ALGORITHMS AS ALLIES:
REGULATING NEW TECHNOLOGIES IN THE FIGHT FOR WORKPLACE EQUALITY

*Jack Hensler**

ABSTRACT

Applying algorithms in recruitment and hiring can implicate anti-discrimination law by concealing encoded bias in data processing and delivery. This Comment breaks down, through an extensive hypothetical, the ways in which an algorithm can violate Title VII. It then assesses how a court would likely resolve this form of discrimination.

My research cautions against the adoption of an overly broad regulation, like Europe's General Data Protection Regulation. The federal framework of anti-discrimination law, given its history of incorporating statistical analysis, appears well-suited to address algorithmic discrimination. Still, the feasibility of auditing for discriminatory impacts before they occur presents a compelling justification for proactive regulation.

TABLE OF CONTENTS

I. INTRODUCTION

Should Congress look to Europe for solutions to protect against algorithmic discrimination, or can a moderate adaptation of federal law adequately address this challenge? Machines do not conform to human notions of equality.[1] An algorithm[2] does not recognize the optics of its results or appreciate historical narrative. Emerging technologies, such as artificial intelligence (AI)[3] and machine learning[4]

*J.D. Candidate, Temple University Beasley School of Law, 2020; B.A., Economics and History, University of Notre Dame, 2015. I wish to thank my advisor, Professor Erika Douglas, for her insightful feedback and Professor Brishen Rogers for his classes in labor and employment law. To my family, friends, and classmates who generously and patiently listened to my musings on algorithmic discrimination—thank you.

1. *See* Francis F. Steen, *The Paradox of Narrative Thinking*, 3 J. CULTURAL & EVOLUTIONARY PSYCHOL. 87, 88 (2005) (explaining how humans, in contrast with computers, think in narrative forms). While a computer detects patterns as they exist in the present, human conceptualizations of justice and equality rely heavily on historical narrative. Douglas J. Sylvester, *Myth in Restorative Justice History*, 2003 UTAH L. REV. 471, 475 (2003) (discussing how scholars employ historical narrative in argument).

2. *See generally* Paul Zandbergen, *What Is a Computer Algorithm? - Design, Examples & Optimization*, STUDY.COM, https://study.com/academy/lesson/what-is-a-computer-algorithm-design-examples-optimization.html#/transcriptHeader (last visited Sept. 9, 2019) (explaining that algorithms are procedures that allow computers to solve a wide array of problems and describing algorithms as standard or unambiguous because they will repeat the same procedure without subjective considerations).

3. *See generally* B.J. Copeland, *Artificial Intelligence*, ENCYCLOPÆDIA BRITANNICA (May 9, 2019), https://www.britannica.com/technology/artificial-intelligence (defining artificial intelligence as the capacity of computers to perform tasks usually associated with intelligent beings).

4. *See* MEHRYAR MOHRI ET AL., FOUNDATIONS OF MACHINE-LEARNING 1–2 (2d ed. 2018)

in particular, increase the likelihood that algorithms will circumvent anti-discrimination law[5] by concealing encoded bias.[6] This Comment considers how machine learning influences corporate decision-makers in recruiting and hiring. In three parts, it explores (i) the ways in which machine learning can violate Title VII of the Civil Rights Act of 1964 ("Title VII" or "the Act") through flawed data mining and model formation; (ii) how the design choices surrounding data presentation can encourage discrimination; and (iii) why now is the time to get ahead of algorithmic discrimination through proactive regulations.

AI applications, such as driverless cars[7] and predictive healthcare technology,[8] offer solutions to some of society's most intractable problems. Driverless cars are expected to dramatically reduce carbon emissions, improve safety for drivers, and decrease traffic volume.[9] Machine learning tools in healthcare, such as IBM's Watson, assist in determining treatment options for certain cancers.[10] Predictive analytics[11] outperform a strict application of "clinical decision support tools"[12] in predicting in-hospital mortality rates for patients with sepsis.[13] In a less corporate

(defining machine learning as applying computational methods to process past information and identify patterns, with the purpose of making accurate predictions or optimizing performance).

5. *See generally Laws Enforced by EEOC*, U.S. EQUAL EMP. OPPORTUNITY COMMISSION, https://www.eeoc.gov/laws/statutes/ (last visited Sept. 9, 2019) (providing a full list of anti-discrimination statutes in the United States).

6. *See* Gideon Mann & Cathy O'Neil, *Hiring Algorithms Are Not Neutral*, HARV. BUS. REV. (Dec. 9, 2016), https://hbr.org/2016/12/hiring-algorithms-are-not-neutral (discussing how algorithms may appear neutral but actually take on human biases because they mimic human decision-making).

7. *See generally* Cadie Thompson, *8 Ways Self-Driving Cars Will Drastically Improve Our Lives*, BUS. INSIDER (Dec. 14, 2016, 10:26 AM), https://www.businessinsider.com/how-driverless-cars-will-change-lives-2016-12 (discussing the potential benefits of self-driving cars).

8. *See, e.g.*, R. Andrew Taylor et al., *Prediction of In-Hospital Mortality in Emergency Department Patients with Sepsis: A Local Big Data-Driven, Machine Learning Approach*, 23 ACAD. EMERGENCY MED. 269, 274–75 (2016), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5884101/.

9. *See* Thompson, *supra* note 7 (examining the tremendous benefits that driverless cars will have for society and the environment).

10. *See* Dan Shewan, *10 Companies Using Machine Learning in Cool Ways*, WORDSTREAM, https://www.wordstream.com/blog/ws/2017/07/28/machine-learning-applications (last updated Aug. 12, 2019) (discussing, among other applications, machine learning in treatment recommendations for certain types of cancer).

11. *See generally* Seewon Ryu, *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie or Die*, 19 HEALTHCARE INFORMATICS RES. 63, 63 (2013) (reviewing ERIC SEIGEL, PREDICTIVE ANALYTICS: THE POWER TO PREDICT WHO WILL CLICK, BUY, LIE OR DIE (2013)) (describing how predictive analytics in the healthcare context requires analyzing patient data for health signals).

12. *See generally* Jennifer Bresnick, *Understanding the Basics of Clinical Decision Support Systems*, HEALTH IT ANALYTICS (Dec. 12, 2017), https://healthitanalytics.com/features/understanding-the-basics-of-clinical-decision-support-systems (describing clinical decision support systems as tools meant to assist healthcare professionals in dealing with enormous amounts of medical data and detecting problems).

13. Taylor et al., *supra* note 8, at 276.

and more familiar setting, platforms like Netflix,[14] Spotify,[15] and Amazon[16] use AI to make personalized recommendations based on prior selections and search history. Evidence of AI's capacity to materially improve people's lives abounds in fields from infrastructure to healthcare.[17]

Beyond these fields, AI offers promising solutions to a significant issue in employment: eliminating unconscious bias in recruiting and hiring.[18] Companies like Pymetrics[19] and HireVue[20] apply machine learning to recognize and combat unconscious bias that can lead to inequitable hiring.[21] Because machines are not trained in the same way that individuals are socialized,[22] implementing an algorithm to assess and rank potential job candidates can remove bias from résumé reviews or interviews and facilitate the hiring of a diverse workforce.[23]

Like any innovative solution, AI also presents a new set of challenges; regulators and courts must keep pace with the rapid innovation that has transformed the hiring process.[24] In a recent national survey, 55% of human resources managers

---

14. *See* Matt Burgess, *This Is How Netflix's Secret Recommendation System Works*, WIRED (Aug. 18, 2018), https://www.wired.co.uk/article/netflix-data-personalisation-watching (providing examples of the ways that Netflix uses algorithms to customize what subscribers watch).

15. *See* Eric Boam, *I Decoded the Spotify Recommendation Algorithm. Here's What I Found.*, MEDIUM (Jan. 14, 2019), https://medium.com/@ericboam/i-decoded-the-spotify-recommendation-algorithm-heres-what-i-found-4b0f3654035b (examining Spotify's use of machine-driven algorithms to provide users with personalized playlists and recommendations).

16. *See* Tom Buckland, *Amazon A9 Optimization & Algorithm*, AMAZON SEO CONSULTANT (Jan. 17, 2018) https://amazonseoconsultant.com/amazon-a9-optimization/ (last updated Aug. 15, 2019) (detailing Amazon's A9 algorithm and its effects on searches through Amazon's search engine).

17. *See* Thompson, *supra* note 7 (forecasting reduced traffic among other benefits of self-driving cars); *see also* Shewan, *supra* 10 (identifying the use of machine learning in cancer treatment).

18. *See* Audrey J. Lee, Comment, *Unconscious Bias Theory in Employment Discrimination Litigation*, 40 HARV. C.R.-C.L. L. REV. 481, 482 (2005) (discussing the subtler nature of the barrier that unconscious bias imposes in the workplace).

19. *See generally Science*, PYMETRICS, https://www.pymetrics.com/science/ (last visited Sept. 27, 2019) (introducing company strategies of using machine learning to remove bias from corporate hiring).

20. *See generally Bias, AI Ethics, and the HireVue Approach*, HIREVUE, https://www.hirevue.com/ (last visited Sept. 11, 2019) (discussing the company's commitment to promoting diversity and fairness in the hiring process).

21. *See, e.g.*, *Science*, *supra* note 19 (discussing Pymetrics' commitment to produce a bias-free algorithm).

22. *See* Lee, *supra* note 18, at 483–84 (providing an overview of the social science reasons why individuals tend to group similar objects, resulting in stereotypes).

23. *See, e.g.*, *Science*, *supra* note 19 (explaining the proactive steps taken to ensure the algorithms offer an unbiased method of candidate review)*; see also* PYMETRICS, GENDER EQUALITY WHITEPAPER: METHODICALLY BREAKING THE GLASS CEILING (explaining how the Pymetrics methodology facilitates gender equality in hiring) (available upon request at https://www.pymetrics.com/science/).

