COLORADO TECHNOLOGY LAW JOURNAL

University of Colorado Law School
320-L Wolf Law Building, 401 UCB, Boulder, CO 80309-0401
1.303.735.1032 • ctlj@colorado.edu • http://ctlj.colorado.edu

POSTMASTER: Please send address changes to the address above.

The Colorado Technology Law Journal (ISSN 2374-9032) is an association of students sponsored by University of Colorado Law School and the Silicon Flatirons Center for Law, Technology, and Entrepreneurship.

Subscriptions

Issues are published semiannually. Domestic volume subscriptions are available for $50.00. City of Boulder subscribers add $3.74 sales tax. Boulder County subscribers outside the City of Boulder add $2.21 sales tax. Metro Denver subscribers outside of Boulder County add $1.85 sales tax. Colorado subscribers outside of Metro Denver add $1.31 sales tax. International volume subscriptions are available for $55.00. Inquiries concerning ongoing subscriptions or obtaining an individual issue should be directed to ctlj@colorado.edu or by mail to the address above. Back issues in sets, volumes, or single issues may be obtained from:

William S. Hein & Co., Inc.
1285 Main Street, Buffalo, NY 14209
p: 1.716.882.2600 • http://www.wshein.com http://heinonline.org (for back issues in electronic format)

Cite as: 21 COLO. TECH. L.J. __ (2023).

THE DISPARATE IMPACT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

CHRISTINE POLEK* AND SHASTRI SANDY**

The use of artificial intelligence (AI) models to make decisions about actual or potential employees and consumers carries the risk of "disparate impact"—unintentional discrimination. Employment and credit markets are of particular interest to the legal community because both are subject to regulations that prohibit discrimination by classes such as age, color, disability, genetic information, national origin, race, religion, sex, and veteran status. Consequently, federal regulators of these sectors have increased their technical capabilities and indicated an interest in AI-related enforcement to help protect employees and customers.

If left unchecked, discrimination from the increasing use of AI technology to make decisions could result in substantial fines to companies, class-action lawsuits, and AI regulations. This article describes ways to identify, remedy, and reduce the potential for bias in AI applications, with a focus on employment and credit markets. We describe how the use of AI decision models by employers and lenders can lead to bias as well as techniques that could be used by regulators and litigators to identify disparate impact.

* Christine Polek, Ph.D.* Principal, The Brattle Group. Dr. Polek has experience in investigations and litigation involving income, labor, and tax issues, including discrimination in employment and valuation.

** Shastri Sandy, Ph.D.** Senior Associate, The Brattle Group. Dr. Sandy has experience in investigations and litigation involving income, credit markets, and protection of consumers. He has worked with diverse and big data and reviewed complex computer programs and models.

INTRODUCTION

Advancements in computing technology are making it easier for organizations to collect and analyze large quantities of data for use in complex decision models. The most common of these decision models involve artificial intelligence (AI) and machine learning (ML).[1] These models are often a "black box," allowing for limited

1. For ease of exposition, we will use AI to refer to both AI and ML models. As discussed later, ML is a field within AI. *See infra* Section II.

transparency on how decisions are made and bringing the potential for unintended consequences.[2]

Concerns over how these AI-based decision models may affect public interests motivated several United States governmental agencies—including the National Economic Council and the Office of Science and Technology Policy—to study these issues, culminating in the release of the so-called "Big Data Report." The report recognizes that these AI decision models "have the potential to eclipse longstanding civil rights protections in how personal information is used in housing, credit, employment, health, education, and the marketplace."[3] Furthermore, to help mitigate and prevent civil rights infractions, the report recommends the federal government increase its technical capabilities to investigate discrimination protected classes could face due to reliance on these algorithms for decision-making purposes.[4]

Consistent with these recommendations, regulators and advocacy groups have begun to make a more detailed examination of whether AI models could produce decisions that are discriminatory towards protected classes. Greater scrutiny by regulators and advocacy groups increases the risk of legal and regulatory actions for companies who utilize these AI models.

In this paper, we examine bias in automated decision-making and its regulatory oversight in employment and credit markets. First, we define key terms and concepts in the literature, then provide examples of how AI models may contribute to or mitigate disparate impact related to discrimination. We also provide examples of regulatory investigations into the use of these decision models within labor and credit industries. We then describe the quantitative and qualitative analyses typically applied by economists to identify and mitigate such AI-driven unintentional bias. We conclude with suggestions of where biases in AI decision-making may be headed in the future.

---

2. Computer algorithms today can process diverse personal data—such as facial expressions and other characteristics from pictures and videos, preferences from cell phone data, demographic attributes, and shopping patterns—and make inferences about future behaviors. Computer algorithms set out the rules by which to process input data to support decisions such as prioritization, classification, association, and filtering. *See infra* Section II.

3. EXEC. OFF. PRES., BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES III (2014), https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf [https://perma.cc/P9TS-JRNN] (opening letter from John Podesta, Counselor to the President, Penny Pritzker, Secretary of Commerce, Ernest J. Moniz, Secretary of Energy, John Holdren, Director, Office of Science & Technology Policy, and Jeffrey Zients, Director, National Economic Council).

4. *Id.* at 65.

I. BACKGROUND

This section defines key terms that will be essential for further understanding of this article.

A. Artificial Intelligence

AI refers to the design of computational systems that behave intelligently, like humans, and can perform complex tasks. That is, machines that can learn, reason, and act for themselves.[5] Similar to humans, these machines can make their own decisions when they encounter novel situations.[6] For example, one application of AI is natural language processing (NLP), which refers to programming computers to understand and analyze text and spoken words, similar to humans.[7] Applications that use NLP include spam filters and website chatbots.[8]

B. Machine Learning

In ML, a field within AI, computers "program themselves through experience."[9] This capability eliminates the explicit need for intensive programming.[10] Arthur Samuel, an AI pioneer, helped popularize ML. He presented the example of machine learning where "a computer can be programmed so that it will learn to play a better game of checkers than can be played by the person who wrote the program."[11]

5. Sara Brown, *Machine Learning, Explained*, MIT SLOAN SCH. MGMT. (Apr. 21, 2021), https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained [https://perma.cc/T28D-FSAR].

6. Karen Hao, *What Is AI? We Drew You a Flowchart to Work It Out*, MIT TECH. REV. (Nov. 10, 2018), https://www.technologyreview.com/2018/11/10/139137/is-this-ai-we-drew-you-a-flowchart-to-work-it-out/ [https://perma.cc/PS6R-A93E].

7. Brown, *supra* note 5; *see infra* Section III.A (discussing how NLP may be used to mitigate discrimination).

8. David Jurafsky & James H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition 2 (Dec. 30, 2020) (unpublished third edition manuscript) (on file with author).

9. Brown, *supra* note 5 (explaining how these computer programs can adapt when provided with new data. In ML, the program can train itself using an input test dataset. This dataset could take varying formats including tabular inputs, pictures, and videos. Using this input, the program trains itself to recognize patterns and make decisions).

