a human review of the decision, Zorg Industries and companies like it will be forced to explain and be held immediately accountable for their use of ADM systems. Flaws may be discovered sooner, and applicants, employees, and employers could avoid costly and prolonged litigation by fixing issues earlier.

Lastly, transparency with respect to ADM systems' functionality will likely foster the development of explainable AI, or XAI. True black box systems whose outputs cannot be explained will be deterred from participation in the market if employers and software designers cannot meet the legal requirements associated with deploying them for use in the real world. Systems which can process input and meaningfully explain outputs should find higher adoption. This, in turn, will promote investment and development in such systems, and should benefit both employers and employees in the long run.

5. What Early Opt-Out Systems, such as the New York Ordinance, Get Wrong

[195] Kroll et al. acknowledge that "full or partial transparency can be a helpful tool for governance," but they also argue that transparency is not sufficient to provide accountability. Kroll et al., *supra* note 131, at 657–58. A primary contention that they make is that full transparency of computer code as well as key inputs and outputs may lead to individuals gaming the algorithms. *Id.* at 658. This argument may be persuasive for AI used fraud or tax evasion, but it does not appear well-suited to employment scenarios in which employees and job applicants have a legitimate expectation in understanding the standards judging their job performance.

Assuming that applicants and employees receive timely notice of the use of an ADM system to evaluate them or make an adverse action decision, an alternative approach is to allow the employees to opt out of the ADM process altogether. This, for example, is the approach that the New York AI Ordinance has taken, and the approach that data privacy regimes provide to consumers when they do not wish their personal data to be processed by automated means or through profiling. Allowing an early opt-out option for employees, however, is a mistake.

First, allowing early employee opt-outs will likely corrupt the integrity of the data upon which these systems rely for their performance and improvement.[196] In the case of an applicant tracking system, if a large group of applicants opts out due to selection bias and mistrust of AI or ML systems, removal of that group and their corresponding characteristics may skew the data for the remaining population. Where a legal analysis may be done later for auditing, internal evaluation of the system's performance, or due to litigation, the remaining data will not tell the full story of how those applicants would have been evaluated or how the ADM system would have treated them.

Second, machine learning systems, when programmed and utilized correctly, improve with successive iterations, learning and fine-tuning mistakes through trial and error. Again, if a specific and self-selected population removes itself through an opt-out mechanism, a ML system may never learn to properly assess their characteristics or

[196] Joseph W. Sakshaug, Alexandra Schmucker, Frauke Kreuter, Mick P. Couper & Eleanor Singer, *Evaluating Active (Opt-In) and Passive (Opt-Out) Consent Bias in the Transfer of Federal Contact Data to a Third-Party Survey Agency*, 4 J. SURV. STAT. & METHODOLOGY 382, 386, 402–03 (2014) (finding that opt-in and opt-outs increase selection bias and decrease sample size, with opt-ins having the greater detrimental effect on selection bias); EXEC. OFFICE OF THE PRESIDENT, BIG DATA: A REPORT ON ALGORITHMIC SYSTEMS, OPPORTUNITY, AND CIVIL RIGHTS 8 (May 2016), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf (noting that selection bias may occur in the inputs for algorithms "where the set of data inputs to a model is not representative of a population and thereby results in conclusions that could favor certain groups over others."); Ignacio N. Cofone, *Algorithmic Discrimination Is an Information Problem*, 70 HASTINGS L.J. 1389, 1402–03 (2019) ("Oftentimes, the data fed to algorithms suffer from a self-selection problem."); Bent, *supra* note 12, at 812 ("If, for example, data are more readily available for men than women, or for younger applicants than older applicants, the algorithm may unintentionally disfavor the underrepresented group due to the data's inaccurate reflection of the relevant population.").

performance. The systems cannot learn from absent data. On the other hand, if given the chance to evaluate the data of the employees or applicants, make a decision based on the available data, and then have that decision reviewed and corrected by human oversight, an ADM system may be able to learn from and improve on its mistakes in the future.