24. *See More Than Half of HR Managers Say Artificial Intelligence Will Become a Regular Part of HR in Next 5 Years*, CAREERBUILDER (May 18, 2017), http://press.careerbuilder.com/2017-05-18-More-Than-Half-of-HR-Managers-Say-Artificial-Intelligence-Will-Become-a-Regular-Part-of-HR-in-Next-5-Years (discussing how, as of 2017, 13% of HR managers had already

claimed AI would become a substantial part of their jobs within five years.[25] While many companies plan to use AI in a modest fashion to serve as a first round of sorting through candidate résumés, others believe that AI will lead to a wholesale transformation of the hiring process.[26]

Ahead of the curve in adopting new technologies, in 2014, Amazon began developing a candidate assessment algorithm that processed large amounts of data and provided an output of what it considered the best résumés.[27] It identified key terms in résumé and assigned values to these terms corresponding to whether they were predictive of success or failure.[28] While an oft-cited advantage of algorithms in the hiring process is their ability to remove bias, this algorithm began to internalize the bias of the data it processed and developed a strong preference for male candidates.[29] The algorithm tended to prefer masculine verbs such as "executed" and "captured" and penalized résumé that showed membership in women's organizations.[30] While engineers likely initiated this endeavor with good intentions, they did not account for the algorithm inheriting bias from the data it processed.[31] The project was dropped in 2017.[32]

As the programmers at Amazon came to understand, algorithmic discrimination can stem from multiple causes.[33] The model might not be able to draw accurate conclusions due to insufficient data,[34] or it might inherit human biases by picking up discriminatory patterns.[35] Algorithms can also draw adverse inferences about

reported evidence of regular application of artificial intelligence in human resources)*; see also* Ben Dattner et al., *The Legal and Ethical Implications of Using AI in Hiring*, HARV. BUS. REV. (Apr. 25, 2019), https://hbr.org/2019/04/the-legal-and-ethical-implications-of-using-ai-in-hiring (contemplating the potential impact that using AI in hiring may have on federal and state employment laws).

25. *More Than Half of HR Managers Say Artificial Intelligence Will Become a Regular Part of HR in Next 5 Years*, *supra* note 24.

26. *See* Aigerim Berzinya, *Artificial Intelligence in Human Resources: Key Innovation Trends*, BIG DATA MADE SIMPLE (Dec. 19, 2018), https://bigdata-madesimple.com/artificial-intelligence-in-human-resources-key-innovation-trends/ (emphasizing the increasing importance of automation in HR departments).

27. *See* Jordan Weissmann, *Amazon Created a Hiring Tool Using A.I. It Immediately Started Discriminating Against Women*, SLATE (Oct. 10, 2018, 4:52PM), https://slate.com/business/2018/10/amazon-artificial-intelligence-hiring-discrimination-women.html (detailing Amazon's use of AI to rate job candidates and expedite the hiring process).

28. *See id.* (describing Amazon's machine learning algorithm's reliance on key terms from the résumés of past applicants in searching for new hires).

29. *Id.*

30. *Id.*

31. *See id.* (discussing how the engineers failed to consider the effects of training the algorithm on past applicants—the majority being men).

32. *Id.*

33. See *infra* Part II for a technical analysis of the ways in which algorithms can discriminate.

34. *See* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 675 (2016) (including the use of inadequate underlying data sets among the reasons why algorithms sometimes discriminate).

35. *See* Anup Kumar, *Basics of Machine Learning*, GALAXY TRAINING, https://galaxyproject.github.io/trainingmaterial/topics/statistics/tutorials/machinelearning/tutorial.html (last visited Sept. 12, 2019) (describing how machine learning algorithms analyze user data to

those who do not regularly use technology by penalizing blank values.[36] And, almost by definition, machine learning algorithms tend to perpetuate existing trends and biases.[37] This leaves us with a question: when algorithms retain an inequitable status quo, should lawmakers task data scientists with manipulating models to ensure equitable outcomes?[38]

My research suggests that the courts stand well-equipped to handle cases of algorithmic discrimination in hiring and recruitment.[39] Detailed analysis of statistics is not a new phenomenon in the Title VII context.[40] For decades now, the courts have scrutinized hiring policies through extensive data or regression analysis.[41] While AI may not eliminate bias in recruiting and hiring, there is little evidence that machines will amplify the complexities of judicial resolution.[42]

My research also cautions against rushing to adopt a broad regulation on data processing that lacks specificity.[43] Several scholars have advocated for a wholesale adoption of Europe's General Data Protection Regulation (GDPR), or something similar, in the United States.[44] But, at least for algorithmic discrimination in private employment, a broad adoption would not offer meaningful improvements to the existing Title VII framework.[45] Additionally, a regulation modeled after the GDPR could discourage innovative solutions that show promise in battling subconscious bias among employers.[46] However, the GDPR's provisions that (i) distinguish types

---

learn features and create predictions).

36. Kishan Maladkar, *5 Ways to Handle Missing Values in Machine Learning Datasets*, ANALYTICS INDIA MAG. (Feb. 9, 2018), https://www.analyticsindiamag.com/5-ways-handle-missing-values-machine-learning-datasets/ (discussing methods of dealing with inaccurate, deleted, or blank data).

37. *See* Barocas & Selbst, *supra* note 34, at 691 (indicating that data mining can lead to prejudicial algorithms despite the use of genuinely relevant criteria).

38. See *infra* Part IV for a discussion of possible justifications for imposing affirmative obligations on data processors.

39. *See, e.g.*, Hazelwood Sch. Dist. v. United States, 433 U.S. 299, 307–13 (1977) (basing its ruling in large part on an in-depth statistical analysis of this systemic disparate treatment case).

40. *See, e.g.*, *id.* (showing the use of data analysis in a U.S. Supreme Court case from 1977).

41. *See id.* (using primarily statistics in deciding an employment discrimination claim)*; see also* Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 339 (1977) (discussing the importance of statistics in disputed discrimination cases).

42. *See Int'l Bhd. of Teamsters*, 431 U.S. at 339–40 (suggesting that statistics are important in cases involving discrimination but still may be refuted like ordinary evidence).

43. See *infra* Part II for a discussion of potential regulatory solutions.

44. *See, e.g.*, *Copy That: America Should Borrow from Europe's Data-Privacy Law*, ECONOMIST (Apr. 5, 2018), https://www.economist.com/leaders/2018/04/05/america-should-borrow-from-europes-data-privacy-law (suggesting that America should copy the European Union's GDPR approach)*;see also* Gus Rossi, *Is the GDPR Right for the United States?*, PUB. KNOWLEDGE (Apr. 9, 2018), https://www.publicknowledge.org/is-the-gdpr-right-for-the-united-states/ (providing a more cautious view towards the United States adopting the GDPR).

45. See *infra* Part II.C for a discussion of Title VII's ability to resolve issues of algorithmic discrimination.

46. *See* Seth P. Berman, *GDPR in the U.S.: Be Careful What You Wish For*, GOV'T TECH. (May 23, 2018), www.govtech.com/analysis/GDPR-in-the-US-Be-Careful-What-You-Wish-For.html (explaining how some of the rights espoused by the GDPR conflict with the U.S. legal system).

of companies based on how they interact with data[47] and (ii) set up a system for proactive monitoring[48] are small pieces that can patch up existing regulatory gaps in the United States.

Despite the drawbacks of an overly broad regulation, the increased application of AI offers compelling reasons for regulating its use in recruitment and hiring, primarily because of improved feasibility.[49] The time is ripe for such a regulation for three reasons. First, the data and the resources for analyzing potential disparate impacts are readily available for proactive auditing.[50] Second, federal courts have created workable standards for assessing data analysis, such as the two standard deviations rule for systemic disparate treatment[51] and the four-fifths rule for disparate impact analysis.[52] Third, the Equal Employment Opportunity Commission (EEOC) is fully capable of performing the type of statistical analysis that a regulation would require.[53] An effective regulation should impose an obligation on data processors[54] to proactively audit their models,[55] correct for bias, and verify that they have a sufficiently large and diverse data pool to provide for trustworthy models.[56]

The following sections analyze the extent to which Title VII addresses different types of algorithmic discrimination. Part II provides an overview of Title VII and, through an extensive hypothetical, explores how an algorithm can implicate the Act. Part III explores how data delivery and the presentation of results can encourage discrimination and potentially incur liability under Title VII. Part IV then considers why, despite the limited risk that algorithms will successfully circumvent Title VII,

47. *See* 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), 2016 O.J. (L 119) 33 [hereinafter GDPR] (distinguishing different types of data entities, such as controllers from processors).

48. *Id.* at 53.

49. *More Than Half of HR Managers Say Artificial Intelligence Will Become a Regular Part of HR in Next 5 Years*, *supra* note 24 (offering evidence of increased use).

50. *See* Michiel Rhoen & Qing Yi Feng, *Why the 'Computer Says No': Illustrating Big Data's Discrimination Risk Through Complex Systems Science*, 8 INT'L DATA PRIVACY L. 140, 140–41 (2018) (noting the increasing quantifiability of the world, which the authors refer to as datafication).

51. *See* Hazelwood Sch. Dist. v. United States, 433 U.S. 299, 308 n.14 (1977) (discussing how a disparity as great as two standard deviations can provide the basis for an inference of discrimination).

52. *See* Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607.4(D) (2015) (explaining the four-fifths rule and how it provides evidence of an adverse impact).

53. *See generally id.* (referencing statistical benchmarks of fairness throughout the guidelines).

54. Throughout this Comment, the term "data processors" will refer to companies that perform operations on data to assess and rank applicants. *See* GDPR, *supra* note 47, at 33, art. 4 ("'[P]rocessor' means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller.").

55. *See, e.g.*, GDPR, *supra* note 47, at 53–54, art. 35 (requiring audits, or data protection impact assessments, whenever there is an algorithm that imposes a high risk of legal implications).

56. *See id.* at 54–55 (requiring the organization of a supervisory authority which would determine the specifics for such monitoring).

now is the time to get ahead of discrimination in the American workplace.

II. DISCRIMINATION IN DATA PROCESSING

When a company deploys algorithms in hiring and recruitment, courts may be called upon to assess compliance with Title VII.[57] Under Title VII, discrimination occurs when a decision exhibits a legally unjustified preference that disadvantages an individual or class of individuals based on some protected characteristic.[58] Title VII prohibits discrimination in employment and, as the primary equal opportunity legislation in the United States,[59] provides broad examples of methods of proof in anti-discrimination law.[60]

A. Title VII Background

Title VII's coverage is limited to specific types of actors, categorized as "employers, labor organizations and employment agencies."[61] The statute defines an employer as having "fifteen or more employees for each working day in each of twenty or more calendar weeks in the current or preceding calendar year."[62] This employer numerical requirement creates a safe harbor for small businesses, but the Act does not impose a similar threshold on the U.S. government or employment agencies.[63] The Act defines an employment agency, regardless of its number of employees, as a company that regularly refers employees to employers.[64] The Act specifically prohibits employment agencies from taking adverse actions, referring, or *classifying* on the basis of protected characteristics.[65]

Under Title VII, an individual plaintiff can prove discrimination under either the theory of disparate treatment, which deals with intentional discrimination, or disparate impact, which deals with legally unjustified discriminatory results.[66] Because a machine cannot form intent in the same way as a human, at first blush, the disparate treatment approach seems like a misfit in the context of algorithmic discrimination.[67] However, disparate treatment appears to be the most viable theory

---

57. *See* 42 U.S.C. § 2000e-5(a)–(f) (2012) (explaining the procedures for enforcement).