10. Arthur L. Samuel, *Some Studies in Machine Learning Using the Game of Checkers*, 3 IBM J. 210, 211 (1959).

11. *Id.*

C. Disparate Impact and Disparate Treatment

Disparate impact is defined as a test or other tool used for selection that, though appearing neutral, actually has an adverse effect on a particular protected class of individuals.[12] The U.S. Supreme Court has ruled anti-discriminatory statutes—such as Title VII of the Civil Rights Act of 1964, the Age Discrimination in Employment Act of 1967, and Title VIII of the Civil Rights Act of 1968 (also known as the Fair Housing Act)—allow for disparate impact claims.[13]

The Supreme Court addressed the concept of disparate impact in *Griggs v. Duke Power Co.* In its 1971 decision, the Court stated "practices, procedures, or tests neutral on their face, and even neutral in terms of intent, cannot be maintained if they operate to 'freeze' the status quo of prior discriminatory employment practices."[14]

Other statutes and regulations also seek to prohibit discrimination. For example, fair lending laws and regulations such as the Equal Credit Opportunity Act (ECOA) seek to prohibit discrimination in credit transactions.[15] These laws define a policy or practice as having a disparate impact "[w]hen a lender applies a racially or otherwise neutral policy or practice equally to all credit applicants, but the policy or practice disproportionately excludes or burdens certain persons on a prohibited basis."[16]

Disparate impact is sometimes referred to as unintentional discrimination, implying that, while policies and procedures are the "same" for everyone, protected classes are adversely affected in the implementation of those policies and procedures. Intentional discrimination, on the other hand, is described by the term "disparate treatment."[17] An example of disparate treatment is when an organization's hiring practices are designed to deliberately eliminate candidates of a protected class. Both disparate impact and disparate treatment fall under the umbrella of discriminatory practices.

---

12. Definition: Disparate Impact, CORNELL L. SCH.: LEGAL INFO. INST., https://www.law.cornell.edu/wex/disparate_impact [https://perma.cc/C5CC-HEKD] (Nov. 11, 2022); Ricci v. DeStefano, 557 U.S. 557, 577 (2009).

13. Texas Dep't of Hous. and Cmty. Affs. v. Inclusive Cmty. Project, Inc., 576 U.S. 519, 545–46 (2015).

14. Griggs v. Duke Power Co., 401 U.S. 424, 430 (1971).

15. Equal Opportunity Credit Act of 1974, 15 U.S.C. § 1691 et seq. (as amended 2014).

16. FDIC, CONSUMER COMPLIANCE EXAMINATION MANUAL, IV–1.3 (Mar. 2021), https://www.fdic.gov/resources/supervision-and-examinations/consumer-compliance-examination-manual/documents/4/iv-1-1.pdf [https://perma.cc/X8N8-XBHH].

17. Joseph A. Seiner, *Disentangling Disparate Impact and Disparate Treatment: Adapting the Canadian Approach*, 25 YALE L. & POL'Y REV. 95, 96 (2006).

D. Protected Classes

Protected classes are groups of individuals who have a common trait and are legally protected from discrimination because of that trait. The Equal Employment Opportunity Commission (EEOC) is an example of an organization that seeks to enforce federal laws to prevent employment discrimination among protected classes.[18] The EEOC lists eight protected classes regarding employment discrimination.[19]

Acts by Congress also seek to prevent discrimination of protected classes. The Fair Housing Act and Fair Housing Amendment Acts (collectively referred to in this paper as FHA&A) list seven protected classes in housing discrimination,[20] while the ECOA lists six protected classes in credit discrimination.[21] Collectively, the protected classes under these statutes and the Vietnam Era Veterans' Readjustment Assistance Act (VEVRAA) are age, color, disability, familial status, genetic information, national origin, race, religion, sex, and veteran status.[22] Further discussion of these classes can be found in the Appendix.

II. HOW AI CAN CONTRIBUTE TO DISPARATE IMPACT

Discrimination due to faulty data analysis is not a novel concept. For example, Frederick Hoffman, citing statistical analysis, claimed in 1896 that the "gradual extinction [of Black people] is only a question of time."[23] By insinuating Black people had weak

18. *Commissioner Charges and Directed Investigations*, EEOC https://www.eeoc.gov/commissioner-charges-and-directed-investigations [https://perma.cc/HM9F-T4ZA]; *All Charges Data Infographic*, EEOC (2019) https://www.eeoc.gov/sites/default/files/2020-06/OEDA_All%20Charges%20Infographic_052620.pdf [https://perma.cc/B9QH-CHYU] (enforcement actions are primarily initiated through individual complaints and secondarily through separate investigatory tools that originate from the Commission without individual complaints. With regards to individual complaints, the EEOC received 72,675 complaints from individuals alleging discrimination against them in 2019 and initiated a total of 64 Commissioners Charges and Directed Investigations.).

19. *3. Who Is Protected from Employment Discrimination?*, EEOC https://www.eeoc.gov/employers/small-business/3-who-protected-employment-discrimination [https://perma.cc/24SD-MTS7] (stating that the eight classes are: age, color, disability, genetic information, national origin, race, religion, and sex).

20. Fair Housing Act of 1768, 42 U.S.C. § 3604(a)–(e), amended by Fair Housing Amendments Act of 1988, 42 U.S.C. § 3604(a)–(e) (indicating seven classes: color, handicap (disability), familial status, national origin, race, religion, and sex); *see* 42 U.S.C. § 3602.

21. 15 U.S.C. § 1691(a)(1) (indicating six classes: age, color, disability, national origin, race, religion, and sex (including marital status)).

22. *See* Vietnam Era Veterans' Readjustment Assistance Act, Pub. L. No. 92-540, 86 Stat. 1074 (codified as amended at 38 U.S.C. § 4211 et seq.).

23. Frederick L. Hoffman, *The Race Traits and Tendencies of the American Negro*, 11 AM. ECON. ASS'N 1, 329 (1896).

health profiles, Hoffman's work made it difficult for them to obtain insurance.[24] Part of Hoffman's faulty analysis was caused by confusing correlation and causation.[25]

Today, the ways in which AI-driven data analysis can unintentionally lead to discriminatory decisions are mostly well known.[26] Generally, AI bias originates from incomplete training data or reliance on data that is spuriously correlated with protected classes.[27] By training an algorithm on unrepresentative or incomplete training data (such as only using "white-sounding" names), the model learns to automatically filter the subset of data not seen (such as "Black-sounding" names).[28] Reliance on information reflecting historical inequalities can introduce unanticipated bias in decision making; a seemingly neutral variable (such as a zip code) might be so highly correlated with a legally protected class (such as race) that it ends up serving as a proxy for that class.[29]

Unintentional AI-driven bias has been identified in a wide range of applications. Academic studies, for example, have identified biased outcomes from the use of algorithms in healthcare,[30]

24. Megan J. Wolff, *The Myth of the Actuary: Life Insurance and Frederick L. Hoffman's Race Traits and Tendencies of the American Negro*, 121 PUB. HEALTH REP. 84, 86 (2006).