B. *Forms and Benefits of Human Oversight*

AI and machine learning systems work most effectively when paired with the innate abilities of humans. Remember, for example, that AI systems are typically poor at basic human activities, tasks such as perception of the real world, intuition, common sense, reasoning by analogy, and transfer learning. AI and machine learning, however, have capacities that go far beyond human abilities for understanding and processing large amounts of data. Human oversight, therefore, can work in conjunction with notice and transparency obligations.[197]

Human oversight, or a "human in the loop," can take many forms. As discussed previously, invocation of human review prior to the ADM system making an adverse employment decision is not the optimal approach. When employees and applicants have been subject to an adverse action, and they are subsequently provided with notice and the right to request human review of that decision, human oversight may have the benefit of remedying an unjust or inaccurate decision soon after it is made, and without the need for litigation.

Human oversight may also be utilized for the ongoing monitoring, review, and correction of ADM systems during their performance. This is the approach taken by the EU Artificial Intelligence Proposal, which requires ongoing monitoring by humans of the processes and results of a high-risk AI system, as well as human ability to stop such systems. The proposal states:

> High-risk AI systems should be designed and developed in such a way that natural persons can oversee their functioning. For this purpose, appropriate human oversight measures should be identified by the provider of the system before its placing on the market or putting into service. In particular, where appropriate, such measures should guarantee that the system is subject to in-built

[197] Kroll et al., *supra* note 131, at 639 ("That is, while transparency of a rule makes reviewing the basis of decisions more possible, it is not a substitute for individualized review of particular decisions.").

> operational constraints that cannot be overridden by the system itself and is responsive to the human operator, and that the natural persons to whom human oversight has been assigned have the necessary competence, training and authority to carry out that role.[198]

The rationale for this approach, as outlined by the Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission, is that AI should respect human autonomy and should be designed to work in conjunction with humans instead of being designed to replace them. The group states:

> The fundamental rights upon which the EU is founded are directed towards ensuring respect for the freedom and autonomy of human beings. Humans interacting with AI systems must be able to keep full and effective self-determination over themselves, and be able to partake in the democratic process. AI systems should not unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans. Instead, **they should be designed to augment, complement and empower human cognitive, social and cultural skills. The allocation of functions between humans and AI systems should follow human-centric design principles and leave meaningful opportunity for human choice. This means securing human oversight over work processes in AI systems.** AI systems may also fundamentally change the work sphere. It should support humans in the working environment, and aim for the creation of meaningful work.[199]

ADM systems may significantly benefit from human oversight, and systems built with human-centric designs may be more

[198] E.U. AI Proposal rec. 48.

[199] INDEPENDENT HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, ETHICS GUIDELINES FOR TRUSTWORTHY AI 12, 14 (2019) (emphasis added), https://digital-strategy.ec.europa.eu/library/ethics-guidelines-trustworthy-ai.

fair and may outperform alternative systems without such safeguards. As we have seen, current artificial intelligence and machine learning systems lack capacities such as basic common sense and the ability to reason by analogy. They cannot always independently identify their own errors, such as when they draw a discriminatory inference about employees. Many companies have discovered systematic issues of discrimination after trialing such programs, but these investigations require competent human review. This is consistent with the trend observed by most artificial intelligence researchers: current AI technology and humans tend to complement each other. Perhaps at some point in the future these trends will change. AI may surpass humanity at the ability to perceive and reason about our everyday world, and it may be that general AI systems will eventually be able to effectively oversee more specialized systems. Until that time arrives, requiring human oversight of ADM systems will be the best approach to ensuring such systems continue to comport with our ideals of fairness.