58. See *infra* Part II.A for an in-depth analysis of Title VII.

59. *Title VII of the Civil Rights Act of 1964*, KPPB L., https://www.kppblaw.com/practices/labor-employment/employment-litigation/title-vii/ (last visited Sept. 25, 2019).

60. *See EEOC Litigation Statistics: FY 1997 through FY 2017*, U.S. EQUAL EMP. OPPORTUNITY COMMISSION, https://www.eeoc.gov/eeoc/statistics/enforcement/litigation.cfm (last visited Sept. 1, 2019) (providing a database of Equal Opportunity employment litigation statistics).

61. *See* 42 U.S.C. § 2000e (promulgating employment rules only onto employers, labor organizations, and employment agencies).

62. *Id.* § 2000e(b).

63. *Id.* § 2000e.

64. *Id.* § 2000e(c).

65. *Id.* § 2000e-2(b) (indicating race, color, religion, sex, or national origin as protected characteristics).

66. *See id.* § 2000e-2(k) (setting forth the burden of proof for a disparate impact case)*; see also* Barocas & Selbst, *supra* note 34, at 694 (providing an overview of the available theories for plaintiffs under anti-discrimination law).

67. *See* Steen, *supra* note 1, at 87–88 (discussing narrative forms of thought as a uniquely

for cases involving the presentation and design of results, such as filtering results on the basis of race or sorting by age.[68]

Disparate treatment requires proving intentional discrimination, but not animus or bad faith, so employers can still face liability despite their benign motivations.[69] For example, in *International Union v. Johnson Controls*, an employer sought to exclude women capable of bearing children from jobs that involved lead exposure.[70] Although the company may have had the best of intentions towards its female employees and their potential children, the Supreme Court found that "the absence of a malevolent motive does not convert a facially discriminatory policy into a neutral policy."[71] The Court found the company in violation of Title VII for discriminating on the basis of sex.[72]

Functionally, courts examine disparate treatment claims under the *McDonnell-Douglas* burden-shifting framework.[73] According to the framework, a plaintiff must first establish a prima facie claim.[74] To do this, the plaintiff must demonstrate that he or she belonged to a protected group, that he or she was qualified for the rejected position, and that the employer continued accepting applications after passing over the plaintiff.[75] If the employer is able to demonstrate a legitimate nondiscriminatory reason for its decision, then the plaintiff has the burden of showing that such a reason was merely pretext.[76]

When multiple reasons contribute to a failure to hire, the plaintiff can succeed by showing that the discrimination was a "motivating factor" in the defendant employer's decision.[77] Once the plaintiff has proven that the employer intentionally discriminated against the plaintiff because of the plaintiff's status as a member of a protected class, the employer can only escape liability by proving that the discrimination was based on a bona fide occupational qualification, an extremely narrow exception.[78]

---

human trait not found elsewhere).

68. See *infra* Part II for further discussion of disparate treatment cases.

69. 499 U.S. 187, 198–200 (1991).

70. *Id.* at 190; *International Union v. Johnson Controls*, 499 U.S. 187, 198–200 (1991).

71. *Id.* at 199.

72. *Id*.

73. *See* McDonnell Douglas Corp. v. Green, 411 U.S. 792, 802 (1973) (providing the burden-shifting framework)*;see also* Barocas & Selbst, *supra* note 34, at 696 (explaining *McDonnell-Douglas* framework).

74. *McDonnell Douglas Corp.*, 411 U.S. at 802.

75. *See id.* (describing how a plaintiff must prove his or her Title VII claim).

76. *See* Barocas & Selbst, *supra* note 34, at 696–97 (describing the significance of proving that a defendant's given justification is merely pretext).

77. *See id.* at 697 (quoting 42 U.S.C. § 2000e-2(m) (2012)) (explaining the mixed-motive framework for trying disparate treatment cases).

78. *Compare* Int'l Union v. Johnson Controls, 499 U.S. 187, 190 (1991) (suggesting that the potential harm to future offspring did not qualify as a bona fide occupational qualification), *and* Michael J. Frank, *Justifiable Discrimination in the News and Entertainment Industries: Does Title VII Need a Race or Color BFOQ*, 35 U.S.F. L. REV. 473, 473–74 (2001) (stating that Title VII does not allow for race to serve as a bona fide occupational qualification), *with* Dothard v. Rawlinson, 433 U.S. 321, 335 (1977) (holding that gender was a bona fide occupational qualification for prison

Disparate treatment can occur individually[79] or systemically.[80] Plaintiffs who allege systemic disparate treatment may use statistics to establish an inference of discriminatory intent.[81] To determine whether a demographic imbalance in employment is statistically and legally significant, the court compares the concentration of current employees who belong to the protected group to the concentration of qualified potential applicants within the relevant labor pool.[82] In *Hazelwood School District v. United States,* the Supreme Court offered a framework for employers. It suggested that if the percentage of employees within the protected group were more than two or three standard deviations below the mean as compared to the percentage in the relevant labor pool, then the court would likely presume discrimination.[83] Such statistical baselines can provide valuable guidance to data processors.[84]

While applying statistics to disparate treatment cases is viable only under certain circumstances, considering which statistical correlations constitute discrimination is a common task under disparate impact analysis.[85] When proceeding on the theory of disparate impact, "the gravamen of the offense is the *unjustified impact* of the policy or action."[86] In other words, the plaintiff must show that a facially neutral practice erects a barrier that prohibits a protected group from achieving equitable employment results.[87] In *Griggs v. Duke Power Co.*, the Court held that, when viewed in the context of the history of segregated schooling, requiring a high school diploma and a median score on a generic intelligence test created an "artificial, arbitrary, or unnecessary barrier to employment."[88] In that same case, the Supreme Court also laid out the basic framework for proving disparate impact.[89]

After a prima facie disparate impact showing by the plaintiff, the burden shifts to the defendant to demonstrate that the disputed requirement bears a "relationship

---

guards, based on the security of the inmates).

79. *See, e.g.*, *McDonnell Douglas Corp.*, 411 U.S. at 792 (providing an example of individual discrimination).

80. *See, e.g.,* Hazelwood Sch. Dist. v. United States, 433 U.S. 299, 299 (1977) (providing an example of systemic discrimination).

81. *See* SAMUEL ESTREICHER ET AL., CASES AND MATERIALS ON EMPLOYMENT DISCRIMINATION 92–102 (5th ed. 2016) (discussing how statistical analysis may be used by courts to evaluate potentially discriminatory practices in hiring).

82. *See Hazelwood School Dist.*, 433 U.S. at 311 n.17 (explaining the legal significance of using standard deviations as predictors of discrimination).

83. *Id.*

84. See *infra* Part IV for a discussion of effective data processing regulations to prevent employment discrimination.

85. *Compare* Connecticut v. Teal, 457 U.S. 440, 440 n.4 (1982) (basing much of its disparate impact analysis on statistics) *with* N.Y. Transit Auth. v. Beazer, 440 U.S. 568, 584 (1979) (conceding that statistical evidence may establish a prima facie case against an employer, but emphasizing that the statistics relied upon must pertain to the protected characteristics at issue).

86. ESTREICHER ET AL., *supra* note 81, at 123 (emphasis added).

87. Griggs v. Duke Power Co., 401 U.S. 424, 429–30 (1971).

88. *Id.*, at 431.

89. *Id.* at 432–34.

to successful performance of the specific job for which it was used."[90] To provide guidance for employers, the EEOC has set forth three methods for validating a selection procedure: "criterion-related, content, and construct."[91] Validating selection procedures ensures that employers, if sued for disparate impact, can avail themselves of the business necessity defense.[92] Additionally, the criterion-related validity method directly relates to machine learning, because it entails using empirical data to prove a criteria's correlation with, or predictive capacity of, a particular job skillset.[93]

Under the criterion-related validity standard, tests that tend to exclude a protected group do not satisfy the business necessity requirement unless "predictive of or significantly correlated with important elements of job performance."[94] In *Albemarle Paper Co. v. Moody*, validity tests revealed that a potentially discriminatory employment testing procedure was statistically correlated to job relatedness in only a few lines of work.[95] The Court noted that the "odd patchwork of results" did not justify the broad application of the employment tests for all skilled positions.[96] Normally, an employer can comply with the criterion-related guideline by comparing the performance ratings of high-level employees to the scores they received on the disputed test.[97] The employer must then show (i) sufficient correlation between the predictors and (ii) statistical significance, or a substantial likelihood that the patterns do not exist only by chance.[98]

EEOC guidelines, which are granted a degree of deference by courts, state that acceptable criteria can include "measures other than the actual work proficiency, such as training time, supervisory rations, regularity of attendance, and tenure," suggesting that the employer has a degree of discretion so long as the criteria appear legitimate.[99] Even if the employer is able to demonstrate that the test or required qualification is sufficiently correlated with or predictive of job-related proficiency, a plaintiff can put forth a less discriminatory alternative.[100] If the employer refuses to acquiesce to this alternative, then the court may impose liability on the employer.[101]

In summary, disparate treatment requires proving intent while disparate impact

90. *Id.* at 432.

91. 29 C.F.R. § 1607.5(B) (2015).

92. *But see id.* § 1607.3(A) (stating that selection procedures that do not comply with the uniform rules and have an adverse impact are considered discriminatory unless justified).

93. *Id.* § 1607.5(B).

94. *Id.*

95. 422 U.S. 405, 411 (1975); Albemarle Paper Co. v. Moody, 422 U.S. 405, 411 (1975).

96. *Id.* at 431–32.

97. *See* ESTREICHER ET AL., *supra* note 81, at 162 (explaining how a potentially discriminatory test may be deemed valid using statistics).