25. *Id.* at 85 ("Hoffman had made the mistake of aggregating his data, thereby obscuring any relationship between cause and effect other than the single commonality of race itself . . . On the grounds of his methods alone, the bulk of Hoffman's claims and conclusions could be easily toppled."). Causation indicates that action A (or a change in variable A) results in outcome B (or results in a change in variable B). Correlation refers to a linear relationship—size and direction—between two variables. Evidence of correlation between two variables does not imply causation between the variables.

26. Eirini Ntoutsi et al., *Bias in Data-Driven Artificial Intelligence Systems—An Introductory Survey*, 10 WILEY INTERDISC. REVS. DATA MINING AND KNOWLEDGE DISCOVERY, May/June 2020, No. e1356, at 1, 2.

27. *See, e.g.*, Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 688 (2016) (explaining the problems relate to (i) how the "target variable" (data the algorithm is trying to identify) and the "class labels" (categories of target) are defined; (ii) labelling the training data; (iii) collecting the training data; (iv) feature selection; and (v) proxies).

28. *See* Janice Gassam Asare, *Are Job Candidates Still Being Penalized for Having 'Ghetto' Names?*, FORBES (Feb. 20, 2020, 11:09 AM), https://www.forbes.com/sites/janicegassam/2020/02/20/are-job-candidates-still-being-penalized-for-having-ghetto-names/?sh=48b9040450ed [https://perma.cc/9M94-EAFH].

29. For a discussion of this point in the insurance context, *see* Daniel Schwarcz, *Ending Public Utility Style Rate Regulation in Insurance*, 35 YALE J. REG. 941, 978 (2018).

30. A 2019 study concluded a healthcare model affecting millions of patients displayed evidence of racial bias, with Black patients receiving less care than equally sick white patients. According to the study, the healthcare algorithm predicted heath costs and used this outcome as a proxy for the need of healthcare. However, holding healthcare need constant, less money is spent on Black patients (lower healthcare costs) than white patients. *See* Ziad Obermeyer et al., *Dissecting Racial Bias in an Algorithm Used to Managed the Health of Populations*, 366 SCI. 447, 447 (2019).

criminal risk assessment,[31] and facial recognition applications,[32] among others.[33]

Below, we present examples of biases introduced through algorithms and AI tools in two areas that are subject to anti-discrimination laws: employment and credit markets.

A. Bias in Employment

1. Examples of how AI may be applied in employment

AI is used for recruiting in "four general sets of activities: outreach, screening, assessment, and coordination."[34] As companies hire for multiple positions, AI efficiently organizes potential candidates away from less-suitable applicants. However, algorithms used in recruiting are at risk of reproducing bias from the real world.[35]

A well-publicized example is Amazon's attempt in 2014 to develop an automated system for hiring. Amazon's hiring algorithm

31. Risk assessment tools are used in the criminal justice system to predict the recidivism (tendency of a convicted individual to relapse into criminal behaviors) risks of offenders. These tools can influence bail and prison sentences. A 2019 study claims one of these decision models, the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) tool, overestimated the recidivism rates for Hispanic people. The study compared the COMPAS expected risk of general and violent recidivism against the observed rates for individuals arrested in Broward County, Florida. The study's findings suggested that—for both types of recidivism, general and violent—the COMPAS expected risk score tends to be higher than the observed rates for Hispanic people. *See* Melissa Hamilton, *The Biased Algorithm: Evidence of Disparate Impact on Hispanics*, 56 AM. CRIM. L. REV. 1553, 1565 (2019).

32. A 2021 study by Twitter claimed its automated image cropping system, which was implemented by AI, may have contributed to disparate impact where "Twitter's cropping system favors cropping light-skinned over dark-skinned individuals and favors cropping woman's bodies over their heads." A possible explanation by Twitter, and one consistent with other studies, is that the model was trained to "accommodate whiteness." *See* Kyra Yee et al., *Image Cropping on Twitter: Fairness Metrics, their Limitations, and the Importance of Representation, Design, and Agency*, 5 PROC. OF THE ACM ON HUM. COMPUT. INTERACTION 1, 21 (2016).

33. Bias has, also, been identified in online ads, word association, and criminal justice algorithms against African Americans. *See, e.g.*, Nicol Turner Lee et al., *Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms*, BROOKINGS INST. (May 22, 2019), https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/ [https://perma.cc/X397-P3RS].

34. *See* J. Stewart Black & Patrick van Esch, *AI-Enabled Recruiting: What Is It and How Should a Manager Use It?*, 63 BUS. HORIZONS 215, 218 (2020).

35. *See* Miranda Bogen, *All the Ways Hiring Algorithms Can Introduce Bias*, HARV. BUS. REV. (May 6, 2019), https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias [https://perma.cc/9PXH-TW93].

was "trained" on past resumes gleaned from a ten-year period, during which men applied, and were hired, more often than women.[36] Consequently, Amazon's algorithm contributed to the systematic discrimination against women applying for technical jobs.[37]

The data used to train the algorithm may have accurately represented Amazon's historical hiring practices, where significantly more male applicants led to more male hires.[38] However, by failing to examine how the training data could be strongly correlated with protected classes, such as sex, the model's decision outcome from its data analysis was to rank female candidates lower than male candidates.[39] Amazon ultimately abandoned the project.[40]

More recently, in 2019, the non-profit Electronic Privacy Information Center filed a complaint with the Federal Trade Commission against HireVue—a provider of software that evaluates job candidates based on an algorithmic assessment.[41] The complaint alleged HireVue's software, which analyzed a person's facial expressions in a video to discern certain characteristics, was discriminatory.[42] In early 2021, HireVue announced it would stop using the technology.[43]

2. Anti-discrimination oversight in employment

In 2021, the EEOC, which enforces laws that ban employment discrimination, announced an "initiative to ensure that AI and other emerging tools used in hiring and other employment decisions comply with federal civil rights laws."[44] The EEOC has already in-

36. Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women*, REUTERS (Oct. 10, 2018, 5:04 AM), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G [https://perma.cc/5AZJ-N9MY].

37. *Id.*

38. *Id.*

39. *Id.*

40. *Id.*

41. *HireVue, Facing FTC Complaint from EPIC, Halts Use of Facial Recognition*, ELEC. PRIV. INFO. CTR. (Jan. 12, 2021), https://epic.org/hirevue-facing-ftc-complaint-from-epic-halts-use-of-facial-recognition/ [https://perma.cc/8QKT-QBGR].

42. Electronic Privacy Information Center, *In re* HireVue, Inc.: Complaint and Request for Investigation, Injunction, and Other Relief Submitted 1 (Nov. 6, 2019), https://epic.org/privacy/ftc/hirevue/EPIC_FTC_HireVue_Complaint.pdf [https://perma.cc/8DFH-HNNA].

43. *HireVue, Facing FTC Complaint from EPIC, Halts Use of Facial Recognition*, *supra* note 41.