C. *Independent Audits*

Independent audits like those required by the New York Ordinance, while well-intentioned, are unlikely to be effective. A threshold question arises of how such audits will be practically conducted. As demonstrated by many artificial intelligence and machine learning systems, a fundamental difficulty arises when moving from the learning stages to application to real-world data. This transition exposes the systems to unexpected scenarios and edge cases, often referred to as “long-tail” problems because when plotted on a distribution such scenarios are not often seen except in the “long-tail” of large sets of data.[200] Audits conducted without the benefit of this real-world data likely will not be able to predict or detect the actual problems that the ADM systems will encounter once actually deployed.[201] “Testing of any kind is . . . a fundamentally limited approach to determining whether any fact about a computer system is true or untrue.”[202]

[200] MITCHELL, *supra* note 4, at 100–02.

[201] Kroll et al., *supra* note 131, at 651 (“Even structured ‘audits’ of software systems, in which systems are provided with related inputs and analyzed for differential behavior, cannot provide complete coverage of a program's behavior for the same reason: this methodology explains little about what happens to inputs which have not been tested, even those that differ very slightly.”).

[202] *Id.* at 652.

We must also question the realistic chances for the impartiality and independence of such audits. Without sufficient constraints and guidelines, audits may unfortunately be used in a self-serving manner by the developers to justify their own programs. In January 2021, HireVue made dual announcements in a public statement, announcing that it would no longer use its facial analysis algorithms as part of its applicant assessments, while also releasing a copy of an audit from Cathy O'Neil's organization, ORCAA, which HireVue implied demonstrated a lack of bias in its product.[203] At the time, HireVue stated, "The audit concluded that '[HireVue] assessments work as advertised with regard to fairness and bias issues.'" An independent reviewer, however, concluded that HireVue misrepresented its findings, and ORCAA did not comment on the matter.[204] Of particular concern, HireVue provided a link to the ORCAA report, but required anyone who accessed it to pledge not to publicly disclose its contents.[205] HireVue's actions in this case potentially raise serious implications about companies using an audit for justification of potentially biased and unfair principles, mischaracterizing that AI audit for their own commercial purposes, and also insulating the actual contents of the audit from public review and discussion.

This is not an isolated problem. Alex Engler writes, "HireVue's shady behavior encapsulates the challenges facing the emerging market for algorithmic audits. While the concept sounds similar to well-established auditing practices such as financial accounting and tax compliance, algorithmic auditing lacks the necessary incentives to function as a check on AI applications."[206] Current legislation regarding these audits suffers from a lack of specificity, which compounds the problem that such audits will not be fairly or impartially conducted. For example, the New York Ordinance simply requires employers to conduct a bias audit, which it defines as

[203] *HireVue leads the industry with commitment to transparent and ethical use of AI in hiring*, HIREVUE (Jan. 21, 2022), https://www.hirevue.com/press-release/hirevue-leads-the-industry-with-commitment-to-transparent-and-ethical-use-of-ai-in-hiring.

[204] Engler, *supra* note 6.

[205] As of the time of this writing, the report is available online at https://www.hirevue.com/resources/template/orcaa-report. In order to access or download it, a user must agree to the following statement: "By downloading this document you acknowledge and agree this report is the sole and exclusive intellectual property of HireVue, Inc., and you agree you shall not use, copy, excerpt, reproduce, distribute, display, publish, etc. the contents of this report in whole, or in part, for any purpose not expressly authorized in writing by HireVue, Inc."

[206] Engler, *supra* note 6.

"an impartial evaluation by an independent auditor. Such bias audit shall include but not be limited to the testing of an automated employment decision tool to assess the tool's disparate impact on persons" based on their protected characteristics.[207] The Ordinance says nothing about the characteristics of the auditor, such as required expertise or certification. It does not describe the means by which a bias audit should be conducted, including the applicable statistical analysis or even what might constitute a disparate impact.