98. *Id.*

99. 29 C.F.R. § 1607.5(B)(3) (2015)*; see* Albemarble Paper Co., 422 U.S. at 412 (discussing how the Court of Appeals had deferred to EEOC regulations).

100. 42 U.S.C. § 2000e-2(k)(1)(A)(ii) (2012).

101. *Id.*

focuses primarily on mechanisms and outcomes.[102] Additionally, statistical analysis has historically played an important role in proving both theories.[103] The following section addresses how the theories of disparate treatment and disparate impact, developed in the 1970s to discourage individual bad actors and broad company policies from discriminating on the basis of protected status, can apply to machines.

B. Hiring at Process Perfectionist: A Hypothetical

While examples of algorithmic discrimination abound, this hypothetical serves as an efficient means of surveying various flaws that can occur in machine learning model formation and application.[104] Suppose a prestigious employer called Process Perfectionist offers consulting services and claims that its best performers are creative problem-solvers. Process Perfectionist has an overworked human resources department that receives thousands of applications a year and needs assistance in determining which applicants should receive an interview.[105]

Process Perfectionist reaches out to Alchemist Algorithms, a machine learning company specializing in recruiting and hiring, to assist in determining what makes an ideal applicant. In particular, Process Perfectionist is interested in generating a list of the top 100 candidates. Process Perfectionist has two datasets that it gives to Alchemist Algorithms. Dataset EMPLOYEE_EVAL contains information on current and past employees, including their quarterly review ratings from their managers, pay information, disciplinary history, ranking of their career goals, and application records from years ago. EMPLOYEE_EVAL is considered the "training data,"[106] which Alchemist Algorithms will analyze to discern patterns among top performers and form a model. Dataset APPLICANT_INFO contains information on recent applicants. Using that information, the model will sort the applicants according to how likely they are to become top performers in the future.

1. Defining the Target Variable

Before creating the model, humans at Process Perfectionist set the "target variable," which is the set of characteristics that they believe define a desirable candidate or employee.[107] The characteristics that define the target variable could include the likelihood that an applicant will accept an offer, probabilities regarding an applicant's long-term retention, potential for advancement, or ability to generate

102. *See, e.g.*, McDonnell Douglas Corp. v. Green, 411 U.S. 792, 792 (1973) (describing an instance of disparate treatment)*; see also* Barocas & Selbst, *supra* note 34, at 694 (describing what disparate impact entails).

103. *See* Barocas & Selbst, *supra* note 34, at 708 (detailing how statistical models and tests are used to find discrimination).

104. See *supra* Part I for more examples of algorithmic discrimination.

105. For a discussion of this common problem, see Kayla Kozan, *4 Rules For Optimizing High Volume Recruitment*, IDEAL (Aug. 3, 2017), https://ideal.com/high-volume-recruitment/.

106. *See* Barocas & Selbst, *supra* note 34, at 677, 680 (describing training data as the example from which statistically-significant patterns in a set of data are discerned by data mining processes).

107. *See id.* at 677–78 (defining target variable as the desired outcomes and naming examples).

new business.[108] The characteristics, or data points, that comprise the target variable should represent actual skills required for the job.[109] If instead the target variable includes characteristics of pedigree, such as attendance at a private college, then the final model will take on any biases associated with that trait.[110] For example, attendance at a particular school could be highly correlated with protected classes such as race, gender, or religion.[111]

2. Training and Forming the Model

Once the target variable is set, the computer program will process dataset EMPLOYEE_EVAL to discern patterns among those workers who best matched the target variable's set of desirable characteristics.[112] As the computer program identifies patterns, it will ascribe relative weight to each of these predictive characteristics.[113] The computer program will also identify how good the fit of the model is, or, in other words, the degree to which employee performance is "explained by" the patterns that it has identified.[114] Because the computer model can only search this particular database for patterns, EMPLOYEE_EVAL needs to contain enough data to accurately explain why certain individuals have proven successful.[115] Even for large datasets, the data may be too inconsistent for the model to effectively identify patterns that distinguish high performers from low performers.[116]

a. Hindsight Bias and Omitted Variable Bias

Datasets like EMPLOYEE_EVAL are backward-looking and prone to

---

108. However, it is important to note that under disparate impact doctrine, the attributes that compose the target variable need to be specific to the job under consideration. *See* ESTREICHER ET AL., *supra* note 81, at 162 (explaining that the validity of any one criterion depends upon whether it can predict successful job performance).

109. *See id.* (describing how job performance for applicants in relation to incumbent employees is a valid predictor).

110. *See* Barocas & Selbst, *supra* note 34, at 678–80 (discussing how it is the programmer's job to translate the problem into criteria that would be acceptable to a computer).

111. *See, e.g.*, *id.* at 682 (describing how a computer program for medical school applicants based on previous admissions could disfavor racial minorities and women).

112. *See* SHAI SHALEV-SHWARTZ & SHAI BEN-DAVID, UNDERSTANDING MACHINE-LEARNING: FROM THEORY TO ALGORITHMS 14 (2014) (referring to the "learner's output," or the creation of a prediction rule from the training data entered into the system).

113. *Id.*

114. Thomas J. Campbell, *Regression Analysis in Title VII Cases: Minimum Standards, Comparable Worth, and Other Issues Where Law and Statistics Meet*, 36 STAN. L. REV. 1299, 1302–03 (1984) (offering differences among employees in education and experience as two variables that might explain, or be highly correlated with, variance in salary and benefits).

115. *See* Jason Brownlee, *How Much Training Data is Required for Machine Learning?*, MACHINE LEARNING MASTERY (last updated May 23, 2019), https://machinelearningmastery.com/much-training-data-required-machine-learning/ (indicating that the amount of data gathered must be sufficient to demonstrate relationships between input and output data points).

116. *See* Campbell, *supra* note 114, at 1303 (discussing this issue in the context of judicial use of statistics when there is a lack of fit between the model and the selection process).

"hindsight bias"[117] because they can lead a model to presume that characteristics that led to success in the past will necessarily lead to success in the future.[118] However, in an era of rapidly changing technology, what made employees successful in the past likely differs from the skills that will continue to garner success in the present. In this way, hindsight bias may be understood as a tendency to discount those groups that have been historically excluded from the American workplace.[119]

This issue is compounded if there is a limited sample size of any particular group. If Process Perfectionist tended to exclude certain types of people in the past, like women or racial minorities, then the dataset EMPLOYEE_EVAL could lack sufficient information on these marginalized groups. As a result, the model could place an improper emphasis on irrelevant patterns because of the limited nature of the data.[120] For instance, if the sole woman in a department left within a year, the model may attribute the short retention of that employee to her gender, as it has not assessed any other women.

Because dataset EMPLOYEE_EVAL is limited to human resources department data, it excludes several data points that would prove predictive and therefore presents issues of omitted variable bias.[121] Omitted variable bias exists when a variable that is not available for the model or within a dataset is correlated with both the dependent target variable and another independent variable used in the model.[122] The variable that is included in the model will appear overly predictive, because it absorbs a degree of the omitted variable's predictive capacity.[123]

For example, suppose the EMPLOYEE_EVAL dataset contains information on employees' genders, but not their college majors. If the job is a type that benefits from a science or engineering degree, and the model contains mostly males who

117. Cathy O'Neil, *Amazon's Gender-Biased Algorithm Is Not Alone*, BLOOMBERG (Oct. 16, 2018, 9:00 AM), https://www.bloomberg.com/opinion/articles/2018-10-16/amazon-s-gender-biased-algorithm-is-not-alone (discussing the limits and dangers of analyzing past information to predict the future).

118. *See* Mayukh Bhaowal, *Back to the Future: Demystifying Hindsight Bias*, INFOQ (May 9, 2018), https://www.infoq.com/articles/data-leakage-hindsight-bias-machine-learning (using the example of the Titanic to illustrate hindsight bias).

119. *Id.*; *see Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues*, FED. TRADE COMMISSION 8 (2016), https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf (discussing potential risks, such as failure to give credit, based on biased data).

120. *See* Brownlee, *supra* note 115 (explaining that a lack of data will preclude relationships between input and output values from being captured); *see also* Bryce W. Goodman, *A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection* 3 (2016) (unpublished manuscript) (on file with Oxford Internet Institute) (discussing circumstances in which membership in a protected category is genuinely predictive of performance).

121. *See Omitted Variable Bias: A Comprehensive Econometrics Review*, ALBERT (Sept. 20, 2016, 3:00 PM), https://www.albert.io/blog/omitted-variable-bias-econometrics-review/ (demonstrating, through statistical analysis, that the omission of an important independent variable from a learning model will cause the model to misidentify causal relationships in the data).

122. *Id.*

123. *See id.* (assuming that both the included and omitted variable have a positive correlation with the target variable).

possess those degrees, then the model will encode the predictive capacity of the college degree into a variable that is actually included in the model, such as gender.[124] The model will then suggest that being a male is an indicator of success, even though it is really the possession of a science or engineering degree that predicts success.[125] Therefore, should a female engineer apply for that job, her application is likely to be improperly discounted on the basis of her sex.

b. Emergent Properties

The concept of emergent properties in computer science offers a particularly useful way to consider the risk of bias in algorithmic decision-making.[126] An emergent property has been defined as "a property displayed by a complex system that is not directly predictable from the properties of that systems' [sic] elements."[127] In other words, humans create complex systems[128] when they group together to form organizations: corporations, parishes, unions, etc.[129] These composite organizations interact with each other unpredictably, insofar as an observer's understanding of the system is limited to familiarity with its constituent members.[130]

In our current data-rich environment, processing oftentimes reveals that individuals have a tendency to group together based on shared characteristics.[131] This may lead them to the same chat rooms, Facebook groups, or shopping forums.[132] With such large amounts of available data and high processing speeds, facially neutral data can prove highly correlated with protected characteristics.[133] Suppose several employees and applicants for Process Perfectionist belong to a popular LinkedIn group called "OUTreach." Membership in this group may appear neutral; however, if the group is comprised mainly of LGBT individuals, then its inclusion in a data model would essentially make sexual orientation a factor in the model.[134] If the general work environment at the employer is hostile to LGBT individuals who then leave the company, then the model would indicate that

124. *See id.* ("[Omitting variable bias from your regression] results in over-estimating (upward bias) or under-estimating (downward) the effect of one of more other explanatory variables.").