44. Press Release, EEOC, *EEOC Launches Initiative on Artificial Intelligence and Algorithmic Fairness* (Oct. 28, 2021), https://www.eeoc.gov/newsroom/eeoc-launches-initiative-artificial-intelligence-and-algorithmic-fairness [https://perma.cc/24HC-LBCR].

vestigated at least two cases involving claims that algorithms unlawfully exclude certain groups of workers during the recruitment process.[45]

While the companies making hiring decisions and using the AI tools are responsible for ensuring that AI tools are not discriminatory, the firms that create hiring algorithms can also be held liable.[46] If the factors considered within the AI algorithm are shown to have a disparate impact on protected groups of job applicants, employers must establish that those factors are both job-related and represent a reasonable measure of job performance.[47] Employers using AI hiring practices, as well as the firm that developed the algorithm, could face liability for any unintended discrimination.[48]

In an effort to prevent certain types of unintentional discriminatory treatment, organizations are developing AI models that enable hiring companies to consider candidates who may not match the characteristics of the current employees.[49] These models are designed to overcome the possibility that hiring algorithms, based on historical datasets that do not reflect current candidates, may be biased.[50] For example, academics at MIT and Columbia University are working on an AI hiring model they claim will lead to more demographic diversity of candidates selected by the AI.[51]

B. Bias in Credit Markets

1. Examples of how AI may be applied in credit markets

AI technologies have the potential to reduce costs, increase efficiency in the underwriting process, provide operational benefits for financial institutions, and provide less bias in credit offerings.

---

45. *See* Gary D. Friedman & Thomas McCarthy, *Employment Law Red Flags in the Use of Artificial Intelligence in Hiring*, AM. BAR. ASS'N (Oct. 1, 2020), https://www.americanbar.org/groups/business_law/publications/blt/2020/10/ai-in-hiring/ [https://perma.cc/JK2J-HA6M].

46. Chris Opfer, *AI Hiring Could Mean Robot Discrimination Will Head to Courts*, BLOOMBERG L. (Nov. 12, 2019, 4:01 AM), https://news.bloomberglaw.com/daily-labor-report/ai-hiring-could-mean-robot-discrimination-will-head-to-courts [https://perma.cc/35YZ-BS36].

47. Manish Raghavan & Solon Barocas, *Challenges for Mitigating Bias in Algorithmic Hiring*, BROOKINGS INST. (Dec. 6, 2019), https://www.brookings.edu/research/challenges-for-mitigating-bias-in-algorithmic-hiring/ [https://perma.cc/TK58-4TF9]; *see, e.g.*, Griggs v. Duke Power Co., 401 U.S. 424, 433 (1971); *see also* Albemarle Paper v. Moody, 422 U.S. 405, 411 (1975).

48. Friedman & McCarthy, *supra* note 45.

49. *See* Danielle Li et al., Hiring as Exploration 1–7 (Nat'l Bureau of Econ. Rsch., Working Paper No. 27736, 2020).

50. *See* Miranda Bogen, *All the Ways Hiring Algorithms Can Introduce Bias,* HARV. BUS. REV. (May 6, 2019), https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias [https://perma.cc/78D5-4KM8].

51. Danielle Li et al., *supra* note 49, at 1.

For example, financial technology (FinTech) companies seek to assess borrower creditworthiness using data, such as cell phone operating systems, education, and social media activity rather than credit scores or banking records.[52] In addition to providing credit to individuals who traditional lenders may not have approved,[53] FinTech firms may help reduce discrimination against protected classes. A 2022 academic study found that, while the federal government's financial support initiative, the Paycheck Protection Program,[54] initially made smaller loan amounts to Black-owned businesses than comparable white-owned businesses, the entry of FinTech lenders reduced this difference.[55]

However, the use of AI technologies also raises fair lending and consumer protection concerns. Rohit Chopra, Director of the Consumer Financial Protection Bureau (CFPB), has noted, "algorithms can help remove bias, but black box underwriting decisions are not necessarily creating a more level playing field and may be exacerbating the biases feeding into them."[56]

Some examples of AI applications in credit markets that demonstrate the potential for bias include the use of AI by Lemonade, an insurance firm. Lemonade stated in its Securities and Exchange Commission (SEC) filings that its complex AI model—which uses close to 1,700 data points to determine its customers' "level of risk"—may lead to unintentional bias and discrimination.[57]

Another widely known example is the Apple credit card, launched in August 2019, which appeared to offer smaller lines of

52. *What Types of Customer Data Do Fintech Firms Use?*, FED. RSRV. BANK OF ST. LOUIS (Apr. 27, 2020), https://www.stlouisfed.org/on-the-economy/2020/april/types-customer-data-fintech-firms [https://perma.cc/6PWD-EB84].

53. For the millions of individuals without banking access and credit scores, obtaining a loan or credit may be difficult. *See* Sameepa Shetty, *Start-Up Uses Mobile Data as a Credit Score for the Global Unbanked*, CNBC (Jan. 3, 2020, 7:29 AM), https://www.cnbc.com/2020/01/03/start-up-uses-mobile-data-as-a-credit-score-for-the-global-unbanked.html [https://perma.cc/GF8G-RET2].

54. The federal government established the Paycheck Protection Program through the Coronavirus Aid, Relief, and Economic Security Act (CARES Act). The Program's objective was to provide financial support to small businesses, "eligible nonprofit organizations, Veterans organizations," eligible "Tribal businesses," as well as eligible self-employed or independent contractor individuals. *See* Paycheck Protection Program, U.S. DEPT. OF THE TREASURY, https://home.treasury.gov/policy-issues/coronavirus/assistance-for-small-businesses/paycheck-protection-program [https://perma.cc/5LW8-43NC].

55. *See* Rachel Atkins et al., *Discrimination in Lending? Evidence from the Paycheck Protection Program*, 58(2) SMALL BUS. ECON. 843, 844 (2021).

56. Lindsay Muniz, *DOJ, CFPB, OCC to Target So-Called Digital Redlining*, COLO. BANKERS ASS'N (Oct. 26, 2021), https://www.coloradobankers.org/news/584714/DOJ-CFPB-OCC-to-Target-So-Called-Digital-Redlining.html [https://perma.cc/V9GN-TJE3].

57. Lemonade, Inc., Registration Statement (Form S-1), SEC. AND EXCH. COMM'N 27, 125 (June 8, 2020), https://www.sec.gov/Archives/edgar/data/1691421/000104746920003416/a2241721zs-1.htm [https://perma.cc/269S-FP5H].

credit to women than to men—even in cases of heterosexual married couples who shared assets where the woman had the higher credit score.[58] Ultimately, it was determined that Goldman Sachs, who was responsible for implementing the algorithm and determining credit offerings, did not discriminate based on sex.

Goldman Sachs was able to identify the factors that led to the credit decisions, such as credit score, indebtedness, income, credit utilization, missed payments, and other credit history elements. These decisions appeared to be consistent with the bank's credit policy, and none of the factors identified were an unlawful basis for a credit determination.[59]

Goldman Sachs' algorithm was determined to be "gender-blind": it did not include race, gender, nor marital status as inputs to the model.[60] Instead, credit score, indebtedness, income, credit utilization, missed payments, and other credit history elements—none of which were viewed as "unlawful bas[e]s" for a credit determination—were factors used to evaluate credit.[61]

Moreover, most credit scoring models are "black box" proprietary algorithms, where the underlying computer logic is not readily available for regulatory and public scrutiny.[62] Companies creating alternative credit scoring models through "proprietary" algorithms treat their "machine-learning tools as closely guarded trade secrets" and may not be explicit as to whether AI does play a role.[63] However, as noted by the National Consumer Law Center (NCLC), "the use of complex, opaque algorithmic models in consumer credit transactions also heightens the risk of unlawful discrimination, and unfair, deceptive, and abusive practices."[64]

---

58. Neil Vigdor, *Apple Card Investigated After Gender Discrimination Complaints*, N.Y. TIMES (Nov. 10, 2019), https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html [https://perma.cc/M3TS-NKFE]; N.Y. STATE DEPT. FIN. SERVS., REP. ON APPLE CARD INVESTIGATION 2 (2021) (the New York State Department of Financial Services review of the data concluded that there was no "evidence of deliberate or disparate impact discrimination but showed deficiencies in customer service and transparency.").

