Arkansas Law Review

Volume 71 | Number 2 Article 7

December 2018

Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices

McKenzie Raub
University of Arkansas, Fayetteville

BOTS, BIAS AND BIG DATA:

ARTIFICIAL INTELLIGENCE, ALGORITHMIC BIAS AND DISPARATE IMPACT LIABILITY IN HIRING PRACTICES*

I. INTRODUCTION

"With artificial intelligence, we are summoning the demon. You know all those stories where there's the guy with the pentagram and the holy water and he's like, yeah, he's sure he can control the demon? Doesn't work out."[1] While this is perhaps dramatic, many Americans share Elon Musk's underlying anxieties about artificial intelligence's increasing proliferation into everyday life.[2] However, few realize the depth of artificial intelligence's involvement in mundane daily activities.[3] Fewer than half of Americans are aware of the existence of "computer programs that can review job

* The author sincerely thanks Danielle Weatherby, Associate Professor of Law, University of Arkansas School of Law, for her thoughtful guidance throughout the drafting and editing this comment. The author also thanks Amanda Beth Hurst, Visiting Professor of Law, University of Arkansas School of Law, Lucas Sheets, Clay Sapp, Pete Brunson, Lacey Johnson, Ryan Smith, and fellow staff editors of the Arkansas Law Review for their encouragement and attention to detail throughout the drafting and editing process. The author would also like to thank Dr. Trisha Posey, Associate Professor of History, Dr. Robert Moore, Associate Professor of History, and Dr. Preston Jones, Professor of History, John Brown University, for teaching the author to approach research and writing with excellence and humility. Finally, the author would like to thank her family for their unwavering encouragement and support.

1. Maureen Dowd, *Elon Musk's Future Shock*, VANITY FAIR, Apr. 2017, at 116, 119.

2. Aaron Smith & Monica Anderson, *Automation in Everyday Life*, PEW RES. CTR. (Oct. 4, 2017), http://www.pewinternet.org/2017/10/04/automation-in-everyday-life/ [https://perma.cc/3A3J-7J5A].

3. *See* Gautam Narula, *Everyday Examples of Artificial Intelligence and Machine Learning*, TECH EMERGENCE (Mar. 29, 2018), https://www.techemergence.com/everyday-examples-of-ai/ [https://perma.cc/MRK4-M5NK].

applications without any human involvement."[4] Despite this, there are a growing number of companies using algorithms and artificial intelligence to simplify hiring.[5] Artificial intelligence developers boast that their programs both streamline and remove bias from recruiting and hiring.[6]

Artificial intelligence has incredible positive societal potential. For example, predictive algorithms are being utilized to increase efficiency in providing necessary resources to abused children.[7] But with that potential for good comes a dark side that cannot be ignored. There is increasing evidence that artificial intelligence is not the unbiased savior it is often heralded to be.[8] Without accountability and responsibility, the use of algorithms and artificial intelligence leads to discrimination and unequal access to employment opportunities.[9] If employers wish to take advantage of the potential efficiency benefits of using artificial intelligence in hiring, they should use caution in selecting a program, encourage the use of responsible algorithms, and push for long term changes in the lack of racial and gender diversity in the technology industry.[10]

---

4. Smith & Anderson, *supra* note 2.

5. Jennifer Alsever, *How AI Is Changing Your Job Hunt*, FORTUNE (May 19, 2017), http://fortune.com/2017/05/19/ai-changing-jobs-hiring-recruiting/ [https://perma.cc/TZQ5-D73D]; Simon Chandler, *The AI Chatbot Will Hire You Now*, WIRED (Sept. 13, 2017, 6:45 AM), https:// www. wired. com/ story/ the- ai- chatbot- will- hire- you- now/ [https://perma.cc/XK5U-5PUP].

6. Chandler, *supra* note 5.

7. Dan Hurley, *Can an Algorithm Tell When Kids Are in Danger?*, N.Y. TIMES (Jan. 2, 2018), https://www.nytimes.com/2018/01/02/magazine/can-an-algorithm-tell-when-kids-are-in-danger.html [https://perma.cc/3XH5-3NZG].

8. Claire Cain Miller, *When Algorithms Discriminate*, N.Y. TIMES (July 9, 2015), https:// www.nytimes.com/ 2015/ 07/ 10/ upshot/ when- algorithms- discriminate.html [https://perma.cc/8CQD-9U2Y].

9. *See* Dipayan Ghosh, *AI is the Future of Hiring, But It's Far From Immune to Bias,* QUARTZ (Oct. 17, 2017), https://work.qz.com/1098954/ai-is-the-future-of-hiring-but-it-could-introduce-bias-if-were-not-careful/ [https://perma.cc/AX9G-B2D2] (writing that "[w]hen AI and recruiting come together thoughtfully and ethically, they can encourage better candidate fits, promote fairer interview screening, and increase overall efficiency. But we must also be mindful of the specter of harms like algorithmic discrimination and implicit harmful bias in AI-enabled recruiting, and do our best to counter them. Nothing is less fair to the people whose livelihoods are at stake").