Lastly, where the auditors are paid by the software developers for their work, they may have little incentive to be critical of their clients' programs. Although a significant portion of the audit may entail a mathematical or statistical analysis, there are numerous subjective decisions to be made that can influence the outcomes. For example, the basic choice of whether to evaluate a program based on a four-fifths rule or a standard deviation analysis will often be determinative based on factors such as the size of the employer and the pool of data.[208] An auditor may therefore be motivated to produce a favorable audit, and this danger is heightened by the regulatory uncertainty regarding the audit requirements and the current low likelihood of enforcement proceedings against customers.[209]

D. *Why Regulate ADM Systems At All: The Market Approach Alternative*

Some may question the necessity for regulating artificial intelligence or ADM systems at all. The benefits of efficiency and even potential non-discrimination may arguably outweigh the costs. Cathy O'Neil, in Weapons of Math Destruction, discusses this argument and juxtaposes it with the case of regulation of sexual orientation as a protected characteristic for employment purposes.[210] In the late 1990s, when President Bill Clinton signed the Defense of Marriage Act ("DOMA"), there was an outpouring of corporate support in favor of the rights of the LGBT community. Many large corporations made

[207] N.Y.C. ADMIN. CODE § 20-870.

[208] Peresie, *supra* note 123, at 783.

[209] Some scholars have argued that independent audits are necessary in order to remediate potential bias and discrimination. Ajunwa, *supra* note 3, at 664–65. Audits may suffer from a lack of adequate data to determine discrimination since applicants and employees are requested, but not required, to provide their demographic information. Professor Ajunwa argues that such information should be legally required of applicants. *Id.* However, this solution only deepens the potential privacy violations and the risks of a cybersecurity breach that exposes applicants and employees' sensitive data.

[210] O'NEIL, *supra* note 1, at 200–02.

public statements and publicized policies pledging that they would not discriminate against applicants or employees based on their sexual orientation, despite no legal obligation to do so. From this, detractors of government regulation may argue or conclude that laws requiring fair and equal treatment of employees based on their protected categories have simply outlived their usefulness.

In the case of sexual orientation, however, corporations did not simply do the right thing out of their public interest, but because of their own economic motivations. Non-discrimination policies related to sexual orientation allow those corporations to compete more effectively for talent, and especially for talent from members of the LGBT community. Conversely, the companies at the time making such public statements and promises had little or nothing to gain from sexual orientation discrimination, and many therefore voluntarily promoted policies of equality.[211]

The situation here is strikingly different. ADM systems promote efficiency and earn companies' money. Employers have little or no interest in self-regulating such systems and turning back the clock to have armies of recruiters and sourcers hired to sort through stack upon stack of applicant resumes.[212] Likewise, they have no interest in transparency or disclosing the basis of their decisions when such practices could subject them to claims of disparate impact discrimination.

ADMs also have a tendency toward self-fulfilling feedback loops which may reinforce employers' use and reliance on biased data.[213] If initially programming an ADM using training data from a historically non-diverse workforce, an employer then continues to discriminate in hiring and promotions, and may continue to train successive iterations of ADM systems on the newly updated, but still biased, data. Without proper regulatory protections, there are few economic incentives for employers to remedy the flaws in these systems. It is therefore incumbent upon legislators to require such disclosures and put in place regulations for the benefit of the applicants and employees.

V. CONCLUSION

Our regulation and relationship to AI, machine learning, and any other type of automated decision-making system in employment

[211] *Id.*

[212] Kim, *supra* note 6, at 894 ("So long as the algorithm is accurate enough to make the employer's process less costly, neither the employer nor the vendor will have sufficient incentive to identify and remove the bias.").

[213] *Id.* at 895–96.

ultimately depends on our perceptions of the usefulness of such products and their relative costs. This should, by many accounts, be a straightforward analysis, analogous to many other cost-benefit analyses that we undertake both overtly and implicitly when weighing regulations for other products and services. We calculate the benefits of mass transportation through automobiles, and weigh those against the relative costs of air pollution, noise pollution, depletion of raw materials, and traffic fatalities. We weigh the benefits of the alternative, riding on horses, and the costs of housing, feeding, tidying up after the horse's natural processes, and potentially being bitten or kicked when the horse has a bad day. We choose cars.