59. N.Y. STATE DEPT. FIN. SERVS., *supra* note 58, at 7.

60. *Id.* at 6.

61. *See id.* at 7. This example highlights ways companies can identify and mitigate disparate impact. *See infra* Section III for more discussion of such approaches.

62. *See* WBG, Credit Scoring Approaches Guidelines 21 (2019), https://thedocs.worldbank.org/en/doc/935891585869698451-0130022020/original/CREDITSCORINGAPPROACHESGUIDELINESFINALWEB.pdf [https://perma.cc/T4VU-6WV6].

63. Mikella Hurley & Julius Adebayo*, Credit Scoring in the Era of Big Data*, 18 YALE J. OF L. & TECH. 148, 158 (2016).

64. Nat'l Consumer L. Ctr., Comment Letter on Request for Information and Comment on the Financial Institutions' Use of Artificial Intelligence, Including Machine Learning (July 1, 2021), https://www.regulations.gov/comment/NCUA-2021-0091-0015 [https://perma.cc/ZC3V-H7EA].

2. Anti-discrimination oversight in credit markets

Federal agencies investigating lending discrimination have long recognized disparate impact in their supervision and enforcement efforts. For example, the Consumer Finance Protection Bureau (CFPB), Department of Justice (DOJ), and Office of the Comptroller of the Currency (OCC) have all stated their interest in "working to rid the market of racist business practices, including those by discriminatory algorithms."[65] The CFPB, in Bulletin 2012–04 on lending discrimination, affirmed its adherence to the fair lending principles outlined in ECOA and Regulation B, as well as its support of the Policy Statement on Fair Lending issued by federal agencies.[66] Consumer and civil rights groups also actively monitor the impact of AI in credit markets, for example, in evaluating housing discrimination by AI-powered advertising platforms and financial discrimination by student loan lenders.[67]

Lending decisions based on AI tools may violate ECOA and/or FHA&A rules and regulations which prohibit discrimination in residential real estate related loans. Examples in this area include disparate impact claims—made by the NCLC and other consumer and civil rights advocates—challenging creditor policies permitting car dealers to markup loan interest rates based on subjective criteria unrelated to creditworthiness,[68] or litigation brought against mort-

65. *CFPB, DOJ and OCC Take Action Against Trustmark National Bank for Deliberate Discrimination Against Black and Hispanic Families*, CFPB (Oct. 22, 2021), https://www.consumerfinance.gov/about-us/newsroom/cfpb-doj-and-occ-take-action-against-trustmark-national-bank-for-deliberate-discrimination-against-black-and-hispanic-families/ [https://perma.cc/37X5-LLYA]. Changes to organizational structures within government regulators signal an increased emphasis in their seeking to determine whether the computer models companies utilize to make decisions have adverse impact on consumers. For example, the CFPB created a new position, chief technologist, to help understand how decision models within financial industries truly work. *See* Georgia Kromrei, *Director Chopra Shakes Up CFPB Leadership with New Appointments*, REVERSE MORTG. DAILY (Oct. 14, 2021, 8:17 PM), https://reversemortgagedaily.com/2021/10/14/director-chopra-shakes-up-cfpb-leadership-with-new-appointments/ [https://perma.cc/2H89-S929].

66. Consumer Fin. Prot. Bureau, Bull. No. 2012-04 (Fair Lending) 1 (2012), https://files.consumerfinance.gov/f/201404_cfpb_bulletin_lending_discrimination.pdf [https://perma.cc/955S-CF5H].

67. Nat'l Consumer L. Ctr., *supra* note 64.

68. *See generally* Closed Cases: Auto Finance Discrimination, NAT'L CONSUMER L. CTR., https://www.nclc.org/our-work/our-services/litigation/closed-cases/ [https://perma.cc/2H3M-9BZD]; JOHN W. VAN ALST, NAT'L CONSUMER L. CTR, TIME TO STOP RACING CARS: THE ROLE OF RACE AND ETHNICITY IN BUYING AND USING A CAR 3 (2019), https://www.nclc.org/wp-content/uploads/2022/08/report-time-to-stop-racing-cars-april2019.pdf [https://perma.cc/55K8-L8V6].

gage lenders whose policies resulted in more expensive loans to protected classes than similarly situated white borrowers.[69] AI models can accelerate existing discriminatory patterns and create systemic risks for credit and housing market consumers, thereby leading to digital redlining. For example, Redfin, a real estate technology company that uses decision models to set housing prices, was sued over allegations that its housing price minimums systematically denied services in areas where a significant portion of residents were non-white.[70]

III. IDENTIFYING DISPARATE IMPACT

Assessing and quantifying unintentional disparate impact resulting from AI-based decision-making tools is a complex technical undertaking, as AI models are inherently difficult to interpret and are frequently opaque. Below, we summarize approaches that have been considered by economists, regulators, companies, litigators, and other interested parties for identifying, remedying, and preventing disparate impact driven by AI processes. We describe both quantitative and qualitative approaches to analyzing disparate impact. Notably, while limitations in data may dictate the relevance of any particular approach, AI itself can be applied pre-emptively to detect or reduce bias.

A. Quantitative Analysis

1. Statistical Analysis

Statistical analysis has frequently been used to evaluate claims of employment discrimination and discrimination in lending practices.[71] Comparisons of averages across class types can provide

69. *See generally, e.g.*, Ramirez v. GreenPoint Mortg. Funding, Inc., 268 F.R.D. 627 (N.D. Cal. 2010); Guerra v. GMAC, L.L.C., 2009 WL 449153 (E.D. Pa. Feb. 20, 2009); Taylor v. Accredited Home Lenders, Inc., 580 F. Supp. 2d 1062 (S.D. Cal. 2008); Miller v. Countrywide Bank, 571 F. Supp. 2d 251 (D. Mass. 2008); Ware v. Indymac Bank, 534 F. Supp. 2d 835 (N.D. Ill. 2008).

70. *See* Andrew Martinez, *Redfin Reaches Settlement in Discrimination Lawsuit*, NAT'L MORTG. NEWS (Apr. 29, 2022, 2:36 PM), https://www.nationalmortgagenews.com/news/redfin-reaches-settlement-in-discrimination-lawsuit [https://perma.cc/AD8K-ARU7].