10. *Id.*

The discussion is divided into three basic parts. Part one will provide a brief overview of artificial intelligence technology, its societal implications, and use emerging uses in hiring. Part two will discuss the potential for Title VII disparate impact arising from the use of artificial intelligence in hiring. Finally, part three will discuss proposed solutions to the challenges associated with the use of artificial intelligence technology, ultimately advocating for an approach that involves careful selection of the artificial intelligence program and balancing the use of artificial intelligence technology with human intuition.

II. ARTIFICIAL INTELLIGENCE TECHNOLOGY

The idea of artificial intelligence is not new, despite the futuristic spin that it is often given in popular culture.[11] In 1947, Alan Turing[12] told a crowd at the London Mathematical Society that "what we want is a machine that can learn from experience."[13] Nearly ten years later, in 1956, a group of mathematicians and scientists at Dartmouth College coined the term "artificial intelligence."[14]

Artificial intelligence has its own language. Broadly, artificial intelligence describes "any technique that enables computers to mimic human intelligence."[15] Machine learning is a subset of artificial intelligence which applies statistical techniques to "enable machines to improve at tasks with experience."[16] Artificial intelligence covers a wide range of

11. *See generally* Gil Press, *A Very Short History of Artificial Intelligence (AI)*, FORBES (Dec. 30, 2016, 9:09 AM), https:// www. forbes. com/ sites/ gilpress/ 2016/ 12/ 30/ a-very-short-history-of-artificial-intelligence-ai/#6fcca48c6fba [https://perma.cc/U9CJ-Y6E7].

12. Alan Turing was a computer scientist considered to be the father of computer science and artificial intelligence. Gil Press, *Alan Turing Predicts Machine Learning and the Impact of Artificial Intelligence on Jobs*, FORBES (Feb. 19, 2017, 1:44 PM), https://www.forbes.com/sites/gilpress/2017/02/19/alan-turing-predicts-machine-learning-and-the-impact-of-artificial-intelligence-on-jobs/#37938d821c2b [http://perma.cc/8FXE-RACZ].

13. *Id.*

14. Matt Vella, *How A.I. is Transforming Our World*, TIME (SPECIAL EDITION), Sept. 29, 2017, at 5, 5.

15. *Id.* at 7.

16. *Id.*

technologies that facilitate computers and robots to solve problems.[17] Within the realm of artificial intelligence, machine learning allows computers to improve at performing tasks with experience.[18]

First, this section will discuss in broad terms how machine learning and big data work together under the umbrella of artificial intelligence technology. The focus will then shift to both the positive and negative societal implications of big data and machine learning in society at large, and how this technology is utilized in the hiring context. Finally, there will be a brief discussion of the challenges stemming from a lack of meaningful diversity in the technology industry and public perceptions of artificial intelligence technology.

A. Big Data & Machine Learning

Deep learning, sometimes referred to as "neural networks",[19] is a 'subset of a subset' of artificial intelligence."[20] Deep learning is an advanced form of machine learning that allows software to "train itself to perform tasks, like speech and image recognition, by exposing multilayered neural networks to vast amounts of data."[21] Neural network software is modeled after the way adjustable networks of neurons in the human brain are believed to function instead of an inflexible set of instructions pre-created by programmers.[22]

Algorithms give computers guidance on how to solve problems.[23] There is no artificial intelligence without

17. Roger Parloff, *The Deep-Learning Revolution*, TIME (SPECIAL EDITION), Sept. 29, 2017, at 10, 14.

18. Vella, *supra* note 14, at 7.

19. Deep learning is a new term to describe an approach to artificial intelligence which is sometimes referred to as neural networks, or neural nets. These terms have been "going in and out of fashion for more than 70 years. Larry Hardesty, *Explained: Neural Networks*, MIT NEWS (Apr. 14, 2017), http://news.mit.edu/2017/explained-neural-networks-deep-learning-0414 [https://perma.cc/QAV9-DWX2].

20. Parloff, *supra* note 17, at 14.

21. Vella, *supra* note 14, at 7.

22. *Id.*

23. Lee Rainie & Janna Anderson, *Code-Dependent: Pros and Cons of the Algorithm Age*, PEW RES. CTR. (Feb. 8, 2017), http://www.pewinternet.org/ 2017/ 02/ 08/code-dependent-pros-and-cons-of-the-algorithm-age/ [https://perma.cc/7CP2-QDJZ].

algorithms.[24] "Algorithms are, in part, our opinions embedded in code."[25] They are "often elegant and incredibly useful tools used to accomplish tasks."[26] These neural networks use "big data," immensely large collected data sets, to analyze and reveal patterns and trends.[27] The development of the internet and advances in computer hardware have allowed programmers to take advantage of the "vast computational power and the enormous storehouses of data—images, video, audio and text files strewn across the internet—that, it turns out, are essential to making neural nets work well."[28]

For deep learning to function, algorithms need to be fed data.[29] Data mining uses algorithms to collect and analyze data.[30] Data mining consolidates massive quantities of data generated on the internet and identifies "interpretable patterns" otherwise too subtle or complex for unaided human discernment.[31] When the data is collected and relationships are identified, it is called a model.[32]

For data mining and deep learning to work, programmers have to translate the problem or desired outcome "into a question about the value of some target variable."[33] Programmers and data miners frequently translate ambiguous problems into questions computers can solve by focusing on the value of a target variable.[34] To create the model, the algorithm is trained to behave in a specific way by the data it is fed.[35] The

24. *Id.*

25. 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 [https://perma.cc/AN4B-RX4B].