125. *See id.* (considering the omission of square feet in home price, which I analogize to the absence of including the type of degree in a model for hiring).

126. *See* Rhoen & Feng, *supra* note 50, at 144 (describing how algorithms can be biased based on protected trait analysis).

127. *Id.* at 142–43.

128. *See About NECSI*, NEW ENG. COMPLEX SYS. INST., http://necsi.edu/about/ (last visited Oct. 1, 2019) (offering a clear and concise definition of complex systems).

129. *See generally* Yaneer Bar-Yam, *Complexity Rising: From Human Beings to Human Civilization, a Complexity Profile*, ENCYCLOPEDIA OF LIFE SUPPORT SYS. (2002), https://static1.squarespace.com/static/5b68a4e4a2772c2a206180a1/t/5bfc60244d7a9c171a603c2b/1543266343802/EOLSSComplexityRising.pdf.

130. *See* Rhoen & Feng, *supra* note 50, at 140 (referring to disorder or non-predictability).

131. *See id.* at 142–43 (referring to this phenomenon of communication and association among human beings as "emergence" in complex systems).

132. *See id.* at 142 (providing similar examples of how humans tend to group together).

133. *See id*. at 140–41 (referring to the concept of datafication).

134. *See id.* at 143 (applying the concept of emergence to artificial intelligence).

membership in OUTreach should be penalized, as it leads to short retention and higher turnover.[135]

c. Data Integrity

Finally, a lack of data integrity could lead Alchemist Algorithm's machine learning program to detect patterns that do not actually exist; that is, the data itself may be inaccurate.[136] For example, if the employees do not respect the system, they may click randomly with inaccurate or blank answers in an effort to get through a particular task.[137] Because the computer program will assume such data is accurate, it will fail to draw any proper connections.[138] Data integrity can also be implicated if the human resources database draws information from other sources that have a similar lack of data integrity.[139] This phenomenon reflects an old adage in the data world: "garbage in, garbage out."[140] If it is clear that the data is inaccurate or blank, then the data processor will need to carefully consider whether that data needs to be deleted or assigned a generic value, such as the mean or mode for that data element.

3. Finalizing and Applying the Model

After it has finished processing the EMPLOYEE_EVAL dataset, Alchemist Algorithm's machine learning program returns predictors for the model that include previous work experience, a particular ranking of career goals, and a high college grade point average.[141] While not bearing directly on the job requirement, those predictors are included in the algorithm because they each have a high correlation with the target variable.[142] As part of the application process, applicants are required to provide this information, which is then stored in the APPLICANT_INFO dataset. The model weighs each data point according to its predictive capacity.[143] At this point, the algorithmic model has been formed. Now it needs to be applied.

Alchemist Algorithms designed the model to target certain applicants for interviews out of thousands who applied, all of whom are recorded in the APPLICANT_INFO dataset. The model processes the information in the

135. *See id.* at 142–43 (referring to an application of the concept of emergence to the hypothetical scenario).

136. *See* Barocas & Selbst, *supra* note 34, at 683–84 (explaining that insofar as inaccurate data is utilized to construct a model, the model will produce unreliable predictive results).

137. *See id.* (discussing why stored data should not be taken as ground truth).

138. *See* Maladkar, *supra* note 36 (discussing methods of dealing with inaccurate, deleted, or blank data).

139. *See* Barocas & Selbst, *supra* note 34, at 684 (indicating that an organization may rely upon bad information collected externally).

140. *Id.* at 683.

141. *See* Iliya Valchanov, *Machine Learning: An Overview*, ORACLE (Jan. 25, 2018), https://www.datascience.com/blog/machine-learning-overview (referring to this aspect as the objective function).

142. *See* Barocas & Selbst, *supra* note 34, at 677 (discussing the implications of improperly defining the target variable).

143. *See* SHALEV-SHWARTZ & BEN-DAVID, *supra* note 112, at 12, 14 (indicating that each classifier is assigned a probability relating to its ability to predict the correct label for a randomly-selected data point).

APPLICANT_INFO dataset and sorts the applicants according to the likelihood that they will become successful employees.[144] The model has now performed its function—sorting the employees according to the likelihood that they will become high performers.[145] Alchemist Algorithms then sends this information back to the recruiting department at Process Perfectionist, highlighting the top 100 candidates. While not binding on the recruiters, those results will likely have a substantial, and even determinative, impact on who will receive interviews.[146] As this example illustrates, algorithmic models are not subject to the same unconscious bias as humans, but they may inherit biases from the composition of the target variable or from the underlying training data they process.[147]

C. Prospects of Judicial Resolution

The precise way in which a machine learning program proves deficient will dictate which Title VII theory is implicated, which can be illustrated by extending the Process Perfectionist hypothetical. After being applied for a year, Process Perfectionist's human resources department finds that the model includes female applicants on the top 100 list at a rate less than 70% of the rate for men. Based on the EEOC's four-fifths rule, which is triggered when the selection rate for a protected class is less than four-fifths of the group with the highest selection rate,[148] these statistics could support a prima facie claim for disparate impact employment discrimination.[149] As discussed previously, this disparity could result from improperly setting the target variable,[150] from using data rendered deficient on the basis of omitted variable bias,[151] or from failing to account for emergent properties,[152] all of which seem to implicate a disparate impact theory.

While it is most likely that Process Perfectionist would be sued on a disparate impact theory, the company could also potentially face a disparate treatment claim. As in the example with the female engineer, if a data scientist does not provide oversight or proper guardrails, the model could encode the predictive capacity of a specific college degree into a gender variable.[153] So long as the model includes data on protected status, data scientists should notice if the model discounts an applicant

144. *See* Valchanov, *supra* note 141 (discussing similar applications in which machine learning is used to predict performance).

145. *Id.*

146. *See More Than Half of HR Managers Say Artificial Intelligence Will Become a Regular Part of HR in Next 5 Years*, *supra* note 24 (providing an overview of the increased use of artificial intelligence in human resource departments).

147. *See* Barocas & Selbst, *supra* note 34, at 671 (discussing broadly the ways in which an algorithm can take on the biases of the data it processes).

148. *See* 29 C.F.R. § 1607.4(D) (2015) ("A selection rate for any race, sex, or ethnic group which is less than four-fifths . . . of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact.").

149. *See, e.g.*, Connecticut v. Teal, 457 U.S. 440, 443 n.4 (showing an example of statistics used to prove disparate impact).

150. See *supra* Part II.B.1 for a discussion of target variable definitions.

151. See *supra* Part II.B.2.a for a discussion of omitted variable bias.

152. See *supra* Part II.B.2.b for a discussion of emergent properties.

153. See *supra* Part II.B.2.a for a discussion on the effect of omitted variables.

explicitly on the basis of a protected trait.[154] If Process Perfectionist's model includes female applicants at only 70% the rate of male applicants, then the company could defend itself only by showing that the disparity was the product of a business necessity.[155] In order to succeed on this defense, the defendant would need to prove (i) the existence of a statistically significant correlation between the predictor and the target variable, and (ii) that each aspect of the target variable is truly job-related and necessary to job performance.[156] This may be feasible if the job requires a particular certificate or degree that males are more likely to possess, such as an engineering degree.[157] But even if the defendant were to successfully establish the existence of a business necessity, the defendant would still have an obligation to implement a less discriminatory alternative model if the plaintiff could point one out.[158]

If Process Perfectionist recognized that the model caused disparate results and attempted to rectify this impact by hiring females at a higher rate after interviews instead of fixing the model, it could still be liable for applying the model.[159] The Supreme Court has found that fixing disparities in the bottom line does not cure an otherwise discriminatory classification.[160] In her article *Data-Driven Discrimination at Work*, Pauline Kim bolsters this line of reasoning and suggests imposing liability for discriminatory classification, regardless of how it actually affects the final decision-making process.[161] While Alchemist Algorithms' machine learning model presents complex issues, the existing Title VII framework can generally resolve them.

III. DISCRIMINATION IN DELIVERY AND DESIGN

Algorithms inform human decision-makers by aggregating data, revealing trends, and predicting outcomes.[162] After Alchemist Algorithms creates a ranked list

154. *See Omitted Variable Bias: A Comprehensive Econometrics Review*, *supra* note 121 (indicating the importance of checking data analysis results for the possibility that the resultant model is subject to omitted variable bias).

155. *See* Albemarle Paper Co. v. Moody, 422 U.S. 405, 434 (1975) (discussing whether tests could be considered genuine business needs by taking into account prospects for promotability).

156. *See id.* (finding that prospects for promotion were not sufficiently job-related to be deemed a business necessity).

157. *See* Brian L. Yoder, *Engineering by the Numbers*, *in* 2017 ASEE PROFILES OF ENGINEERING AND ENGINEERING TECHNOLOGY COLLEGES 15 (American Society for Engineering Education 2017) (providing statistics showing that more male students earn engineering degrees than do female students).

158. *See* ESTREICHER ET AL., *supra* note 81, at 156 (discussing how the 1991 amendments impacted the less discriminatory alternative aspect of litigation).

159. *See* Connecticut v. Teal, 457 U.S. 440, 449–451 (1982) (holding that the employer's genuine attempt to rectify discriminatory effects in the bottom line did not justify a test requirement that had a disparate impact effect).

160. *Id.*

161. *See* Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. 857, 909 (2017) (advocating for liability for classification bias based on close reading of Title VII's text).

162. *See, e.g.*, IDEAL, https://ideal.com (last visited Sept. 20, 2019) (advertising how artificial

of candidates, it could choose several ways to deliver this information.[163] It could use a two-column static report with only the name and the ranking of each applicant. On the other hand, Alchemist Algorithms could deliver an interactive dashboard[164] that allows the recruiter to manipulate the data in various ways, such as filtering by race or gender, or sorting by age.[165] The latter option presents disparate treatment issues.[166]

A. Manipulating Results and Disparate Treatment

Through the delivery and design of their algorithm's output, companies that specialize in machine learning models for recruitment and hiring may enable or encourage discrimination.[167] When a third-party software or machine learning company performs the data processing service, it could face liability in two ways.[168] First, the data processing itself could be flawed, as discussed in Part II of this Comment.[169] Second, even if the data processing is compliant with Title VII, the presentation method could encourage or contribute to the employer's discrimination.[170] Unlike data processing that suffers from omitted variable bias or insufficient data, this type of conduct does not stem from data processing, but rather from how one manipulates and presents the results.[171]

---

intelligence allows recruiters to analyze and optimize data); *About Us*, NIELSEN, https://www.nielsen.com/us/en/about-us/ (last visited Sept. 20, 2019) ("Nielsen is a global measurement and data analytics company that provides the most complete and trusted view available of consumers and markets worldwide.").