71. *See, e.g.*, Peter H. Wingate & George C. Thornton, *Statistics and Employment Discrimination Law: An Interdisciplinary Review*, 19 RSCH. IN PERS. AND HUM. RES. MGMT. 295 (2000); Winnie F. Taylor, *Proving Racial Discrimination and Monitoring Fair Lending Compliance: The Missing Data Problem in Nonmortgage Credit*, 31 REV. BANKING & FIN. L. 199, 212 (2011); *see also* Hanming Fang & Andrea Moro, *Theories of Statistical Discrimination and Affirmative Action: A Survey*, *in* HANDBOOK SOC. ECON. 133 (Jess Benhabib et al. ed., 2011); Palmer Morrel-Samuels, Statistical Analysis in Employment Discrimination: Trends and Implications (2018), https://ssrn.com/abstract=3205709 [https://perma.cc/W2P6-3QQ8].

evidence of disparate impact. For example, if 5 percent of an employer's female applicants are offered jobs, compared to 55 percent of an employer's male applicants, then the employer's hiring practices suggest disparate impact by gender.[72]

The EEOC uses a four-fifths rule of thumb as guidance for identifying disparate impact. Under the four-fifths rule, the EEOC will generally consider a selection rate for any race, sex, or ethnic group less than four-fifths (or 80 percent) of the selection rate for the group with the highest selection rate as a substantially different rate of selection.[73] The EEOC states that the guidance "is not intended as a legal definition" of disparate impact, but rather, provides "a practical means of keeping the attention of the enforcement agencies on serious discrepancies in rates of hiring, promotion and other selection decisions."[74]

Statistical tests that evaluate whether the observed outcome would be unlikely to result from chance include the Fisher's exact test and the chi-square test, which are often considered more informative than the four-fifths rule of thumb.[75]

---

72. Nicholas Schmidt et al., *How Data Scientists Help Regulators and Banks Ensure Fairness when Implementing Machine Learning and Artificial Intelligence Models*, 3 (Conference on Fairness, Accountability, and Transparency, Feb. 23–24, 2018), https://facctconference.org/static/tutorials/schmidt_banking18.pdf [https://perma.cc/FP4Y-W2JL]. Note that, in a case of disparate treatment, in order to ascertain whether an applicant's gender was the cause of being denied employment, the comparison should control for non-gender factors that could influence the hiring decision. However, in a disparate impact case, only non-gender factors that are plausible legitimate business justifications should be included as controls because, in a disparate impact case, the causal role of an unjustified applicant attribute should not be used to explain away a gender disparity in hiring. *See* Ian Ayres, Testing for Discrimination and the Problem of Included Variable Bias 24 (2010) (working paper), https://ianayres.yale.edu/sites/default/files/files/Testing%20for%20Discrimination%20and%20the%20Problem%20of%20_Included%20Variable.pdf [https://perma.cc/K9BH-G85Q].

73. Section 4D, Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607.4(D) (as amended 1981).

74. Questions and Answers to Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures, 44 FED. REG. 11996, 11998 (Mar. 2, 1979).

75. The four-fifths rule does not consider the potential impact of sampling error. The Fisher Exact Test, unlike the four-fifths rule, provides comparisons to a known error rate at the 0.05 level of statistical significance. The chi-square test tests whether two variables are significantly associated: it counts the frequencies of observations in a 4-cell table, where the columns define one variable, the rows define another, and the chi-square statistic computes the likelihood that data from the two variables are related. *See* Scott B. Morris & Russell E. Lobsenz, *Significance Tests and Confidence Intervals for the Adverse Impact Ratio*, 53 PERS. PSYCH. 89, 93–97 (2000); *see also Statistical Sophistication Would Have Provided A Different Liability Answer*, FULCRUM FIN. INQUIRY (Oct. 2013), https://www.fulcrum.com/statistical-sophisticat/ [https://perma.cc/4UGB-VNQH].

Regression analysis can also be used to evaluate disparate impact, estimating the average differential impact on protected classes and whether this estimated impact is statistically significant.[76] For example, regression analysis can help a plaintiff establish a claim of disparate impact under Title VII by showing, even when controlling for legitimate organizational factors, outcomes for individuals of a particular class were different from other employees or applicants at a certain level of statistical significance.[77]

With regression analysis, one needs to consider the appropriateness of the control variables. For example, not including a key variable could potentially lead to omitted variable biases, while adding too many variables could potentially result in over-specification and "multicollinearity."[78] An incorrectly specified regression could result in false negatives (not being able to identify disparate impact) or false positives (incorrectly finding evidence of disparate impact).

Notably, statistical analysis of disparate impact or treatment claims typically requires data about an individual's protected class status. While such data are often available in litigation related to employment and mortgage lending, the absence of such data could hinder "federal regulatory efforts to identify discriminatory lending patterns in non-mortgage credit transactions."[79]

2. Data Analytics

Restrictions on collecting race or gender data in non-mortgage credit markets[80] and regression modeling limitations[81] make less traditional analyses that leverage AI models attractive alternatives. Importantly, as we discuss here, while AI can contribute to discrimination, it can also be used to identify it. AI, accompanied by

76. Regression analysis is a statistical method for determining the relationship that exists in a set of data between a variable to be explained—called the "dependent variable"—and one or more "explanatory variables." *See* Lisa Sullivan, *Correlation and Linear Regression*, B.U. SCH. PUB. HEALTH (Nov. 12, 2022), https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_correlation-regression/bs704_correlation-regression_print.html [https://perma.cc/XQR3-AT4R].

77. Ayres, *supra* note 72, at 8–10.

78. Over-specification occurs when one or more redundant predictor variables are specified. Multicollinearity occurs when independent variables in a regression model are correlated. *See* William H. Greene, *Econometric Analysis*, 57–58, 88, 148–49 (7th ed. 2002).

79. Taylor, *supra* note 71, at 201.

80. *See* U.S. GOV'T ACCOUNTABILITY OFF., GAO-08-698, FAIR LENDING: RACE AND GENDER DATA ARE LIMITED FOR NON-MORTGAGE LENDING 5 (2008).

81. Ayres, *supra* note 72, at 51.

data science[82] and techniques, such as sampling, could be used to review diverse and large amounts of structured and unstructured information. This analysis could help identify trends and relationships across the data that may be correlated with protected classes. Also, while AI models could be one factor contributing to discrimination, it is important to recognize the existence of an AI model should not imply that the outcome is discriminatory. More specifically, research has shown AI, depending how it is used, can have either a positive or a negative effect on disparity.[83] Other research has argued, absent legal constraints, including protected attributes could sometimes reduce discrimination and improve predictive quality.[84]

One form of AI, NLP, could be used to perform lexical processing and sentiment analysis of unstructured text to identify biases. For example, by assigning sentiments within conversations as positive or negative, IBM claims that its NLP, Watson Natural Language Understanding, can detect whether an AI model "inherits discriminatory bias from its input data."[85] Through such an AI analysis, one could potentially identify whether the AI model views a male candidate positively and a female candidate negatively.