As compared to other technologies, however, artificial intelligence triggers a unique and more visceral reaction. The very term is often euphemized to seem more benign to the audience, with speakers deploying terms such as machine learning, neural networks, or even in this article, "automated decision-making." Western culture has a particular fear and fascination with humanity, collectively or individually, being supplanted by machines. Perhaps, in some ways, this is not unfounded. We have seen shifts in our society away from an agrarian economy, and then again away from factory and production work, as the result of advances in technological capacity. This third revolution of information technology, which has only begun, has changed the day-to-day work of nearly everyone on the planet. As the saying goes, however, change is not necessarily bad. Access to a worldwide database of information, to our friends and family through smart phone video calls, and to the tools of work productivity likely outweigh the costs of having to address disinformation campaigns on social media and constantly being forced to hang up on spambot calls. The costs of artificial intelligence, however, seem to strike deeper. People believe the machines will want to, at best, replace us, and at worst to destroy us. The chorus of regulations emerging reflect this, emphasizing that humans continually be kept in the loop of AI systems, beginning at the point of inception, and that those humans have the ability to shut down the AI if it somehow runs amok.

Some of these fears are likely misplaced. Software programs do what they are told. They have no inherent beliefs, desires, or emotions. While programs like Deep Blue, IBM's Watson, and AlphaGo conquered the world's best human competitors in games such as chess, Jeopardy, and Go, the AI systems responsible for such feats of prowess obtained no satisfaction from their victories. They had no pride. They did not even know that they were beating humans. Artificial intelligence software does what it is programmed to do, often in incredibly narrow and rationalist terms, and it should in the future have no animosity towards humans unless those same humans intend it to. Writers like Nick Bostrom may speculate that, given their limited common sense and real-world understanding, such systems could inadvertently run amok even without intent. He gives the example of an AI system programmed to produce as many paper clips as possible, and which begins to consume all the natural resources of earth (including its humans) to do so, only to turn to the stars and other planets to produce yet more paper clips once the earth is exhausted.[214] While thought-provoking, the hypothetical seems conceptually impossible. Such an artificial intelligence system would have to progress to a relative superintelligence, capable of overcoming all other competing AI systems and world defenses, while simultaneously remaining blithely unaware of the real-world implications, costs, and morality of its actions.[215] Accordingly, the warning story seems better fare for a children's parable than a serious building block for discussion of whether and how to address the costs and benefits of such important technology.

Some of these ideas may stem from Western religious and philosophical ideas about the transcendence of the human soul above the physical world and the human body, which may be reframed as an inherent conflict between mankind against soulless machines.[216] The metanarratives of eventual conflict and replacement by artificial intelligence or smart machines are not, however, universal. If we look at Japanese ideas about the relationship between artificial intelligence and humanity, we see that perceptions are shaped by traditional religious ideas such as Buddhism and animism, which focus on the positive interconnections between nature, mankind, and our man-made creations. Animistic principles imply that artificial intelligences may have a spirit or soul. Thus, humans and machines are not fundamentally different, and are not inevitably headed toward conflict. As a result, narratives regarding the eventual fate of AI and man do not center around conflict and replacement, but the constructive relationship possible between the two.[217]

Putting irrational fears aside, current technology does have significant problems. Dark systems, such as those which may invisibly perpetuate bias and historical discrimination, should be addressed. But while the current state and near-future vulnerabilities and uses of ADM systems do merit close scrutiny and regulation, regulators and society should temper the inherent impulse to overreact and overregulate.

[214] NICK BOSTROM, SUPERINTELLIGENCE: PATHS, DANGERS, STRATEGIES 123 (2014).

[215] MARCUS & DAVIS, *supra* note 27, at 196.

[216] MARK COECKELBERGH, AI ETHICS 39–45 (2020).

[217] *Id.* at 47.

While technology-neutral laws that can be applied across all industries may be tempting, they may also not be possible or desirable with respect to technology whose trajectory we cannot accurately predict in the long-term, or where we have existing legal frameworks that have been tailored to meet a specific purpose, such as combating employment discrimination.