163. Krist Wongsuphasawat, *The Challenges of Dataset for Data Visualization Tools*, MEDIUM (Mar. 23, 2017), https://medium.com/@kristw/the-challenges-of-dynamic-data-in-data-visualization-tools-97f5587cbac5 (discussing different visualization choices and how underlying data should influence such choices).

164. For visual examples of interactive dashboards, see *Experience Interactive Dashboards*, IDASHBOARDS, https://www.idashboards.com/dashboard-examples/ (last visited Oct. 1, 2019).

165. *See* Wongsuphasawat, *supra* note 163 (discussing how dynamic datasets are more often used with dashboards and other analytic tools). There are other options as well, such as interactive reports that allow for filtering; *see, e.g.*, Reza Rad, *Dashboard vs Report; Differences at a Glance – Power BI*, RADACAD (Oct. 10, 2016), http://radacad.com/dashboard-vs-report-when-where-why-which-to-use (explaining that data reports are fully interactive and can be filtered by criteria).

166. Int'l Union v. Johnson Controls, 499 U.S. 187, 198–200 (1991).

167. See *infra* Part III.B for a discussion of what facts are likely to influence a court's ruling.

168. *See, e.g.*, Compl. at 1–2, Spees v. Facebook, Inc. (EEOC Sept. 18, 2018) (processing personal data in the process of targeting advertisements), https://www.aclu.org/legal-document/facebook-eeoc-complaint-charge-discrimination*; see also* Amit Datt, Michael Carl Tschantz & Anupam Datta, *Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination,* 1 PROCEEDINGS ON PRIVACY ENHANCING TECH. 92 (Apr. 1, 2015), http://www.andrew.cmu.edu/user/danupam/dtd-pets15.pdf (finding that Google's online advertising system advertises high-paying jobs to women significantly less than to men).

169. See *supra* Part II for a discussion of the ways in which data processing can lead to a discriminatory model.

170. See *infra* Part III.B for a discussion of the potential for secondary liability.

171. Here, the term "data processing" refers to feeding the data through a model for the purpose of gaining a predictive ranking or scoring of the likelihood of being a high performer.

Virtual recruiters like Ideal,[172] Avrio,[173] and HireVue[174] offer data processing and data visualization to better inform recruiters for large employers. Good visualization and interactive displays are hallmarks of successful data presentation,[175] and they provide recruiters with an understanding of the diversity available in their applicant pools.[176]

Machine learning companies market their products and services as giving employers the tools to pick the perfect diverse candidate.[177] As a hypothetical example, a recruiter at Process Perfectionist might filter down a full list of candidates with a few clicks on the dashboard to those who (i) were ranked in the top 20% by the model, (ii) are female, and (iii) belong to a certain race. The dashboard would then return the intersection of these three lists, like the intersection of a Venn diagram with three circles.[178] If a recruiter uses such a list to schedule interviews and hire candidates, then this would implicate disparate treatment doctrine.[179]

Title VII may be ill-equipped to handle this type of issue because of its limited application to specific actors.[180] While Title VII does not allow employment agencies to refer or classify in a discriminatory fashion, it has not traditionally provided for secondary or contributory liability.[181] In *Scaglione v. Chappaqua Central School District*, for example, the court fleshed out the conduct that can lead a company to be held liable as an employment agency.[182] The court held that "one that 'regularly' undertakes to procure employees or employment opportunities . . . [and is] engaged to a significant degree in that activity as their profession or

---

172. *See, e.g.*, *Intelligent Rediscovery*, IDEAL, https://ideal.com/product/sourcing/rediscovery (last visited Sep. 20, 2019) (discussing the candidate rediscovery software).

173. *See, e.g.*, *The Right Candidate. The Right Job. The Right Time.*, AVRIO, https://www.goavrio.com/product/artificial/intelligence/platform (explaining how they apply artificial intelligence and machine learning to find the best candidate match for open jobs and providing screenshots of curated data presentation).

174. *See, e.g.*, HIREVUE, https://www.hirevue.com/ (discussing on its website how it uses artificial intelligence to streamline the hiring process and "gamifies" pre-employment assessments).

175. *See, e.g.*, *Experience Interactive Dashboards*, IDASHBOARDS, https://www.idashboards.com/dashboard-examples/ (last visited Sept. 17, 2019) (displaying examples of visually attractive dashboard GUIs).

176. *See The Right Candidate. The Right Job. The Right Time.*, *supra* note 173 (demonstrating how virtual recruiters showcase the available talent in an applicant pool through data visualization).

177. *See, e.g.*, *Increase Diversity and Mitigate Bias Through Better Decisions*, HIREVUE, https://www.hirevue.com/why-hirevue/foster-diversity (advertising how its software provides a wider array of candidates while reducing bias).

178. *See* Susan Harkins, *How to Create an Effective, User-Friendly Slicer in Excel*, TECHREPUBLIC (Jul. 18, 2016, 2:41 PM), https://www.techrepublic.com/article/how-to-create-an-effective-user-friendly-slicer-in-excel/ (discussing a similar method of filtering data).

179. *See* Int'l Union v. Johnson Controls, 499 U.S. 187, 198–207 (1991) (discussing that intent is the only requirement, and it is satisfied even by intent grounded in good faith).

180. *See* 42 U.S.C. § 2000e (2012) (applying to defined actors under defined circumstances).

181. *See* Greenfield v. Field Enterprises, Inc., 1972 U.S. Dist. LEXIS 15304, 15–17 (N.D. Ill. 1972) (holding that a newspaper was not liable under Title VII for running job advertisements in separate "male" and "female" columns).

182. Scaglione v. Chappaqua Cent. Sch. Dist., 209 F. Supp. 2d 311, 316 (S.D.N.Y. 2002) (citing Brush v. San Francisco Newspaper Printing Co., 315 F. Supp. 577 (N.D. Cal. 1970)).

business" qualifies as an employment agency and may not interfere impermissibly with a plaintiff's employment opportunities.[183] While such a touchstone may exclude businesses like newspapers, it almost certainly covers companies that specialize in recruiting and hiring.[184]

Although the employer has the ultimate say on whether to hire, courts could hold machine learning companies liable when their conduct induces or encourages employer discrimination.[185] As discussed in Part II, Title VII effectively resolves questions of data processing and places liability on the companies with the resources to correct flawed processing.[186] It is less clear whether an employment agency can be held liable for encouraging discrimination, but pertinent court cases suggest that the answer is yes.[187] Although regulation in this area could offer clarity, the fact-specific nature of this inquiry suggests that a solution will need to be crafted case-by-case. By engaging in thorough fact-specific analysis, the courts should attempt to draw a line between *enabling* discrimination and *encouraging* discrimination, and should not tolerate the latter.[188]

In summary, machine learning programs can reveal hidden biases and inform decision-makers, who then have an opportunity to rectify these risks.[189] But even when the data processing has appropriate safeguards, the delivery method can still encourage discrimination. When a specific design induces discriminatory conduct, the company that deployed that algorithm and profits from it should face liability.[190] Unfortunately, delineating conduct that *informs or enables* from conduct that *induces or contributes* is a difficult task that lacks on-point precedent. The following section addresses this inquiry.

B. Purpose, Design, and Precautions Should Dictate Secondary Liability

Through fact-intensive inquiry, this section explores the point at which conduct shifts from *enabling or informing* to *encouraging or contributing* to impermissible discrimination. Conduct that encourages discrimination should be subject to secondary liability as part of a judicially-crafted solution aimed at spreading liability to offending data processors. Although machine learning discrimination cases have not come before the courts, similar cases involving data processors, like those

183. *Id.*

184. *Id.*

185. *See id.* at 318 (holding that allowing an employer to interfere with the employment opportunities of an individual with another employer would be condoning the prohibited conduct).

186. See *supra* Part II for a discussion of discrimination in model formation and application.

187. *See Scaglione*, 209 F. Supp. 2d at 317 (noting that, although case law on the liability of employment agencies is sparse, personnel boards are generally subject to liability).

188. Such an approach is consistent with the federal policy meant to protect innovative internet companies from liability for the conduct of bad actors. *See* 47 U.S.C. § 230(c) (2018) (stating that a provider of an interactive computer service should not be considered a publisher).

189. See *supra* Part II for a discussion of machine learning as a means of battling subconscious bias.

190. *See, e.g.*, *Scaglione*, 209 F. Supp. 2d at 319 (holding that the county's personnel office could be held liable even though the school district was the actual employer).

involving the websites Craigslist,[191] Facebook,[192] and Roommates.com,[193] provide important clues as to how the courts are likely to resolve such issues. In determining whether a data processor should face liability, courts appear to consider (i) whether the primary or exclusive purpose of the delivery platform was to allow users the ability to discriminate unlawfully;[194] (ii) whether the design choices and user interface encourage the employer to make a discriminatory decision;[195] and (iii) whether the company has taken remedial measures to diminish a perceived risk of discrimination.[196]

In 2008, the Seventh Circuit held that Craigslist was not responsible for discriminatory housing ads.[197] Writing for the majority, Judge Easterbrook stated that attorneys "cannot sue the messenger just because the message reveals a third party's plan to engage in unlawful discrimination."[198] The court found that, unlike a newspaper, Craigslist was not the publisher and therefore did not have the responsibility, or even the capacity, to screen the millions of advertisements posted on its website.[199] The Communications Decency Act (CDA), which codified a policy of protecting innovators and insulating them from the actions of bad actors, undergirded the court's decision.[200]

Today, the leading advertisement platforms are far more engaged with their advertisers than Craigslist.[201] In September 2018, the ACLU filed a complaint against Facebook for discrimination in employment advertising on the basis of sex.[202] Facebook requires users to answer several questions when they create an

---

191. *See generally* Chi. Lawyers' Comm. for Civ. Rights Under Law, Inc. v. Craigslist, Inc., 519 F.3d 666 (7th Cir. 2008).

192. Facebook EEOC Compl. at 1–2.

193. *See generally* Fair Hous. Council of S.F. Valley v. Roommates.com, 489 F.3d 921 (9th Cir. 2007) (partially overturned by Fair Hous. Council of S.F. Valley v. Roommates.com, 521 F.3d 1157, 1167 (9th Cir. 2008)).