While AI models allow one to analyze all elements within a large quantity of data, sampling, on the other hand, is a technique where a subset of the population is selected for analysis. To generalize the results of the analysis, the sampled subset must be representative of the population. A properly selected smaller dataset would allow one to check for disparate impact in a timelier and cost-

82. Data science refers to the capture, maintenance, processing, analysis, and communication of data. Through data science, one is able to extract value from data, even large quantities of data. *See What is Data Science?*, U.C. BERKELEY (Nov. 12, 2022), https://ischoolonline.berkeley.edu/data-science/what-is-data-science/ [https://perma.cc/73G2-H67J].

83. *See* Jon Kleinberg et al., *Discrimination in the Age of Algorithms*, 10 J. LEGAL ANALYSIS 113, 163 (2018) (concluding that "[t]he use of algorithms offers far greater clarity and transparency about the ingredients and motivations of decisions, and hence far greater opportunity to ferret out discrimination."); *but see* Andreas Fuster et al., *Predictably Unequal? The Effects of Machine Learning on Credit Markets*, 77 J. FIN. 5, 5 (2020) (finding that Black and Hispanic borrowers are disproportionately less likely to gain from the introduction of machine learning in credit markets).

84. *See*, *e.g.*, Zachary C. Lipton et al., Does Mitigating ML's Impact Disparity Require Treatment Disparity? 5 (Jan. 11, 2019) (working paper), https://arxiv.org/abs/1711.07076 [https://perma.cc/728B-ETPP]. It is notable that some researchers argue that using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models. *See* Indré Žliobaitė & Bart Custers, *Using Sensitive Personal Data May Be Necessary for Avoiding Discrimination in Data-Driven Decision Models*, 24 A.I. & L. 183, 198 (2016).

85. Sean Sodha, *Using Watson NLU to Help Address Bias in AI Sentiment*, IBM WATSON BLOG (Feb. 12, 2021), https://www.ibm.com/blogs/watson/2021/02/watson-nlu-bias-ai-sentiment-analysis/ [https://perma.cc/LAT2-WKG6].

effective manner. This investigation could be conducted using statistical analysis, as described above, or manually, as described below.

3. Manual Data Review

Concerns over computer-generated biases may be resolved through a non-computer-based approach, with humans conducting a manual review of the AI model's decision. While AI programs are designed to think and react like a human, they may not always succeed in this endeavor. Manual reviews may detect outcomes for which the AI program was not designed nor trained to accommodate. Manual reviews can be used to help verify patterns in the data and to help check for cases of false positive and false negative trends.

A manual review of the programming code used to design the AI model could also help identify disparate impact. By reviewing and executing the lines of code one by one—also known as "stepping through the code"—an individual can gain a more complete understanding of the program. One can create an algorithm and flow chart to understand how the program works and identify any potential biases in the system.

B. Qualitative Analysis

1. Documentation of the Model

Despite the name "artificial intelligence," decision models often reflect the beliefs of the model designer. These models could be opaque "black boxes" that make it difficult for those not involved in the model's design to understand why their outcomes could result in a disparate impact.

Documenting the model so it is not a black box, so someone not involved in its design could understand how it operates, could help prevent issues, such as the designer inappropriately weighing inputs. The documentation—including defining the problem and the algorithm used to answer it, assumptions, reasons for selecting the inputs, and validation of the model—should seek to explain how the designer sought to solve a potentially unstructured problem (e.g., determining "good" employees) using data that has been formatted to be analyzed by a computer. Allowing individuals who are not part of the design team to audit the model could help prevent designer bias, both intentional and unintentional. Documentation could, also, help avoid scenarios where only the model creators understand the algorithm and potential pitfalls.

2. Inform Affected Parties of Reasons for Model's Decision

The current focus of regulators has been on developing an understanding of how AI decision models function. The ECOA requires creditors, including those using AI models, to provide applicants with reasons for denial of credit.[86] The CFPB has sought to provide guidance on the types of adverse action notices organizations using AI models should provide to consumers.[87] To be current with regulators' viewpoints, organizations that utilize AI programs should be able to communicate to their consumers the specific reasons the model arrived at an adverse decision, even when the relationship between the reason and the model outcome may not be clear to the consumer.[88]

Furthermore, recording the reasons for the model's positive and negative decisions could help determine if certain factors more heavily influence the model's decision. This information, combined with the model documentation (as described above), could allow one to determine if the model is performing as designed. Likewise, these policies could help determine if one factor plays an outsized role in the model arriving at positive outcomes and another more strongly influencing adverse outcomes.

3. Incorporate Feedback from Affected Parties

One contributing factor to the opacity of a model is the lack of solicitation and incorporation of feedback on the model's performance. Regulators have suggested that companies should inform adversely affected consumers as to why a decision is made and allow consumers the opportunity to respond.

In some situations, bias in the model may not be readily apparent to the organization. Thus, in addition to informing consumers why they received an adverse decision, allowing them to respond provides meaningful feedback on a model. Through this feedback, consumers may help confirm—or dispute—the model's decisions and help reduce opaqueness.

4. Strategic Testing

As stated, decision models are dependent on the input data used to train the model. Further, as these models use more diverse data as inputs, it is critical to understand the role of these inputs.

---

86. 15 U.S.C. § 1691(d).

87. Patrice A. Ficklin et al., *Innovation Spotlight: Providing Adverse Action Notices When Using AI/ML Models*, CFPB BLOG (July 7, 2020), https://www.consumerfinance.gov/about-us/blog/innovation-spotlight-providing-adverse-action-notices-when-using-ai-ml-models/ [https://perma.cc/WGG8-C5Y2].

88. 12 C.F.R. § 1002.9(b)(2)(4) (2011).

Specifically, it is necessary to understand how the input variables could be correlated with protected classes and how various combinations of the input variables could also be correlated. While one could control for the protected classes to help reduce such correlations, data on the protected classes may not always be available. Moreover, there could be a reluctance to eliminate an input that could be correlated with a protected class, as it may reduce the model's precision. One should review whether the variables selected by the model to affect the outcome can be reasonably explained. If the training data is unstructured or needs to be classified, having a diverse team perform and/or review the classification could reduce biases in the labeling of the data.

In addition to reviewing the data used as input, one should also consider how the data is collected and whether the data collection medium itself could lend itself to ignoring certain segments of the population. For example, the city of Boston developed a smartphone app that used GPS and accelerometers data to inform the city of potholes.[89] However, before implementation, the developers recognized that elderly and low-income residents might not own smartphones. Consequently, the app could ignore neighborhoods with concentrations of these residents.[90] Datasets that disproportionately over- or under-represent segments of the population may lead to biased outcomes.

CONCLUSION

Technological improvements have resulted in AI and ML algorithms with the potential to cause unintentional discrimination against consumers and prospective employees. Companies are using more sophisticated computer programs to help decide whether to engage in a housing transaction, hire job candidates, or lend credit to someone. These programs make decisions using more diverse and non-traditional inputs. Consumers, regulators, such as the EEOC and CFPB, and advocacy groups have started to recognize that these AI programs are often trained using data that may be correlated with protected classes. Consequently, while AI models may apply the same policies to everyone, decision outcomes may adversely affect protected classes.

While there have been disparate impact claims brought against lending, housing, and hiring-related firms, given the prevalence of AI programs in these industries—and the growing interests of advocacy groups and regulators of these industries—there is

89. EXEC. OFF. PRES., *supra* note 3, at 51.
90. EXEC. OFF. PRES., *supra* note 3, at 51–52.

anticipation of an increase in regulatory investigations and legal cases related to disparate impact of AI programs.