194. *See id.* at 928 (discussing a hypothetical website called www.harassthem.com).

195. *See* Facebook EEOC Compl. at 13–14 (detailing how Facebook enables job advertisers to exclude women and non-male prospective applicants).

196. *See id*. at 10 (explaining how Facebook has taken steps to stop discrimination against other protected classes but continues to allow employers to exclude female and non-male applicants).

197. Chi. Lawyers' Comm. for Civ. Rights Under Law, Inc. v. Craigslist, Inc., 519 F.3d 666 (7th Cir. 2008); *see also* Amanda Beck, *Craigslist Not Liable for Illegal Ads, Court Says*, REUTERS (Mar. 17, 2008, 8:39 AM), https://www.reuters.com/article/us-craigslist-discrimination/craigslist-not-liable-for-illegal-ads-court-says-idUSN1547054820080317 (providing a general overview of the case).

198. Craigslist, Inc., 519 F.3d at 672.

199. *Id.* at 668.

200. *Id.* at 669; 47 U.S.C. § 230(c)(1) (2018).

201. *See* Craigslist, Inc., 519 F.3d at 668 ("[Craigslist is] in some respects like the classified pages of newspapers, but in others [it] operate[s] like common carriers such as telephone services."); *see, e.g.*, TWITTER BUSINESS, https://business.twitter.com/en.html (last visited Oct. 1, 2019)*; see also* FACEBOOK BUSINESS MANAGER, https://business.facebook.com/(last visited Oct. 1, 2019).

202. *See* Facebook EEOC Compl. at 9 (asserting that Facebook enables advertisers to narrow the audience of their ads).

account.[203] Among these questions, Facebook requires users to list their gender, which then becomes a potential criterion for targeted advertising.[204] The complaint asserts that "through its advertising platform, Facebook enables, encourages, and assists employers[205] to target advertisements and recruitment based on the users' gender, by allowing advertisers to select either "All," "Male," or "Female" users to receive the ad."[206] Two of the three available options likely implicate disparate treatment doctrine.[207] Facebook recently settled, reaching an agreement in which it eliminated the ability of advertisers to exclude users based on gender, age, or other characteristics.[208]

The Facebook Lookalike feature further inserts Facebook into the decision-making process.[209] This feature allows advertisers to target users who share demographic characteristics with the consumers of the advertiser's product.[210] Within the recruiting and hiring area, this type of feature can perpetuate existing biases among Facebook's employer advertisers.[211]

Facebook's platform can also be viewed in a positive light, as its targeted criteria enable advertisers to make better-informed decisions and aim their advertisements at particular applicants.[212] Facebook also allows employers to target applicants they otherwise might not be able to reach, including those belonging to minority groups.[213] Thus, the same conduct that allows some bad actors to discriminate can allow good actors to recognize when their practices prove discriminatory, and then rectify those practices.[214] Notwithstanding the recent settlement, understanding how the court would have ruled requires taking a holistic account of Facebook's advertising platform. First, the primary purpose of Facebook's platform is not discrimination, but rather precision in advertising.[215]

---

203. *Id.* at 3–4.

204. *Id.*

205. In addition to the case involving the ACLU and the employees, the EEOC sent letters to seven employers informing them of their violation of federal law through the use of Facebook job advertising. Kaya Yurieff, *Employers Illegally Used Facebook Ads to Exclude Women and Older Workers, Says EEOC*, CNN BUS., https://www.cnn.com/2019/09/25/tech/facebook-ads-age-discrimination/index.html (Sept. 25, 2019, 5:40 PM).

206. Facebook EEOC Compl. at 1.

207. *See id.* at 1 (explaining the effects of limiting the targeted audience).

208. Galen Sherwin & Esha Bhandari, *Facebook Settles Civil Rights Cases by Making Sweeping Changes to Its Online Ad Platform*, AMERICAN C.L. UNION (Mar. 19, 2019, 2:00 PM), https://www.aclu.org/blog/womens-rights/womens-rights-workplace/facebook-settles-civil-rights-cases-making-sweeping.

209. *See* Facebook EEOC Compl. at 3 (discussing the functionality of the Lookalike feature).

210. *Id.*

211. *See id.* (describing the feature as allowing advertisers to target people like their customers).

212. *See generally* Larry Kim, *5 Ridiculously Powerful Facebook Ad Targeting Strategies*, WORDSTREAM (Aug. 22, 2019), https://www.wordstream.com/blog/ws/2015/01/28/facebook-ad-targeting.

213. *See id.* (discussing the available levels of granularity and how they may be used to target very specific populations).

214. *See id.* (discussing the raw power of the tools to facilitate precision in advertising).

215. *See* Sapna Maheshwari, *Facebook's Ad Profiles Remain Mystery to Many, Survey*

Because this purpose is as consistent with doing good as doing bad, it weighs against an imposition of liability.[216] Second, the way in which Facebook presents a choice of gender without prompting suggests that its design allows and even encourages discrimination.[217] Third, because Facebook has been put on notice by the ACLU, the court would have also taken into account whether Facebook has taken any remedial actions.[218]

While it remains unclear how a court would rule in the machine learning context, it is apparent that the solutions to these problems cannot be empirically formed and require case-by-case resolution. Certainly a line exists separating acceptable from impermissible conduct, and a prudent company will be careful not to go anywhere near that line.[219] Despite this uncertainty, it seems clear that drawing that line depends on whether: (i) the primary or sole purpose of the features is to discriminate; (ii) the features are designed so as to induce improper discrimination; and (iii) the data processor has taken precautions to address any risks of which it is aware.[220]

IV. PROACTIVE PREVENTION

The United States should not follow Europe in adopting a broad and transformative regulation, because the existing anti-discrimination legal framework stands well equipped to handle issues arising from AI in hiring and recruitment. Regardless, the feasibility of preventing algorithmic discrimination before it occurs, coupled with the potential of eliminating subconscious bias, offers compelling justifications for passing an EEOC regulation with a sense of urgency.[221] Effective regulation is currently feasible for three reasons. First, more companies regularly use sophisticated analytics, which means the data is available.[222] Second, the courts and the EEOC have put forth workable statistical guidelines that, if granted greater

---

Reveals, N.Y. TIMES (Jan. 16, 2019), https://www.nytimes.com/2019/01/16/business/media/facebook-advertising-transparency-users.html ("Targeted advertising is the core of Facebook's business, which brings in more than $40 billion in revenue each year.").

216. *See* Fair Hous. Council of S.F. Valley v. Roommates.com, 489 F.3d 921, 928 (9th Cir. 2007) (analogizing to a hypothetical website called www.harassthem.com).

217. *See id.* (discussing basic design decisions that seem to induce discrimination); *see also* Facebook EEOC Compl. at 12 –14 (discussing the design of the gender targeting option).

218. *See* Facebook EEOC Compl. at 10 (alleging that Facebook had learned in 2017 of specific instances of gender discrimination by employment advertisers on its network).

219. *See* Scaglione v. Chappaqua Cent. Sch. Dist., 209 F. Supp.2d 311, 316 (S.D.N.Y. 2002) (citing Brush v. San Francisco Newspaper Printing Co., 315 F. Supp. 577 (N.D. Cal. 1970)) (holding that a county's personnel office, though not the actual employer, could be subject to liability).

220. *See* Facebook EEOC Compl. at 10 (describing how Facebook failed to follow each one of these steps).

221. *See More Than Half of HR Managers Say Artificial Intelligence Will Become a Regular Part of HR in Next 5 Years*, *supra* note 24 (discussing the increased prevalence of artificial intelligence in recruiting and hiring).

222. *See generally* Bernard Marr, *Why Data Is HR's Most Important Asset*, FORBES (Apr. 13, 2018), https://www.forbes.com/sites/bernardmarr/2018/04/13/why-data-is-hrs-most-important-asset/#324000046b0f.

deference by the courts,[223] can serve as bright-line regulations in this sphere.[224] Third, the EEOC has demonstrated a regulatory aptitude for dealing with statistical analysis.[225]

Despite the encouragement among several scholars for broad adoption of the GDPR,[226] my research suggests that, at least when dealing with hiring in the private sector, federal regulators should only borrow specific pieces of the regulation that allow for proactive prevention to fill the existing gaps in federal law.[227] Overregulating data processors in recruiting and hiring would not only deprive employers of an efficient tool for targeting applicants with the best skills, but would also unnecessarily inhibit an important means of fostering greater diversity and inclusion in the workplace.[228] An overly broad adoption of the GDPR would therefore prove counterproductive in the struggle for equitable outcomes in the American workplace.

A. Defining the Contours

An effective regulation should spell out the specific actors and conduct that fall within its ambit.[229] Regulators can borrow the GDPR's method for classifying companies based on the types of activities they perform.[230] The GDPR imposes regulatory obligations based on whether a company is a data controller,[231] data processor,[232] or both. It also requires companies to fulfill these obligations to data subjects,[233] those individuals whose data is processed.

Article 4 of the GDPR offers model definitions for these different types of

---

223. El v. Southeastern Pa. Transp. Auth., 479 F.3d 232, 244 (3d. Cir. 2007) (discussing how the degree of deference given by the courts to EEOC Guidelines has eroded).

224. *See* Hazelwood Sch. Dist. v. United States, 433 U.S. 299, 311 n.17 (providing an overview of the legal significance of standard deviations as predictors of fluctuations)*; see also* 29 C.F.R. § 1607.4D (2015) (providing the four-fifths rule as a guideline for disparate impact analysis).

225. *See id.* § 1607.4D (demonstrating that while the four-fifths rule is not a legal boundary, it serves as an important guideline in determining whether to impose liability).

226. *See, e.g.*, *Copy That: America Should Borrow from Europe's Data Privacy Law*, *supra* note 44 (arguing that the United States should pick and choose certain GDPR provisions to adapt and implement)*;see also* Rossi, *supra* note 44 (providing a more cautious view towards adopting the GDPR).

227. See *supra* Part I.C for an analysis of whether courts are equipped to handle new issues arising from machine learning applications.

228. *See* Lee, *supra* note 18, at 482 (discussing unconscious bias)*; see also* Berman, *supra* note 46 (discussing the potential drawbacks of adopting the GDPR). In particular, substantive rights such as the right to an explanation and the right to be forgotten would require cumbersome regulatory frameworks in an area in which people today do not have a right to an explanation. Margot E. Kaminski, *The Right to an Explanation Explained*, 31 BERKELEY TECH. L. J. 1, 4 (2018).

229. *See, e.g.*, GDPR, *supra* note 47, at 33–34, art. 4 (setting out the different types of data entities).

230. *Id.*

231. *Id.* at 33, art. 4(7).

232. *Id.* at 33, art. 4(8).

233. *Id.* at 33, art. 4(1).

actors.[234] The Article defines processing as the application of computational operations to personal data "such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, [and] use."[235] This definition is broader than that used in Part II of this Comment and would not only cover model formation and application, but would also likely cover post-processing manipulation, such as filtering and sorting.[236] Considering the employer and the machine learning company, collaborative data processors could simplify analysis when one data processor encourages the discriminatory conduct of another.[237]

Under the GDPR's scheme, the data controller "determines the purposes and means of the processing of personal data," and is responsible for placing restrictions on the use of personal data that it shares.[238] Based on the example in this Comment, Process Perfectionist was a data controller, because it stored data on applicants and employees. It therefore owed certain duties to protect private information and draw up contracts with third parties if it shared personal data.[239] Alchemy Algorithms, the company that received that data for model formation and application, played the role of the data processor.[240] Before it could receive data, the company should have been required to sign a data processing agreement, which would specify appropriate uses of the personal data.[241]

The GDPR obliges the controller and processor to execute a written processing agreement that imposes strict contractual requirements with respect to the processor's operations.[242] Such an agreement should include information regarding how to assess liability in the event of a data breach, as well as how to apportion liability in the event of algorithmic discrimination.[243] Requiring a similar agreement in the United States would force the employer, or data controller, to consider who is ultimately responsible for liability should illegal discrimination take place.

B. Required Audits and Proactive Monitoring

The proactive monitoring provisions of the GDPR can serve as a model for regulating machine learning in hiring and recruitment.[244] Among the proactive monitoring provisions are requirements that data controllers retain their records, specify an agreement before sharing data, and submit their processing procedures to

234. *Id.* at 33–34, art. 4.

235. GDPR, *supra* note 47, at 33, art. 4(2).

236. *Id.* See also *supra* Part II for a discussion on how to adapt the current American legal framework on discrimination to the rise of data-driven hiring practices.

237. See *supra* Part III for a discussion of potential liability when a machine learning company encourages an employer's discriminatory conduct.

238. GDPR, *supra* note 47, at 33, art. 4(7).

239. *Id.*

240. *See id.* at 33, art. 4(2), 4(8) (describing processing as the application of computational methods to personal data).

241. *Id.* at 36, art. 6(1)(a).

242. *Id.* at 49, art. 28(3).

243. *See id.* (providing a basic framework for an agreement that could be applied in the United States).

244. GDPR, *supra* note 47, at 49–50, art. 28. *Id.* at 53–54, art. 35.

legally required audits when data processing can result in important legal effects.[245]

Article 35 of the GDPR provides for data protection impact assessments.[246] The data protection impact assessment is the primary proactive tool for detecting and preventing discrimination before it begins.[247] Recital 84 of the GDPR requires that a data processing impact assessment take place whenever the processing of data is "likely to impact the legal rights of persons."[248] In the United States, such a requirement would make the regulation applicable to areas governed by existing anti-discrimination laws.[249]

C. Workable Guideposts

A data protection impact assessment is only valuable if a regulatory agency has established useful benchmarks that, if met, can exempt data processors from litigation.[250] These proactive audits should adopt as benchmarks the EEOC's statistical guidelines and regulations,[251] along with some additional measures of fairness.[252] To ensure that machine learning models are unbiased, data protection impact assessments should require that the model process data with all protected characteristic information.[253] Only then will an auditor be able to discern whether a model is fair and unbiased.

The EEOC or other regulatory agency should perform three primary checks as part of proactive assessments. First, the auditor should check to determine whether a sufficient amount of worker proficiency is explained by the model.[254] This statistical measurement is captured by the R^2 value.[255] Second, the auditor should check the impact on each protected group and work to make the model consistent

---

245. *Id.* at 53–54, art. 35.

246. *Id.*

247. *See* Article 29 Data Protection Working Party, *Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing Is "Likely to Result in a High Risk" for the Purposes of Regulation 2016/679*, at 4 (Apr. 4, 2017), http://ec.europa.eu/newsroom/document.cfm?doc_id=47711 ("In other words, a DPIA is a process for building and demonstrating compliance.").

248. GDPR, *supra* note 47, at Recital 84.

249. *See Laws Enforced by the EEOC*, U.S. EQUAL EMP. OPPORTUNITY COMMISSION, https://www.eeoc.gov/laws/statutes/ (last visited Sept. 17, 2019) (providing a full list of anti-discrimination statutes in the United States).

250. GDPR, *supra* note 47, at 53–54, art. 35 (discussing broad guidelines but not specifying particular criteria).

251. *See* 29 C.F.R. § 1607.4D (2015) (providing the EEOC's four-fifths guideline); Hazelwood Sch. Dist. v. United States, 433 U.S. 299, 311 n.17 (1977) (providing an overview of the legal significance of standard deviations as predictors of fluctuations).

252. *See* Campbell, *supra* note 114, at 1303 (discussing the need for courts to place a greater emphasis on how well the statistical model fits the reality of employers' conduct).

253. *See* GDPR, *supra* note 47, at 38–39, art. 9 (placing restrictions on the processing of sensitive data that would likely inhibit the use of algorithms to foster more inclusive workplaces).

254. *See* Campbell, *supra* note 114, at 1303 (explaining the importance of the good fit of the model).

255. *Id.* at 1310.

with the four-fifths guideline put forth by the EEOC.[256] Third, the regulatory agency should verify that a reasonably diligent investigation was performed in relation to data integrity.[257] In addition to requiring a data processing agreement, these checks should form the rules of the road governing machine learning in hiring and recruitment.

V. CONCLUSION: MODERATE SOLUTIONS IN THE PRIVATE SECTOR

As a crowning achievement of the Civil Rights Movement, Title VII was passed to "achieve equality of employment opportunities and remove barriers that have operated in the past to favor an identifiable group of white employees."[258] In a society in which discrimination is systemic, occurs over the course of generations, and impacts education, employers are not asked to fully rectify an unfair system, but to think carefully before adopting neutral policies that can perpetuate discriminatory patterns. Data analysis has always served as a useful vehicle for detecting and rectifying discriminatory patterns,[259] and the newest generation of analysis, AI, presents a new opportunity to detect and eliminate unconscious bias from the hiring process.[260] Viewed in this light, the equitable interests underlying Title VII do not require an overhaul of federal regulations, but a patching up of existing doctrine that encourages a principled application of AI that eliminates unconscious bias from the hiring process without adding algorithmic bias.[261]

Now is the time to get ahead of algorithmic discrimination and establish clear rules of the road on acceptable conduct among data processors.[262] New technologies offer opportunities to improve regulatory efforts while only marginally increasing the burden on employers.[263] Some transferable regulations have existed for some time, such as the four-fifths rule and validity assessments. Other regulations, such as a goodness of fit measurement and data integrity check, have yet to be implemented.[264] By deferring to clear rules of the road, courts can advance the interest of equality in the American workplace.

---

256. 29 C.F.R. § 1607.4D.

257. See *supra* Part II.B.2.c for a technical discussion of the importance of data integrity. Additionally, if the model is based on a particularly large sample size, the regulatory agency should also check for statistical significance. *See* Hazelwood Sch. Dist. v. United States, 433 U.S. 299, 311 n.17 (1977).

258. Griggs v. Duke Power Co., 401 U.S. 424, 429–30 (1971).

259. *See generally* Hazelwood Sch. Dist., 433 U.S. 299; Connecticut v. Teal, 457 U.S. 440 (1982).

260. *See, e.g.*, *Increase Diversity and Mitigate Bias through Better Decisions*, *supra* note 177; *see also* Nicole Morell, *Using Technology to Combat Bias in Hiring*, MIT NEWS (Mar. 23, 2018), http://news.mit.edu/2018/mit-alumna-stephanie-lampkin-using-technology-to-combat-hiring-bias-blendoor-0323.

261. See *supra* Part II for a discussion of whether the courts are well-equipped to handle issues posed by algorithmic discrimination.

262. *Id.*

263. See *supra* Part IV arguing that a limited adoption of certain GDPR provisions in American law would strengthen anti-discrimination laws without unduly burdening employers.

264. See *supra* Part II.B for a discussion of data integrity. *See* Campbell, *supra* note 114, at 1303 (discussing goodness of fit as an important prerequisite to using a model).

While this Comment suggests moderate solutions to issues involving automation in the private hiring sphere, more drastic measures are likely required to protect against threats of algorithmic discrimination in the public sector.[265] Chief among these drawbacks is automation's interference with the need of citizens to understand why and how decisions are made. In a talk at Harvard Law, Virginia Eubanks, a scholar who addressed data discrimination in the public sector and its potential conflicts with due process, summed up the issue succinctly: discretion is not eliminated when a decision-making process is transferred from a human to a machine; it merely switches from an expert's discretion to a programmer's discretion.[266] Discretion that once resided in a case worker with clinical experience or a judge in the sentencing context may be replaced by the discretion of a data scientist with little to no field experience.[267] Our notions of justice cannot be rendered complacent when displaced by automated decision-making. Scholars should always be wary of automating a system that disadvantages the poor and vulnerable, and society must ensure that the fruits of new technology are shared in a fair manner.

Zeroing in on applications of AI in recruiting and hiring, this Comment suggests that new technologies can actually amplify the sense of ownership that individuals feel over decisions. In private sector hiring, new technology can alert decision-makers to disparate impacts before they occur and provide a platform for careful scrutiny of unconscious bias. Establishing rules of the road to govern this innovation is critical but not daunting.[268] Re-tooling Title VII to deal with algorithmic discrimination requires moderate adjustments and borrowing aspects of Europe's proactive approach from the GDPR.

---

265. *See generally* VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2018); FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2016).

266. Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, YouTube (Oct. 25, 2018), https://www.youtube.com/watch?reload=9&v=4xzDgesK2wU.

267. *Id.*

268. See *supra* Part IV for a discussion of the critical aspects of a proper regulation.