More and more companies are increasing their use of AI models to help with business decisions, thus increasing the likelihood of systematic discrimination against protected classes. Furthermore, qualifying for financial credit, obtaining housing, and securing employment are all connected to one another. It is difficult to rent an apartment if one does not have a job and a favorable line of credit. If one does not have housing, it is difficult to obtain or maintain employment, and without a job or housing as collateral, it is difficult to obtain credit. Furthermore, lending, housing, and hiring have historically faced allegations of discrimination. As the U.S. Supreme Court noted in *Griggs v. Duke Power Co.*, policies and procedures, like those found in AI models, should not allow for the status quo of prior discriminatory practices to continue.[91]

Existing labor, housing, and financial laws allow regulators, advocacy groups, employees, customers, and other stakeholders to bring disparate impact claims against these computer-generated avenues for discrimination. However, determining evidence of discrimination may require a range of skills and techniques. More specifically, identifying, remedying, and preventing disparate impacts from AI models requires a mix of econometric skills to perform robust statistical analyses and familiarity with technology to understand the computer algorithm's design, purpose, testing, and implementation.

APPENDIX

This appendix includes additional information of the protected classes listed *infra* Section II.D. and cites examples where these descriptions arise.

A. Age

The Age Discrimination in Employment Act (ADEA) prohibits discrimination against individuals 40 years or older.[92] The ECOA also prohibits age discrimination, provided the individual is legally able to enter into a contract.[93] The ECOA lists circumstances under which inquiring of an individual's age or use of age in a credit system does not constitute discrimination.[94] While the FHA&A does

91. Griggs v. Duke Power Co., 401 U.S. 424, 430 (1971).
92. 29 U.S.C. § 623 (1967).
93. 15 U.S.C. § 1691(a).
94. *See* 15 U.S.C. § 1691(b)(2)–(4).

not expressly prohibit age discrimination, state laws sometimes have age-related prohibitions against discrimination in housing.[95]

B. Color

Employment, housing, and lending laws all seek to prohibit discrimination based on an individual's skin color.[96]

C. Disability

The FHA&A prohibits discrimination based on disability in all types of housing transactions.[97] The FHA&A defines disability as "a physical or mental impairment[, or record of having such an impairment,] that substantially limits one or more major life activities."[98] The FHA&A provides examples where the term handicap does not apply, such as "the illegal use of or addiction to a controlled substance."[99]

Federal laws prohibit employers from discriminating based on disability. Under the Americans with Disabilities Act (ADA), employers are required to provide reasonable accommodations to employees or job applicants with disabilities, unless doing so would cause undue hardship to the employer.[100] Furthermore, discrimination against an employee based on their relationship with a disabled person is prohibited.[101]

D. Familial Status

The FHA&A prohibits discrimination based on familial status—that is, adverse treatment of a person because they have a family with one or more individuals under the age of 18.[102]

E. Genetic Information

The Genetic Information Nondiscrimination Act (GINA) of 2008 prohibits discrimination based on genetic information by health insurers (Title I of the Act) and employers (Title II of the

95. *See, e.g.*, 775 Ill. Comp. Stat. Ann. 5/1-102(A) (LexisNexis 2022) (stating a policy of prohibiting discrimination in real estate transactions based on, among other things, age).

96. *See* 42 U.S.C. § 2000e-2(a)(1)–(2) (1991) (employment); 15 U.S.C. § 1691(a) (lending); 42 U.S.C. § 3604 (housing); 42 U.S.C. § 3602 (housing).

97. *See* 42 U.S.C. § 3604(f).

98. *See* 42 U.S.C. § 3602(h).

99. *Id*.

100. 42 U.S.C. § 12112(b)(5)(A)–(B) (2009).

101. 42 U.S.C. § 12112(b)(4).

102. 42 U.S.C. § 3602(k).

Act).[103] GINA defines genetic information as information related to genetic tests about an individual, family members, and family medical history.[104] While GINA does not establish disparate impact as a cause of action following genetic discrimination, Section 208 allows for a review of developments in genetics and recommendations to Congress on whether to include a future allowing disparate impact as a cause of action.[105]

F. National Origin

Employment, housing, and lending laws all prohibit discrimination based on ethnicity, accent, country, or the part of the world from which an individual originates.[106] Discrimination may also extend to unfavorable treatment based on marriage to or association with individuals of a certain national origin.[107]

G. Race

Discrimination based on an individual's race or personal characteristics associated with a particular race—e.g., hair texture, facial features—is prohibited in employment, housing, and lending laws.[108] Discrimination may, also, extend to unfavorable treatment based on marriage to or association with individuals of a certain race.[109]

H. Religion

Employment, housing, and lending laws seek to prevent discrimination based on religious beliefs.[110] Discrimination may also extend to unfavorable treatment based on marriage to or association with individuals of a certain religion.[111]

---

103. Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233, 122 Stat. 881 (2008) (codified in scattered sections of 29 and 42 U.S.C.).

104. *See* 42 U.S.C. § 2000ff(4).

105. See 42 U.S.C. § 2000ff-7(a)–(b).

106. *Supra* note 98.

107. 29 C.F.R. § 1606.1 (1980).

108. 42 U.S.C. § 3602(h); *Race/Color Discrimination*, EEOC, https://www.eeoc.gov/racecolor-discrimination [https://perma.cc/E6DX-E7P7] (Nov. 12, 2022).

109. *Race/Color Discrimination*, *supra* note 108.

110. 42 U.S.C. § 3602(h).

111. *Religious Discrimination*, EEOC, https://www.eeoc.gov/religious-discrimination [https://perma.cc/C9P7-3YRY] (Nov. 12, 2022).

I. Sex[112]

Discrimination based on sex is barred by employment, housing, and lending laws.[113] Lending laws include marital status when describing sex as a protected class.[114] Labor laws include an individual's sexual orientation, gender identity, and pregnancy in sex-based discrimination.[115]

J. Veteran Status

VEVRAA seeks to prohibit employment discrimination against protected veterans (Vietnam and non-Vietnam eligible veterans), but only for government contractors and subcontractors.[116] Protected veterans include active-duty wartime or campaign badge veterans, Armed Forces service medal veterans, disabled veterans, and recently separate veterans.[117]

112. See Iris Hentze & Rebecca Tyus, *Sex and Gender Discrimination in the Workplace*, NAT'L CONF. OF STATE LEGISLATURES (Aug. 12, 2021), https://www.ncsl.org/research/labor-and-employment/-gender-and-sex-discrimination.aspx [https://perma.cc/M36X-V7RR] (stating that "various rulings by the [EEOC] extend [the] prohibition on sex discrimination to include prohibit discrimination on the basis of sexual orientation and gender identity").

113. *Supra* note 98.

114. *Supra* note 95.

115. 42 U.S.C. § 2000e(k).

116. *See generally* 38 U.S.C. § 4212(a)(1)–(2).

117. 38 U.S.C. § 4212(a)(3)(A).