26. Rainie & Anderson, *supra* note 23.

27. Vella, *supra* note 14, at 7.

28. Parloff, *supra* note 17, at 11-13.

29. Vella, *supra* note 14, at 7.

30. Alexander Furnas, *Everything You Wanted to Know About Data Mining but Were Afraid to Ask*, ATLANTIC (Apr. 3, 2012), https:// www.theatlantic.com/ technology/ archive/2012/04/everything-you-wanted-to-know-about-data-mining-but-were-afraid-to-ask/255388/ [https://perma.cc/44T9-E7CG].

31. *Id.*

32. Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 677 (2016).

33. *Id.* at 678.

34. *Id.*

35. *Id.* at 680.

kind of data chosen has meaningful consequences for the computer's learning.[36] Labeling "is the process by which . . . training data is manually assigned class labels" by programmers.[37] Sometimes, when a concept is more abstract, like what makes a good employee, programmers and data miners have to make subjective decisions on how to label examples.[38]

The definition of a desirable employee is challenging because it requires prioritization of numerous observable characteristics that make an employee "good."[39] Employers tend to value action-oriented, intelligent, productive, detail-oriented employees.[40] This subjective decision opens the door to potential problems.[41] Essentially, what makes a "good" employee "must be defined in ways that correspond to measurable outcomes: relatively higher sales, shorter production time, or longer tenure, for example."[42] However, the subjective choices made both by the programmers and by the employer in previous hiring decisions are absorbed into the algorithm by way of the data that is used and the subjective labels placed on specific characteristics.[43] Thus, when subjective labels are applied, the results are skewed along the lines of those labels and the data that is utilized.[44] Therefore, it is possible for algorithms and artificial intelligence to inherit prior prejudice and reflect current prejudices.[45]

B. Overarching Societal Implications

At its best, artificial intelligence promotes innovation by increasing efficiency and allowing people to focus on innovation

36. *Id.*
37. Barocas & Selbst, *supra* note 32, at 681.
38. *Id.* at 680.
39. *Id.*
40. Ken Sundheim, *15 Traits Of The Ideal Employee*, FORBES (Apr. 2, 2013, 1:03 AM), https:// www.forbes.com/ sites/ kensundheim/ 2013/ 04/ 02/ 15-traits-of-the-ideal-employee/#9c9c350161f4 [https://perma.cc/QX5B-5B2B].
41. Barocas & Selbst, *supra* note 32, at 679-80.
42. *Id.* at 679.
43. *Id.* at 679-80.
44. *Id.* at 681.
45. *Id.*

instead of mundane tasks.[46] In fact, some liken artificial intelligence to "the new electricity" because it has the power and potential to dramatically transform society in a variety of ways.[47] Artificial intelligence is used to help determine what show to watch next on Netflix, what to listen to next on Spotify, where to go on vacation, and even predict heating and cooling needs in homes.[48]

Unfortunately, the use of algorithms created with good intentions can lead to inadvertent, negative consequences.[49] There are a number of overarching issues with artificial intelligence used to increase efficiency and solve social problems. For example, the choice to use certain data inputs over others can lead to discriminatory outcomes.[50] Without safeguards against poorly designed systems and reckless uses of proxies and data collection, algorithmic flaws could flourish and exasperate existing social divides.[51]

For example, a combination of Facebook likes and digital records of behavior can be used to accurately ascertain a wide range of highly personal, private[52] characteristics, which overlap with protected traits.[53] In a recent study, researchers utilized Facebook likes to predict with a high degree of accuracy sexual

46. Lisa Eadicicco, *He Helped Create the 'Google Brain.' Here's What He Thinks About AI Now*, TIME (Jan. 11, 2017), http://time.com/4631730/andrew-ng-artificial-intelligence-2017/ [https://perma.cc/44HJ-ARJ3]; Forbes Technology Council, *14 Ways AI Will Benefit Or Harm Society*, FORBES (Mar. 1, 2018, 7:00 AM), https://www.forbes.com/sites/forbestechcouncil/2018/03/01/14-ways-ai-will-benefit-or-harm-society/#5bef9f7f4ef0 [https://perma.cc/45RT-RKUE].

47. Eadicicco, *supra* note 46.

48. R.L. Adams, *10 Powerful Examples of Artificial Intelligence In Use Today*, FORBES (Jan. 10, 2017, 8:32 AM), https:// www.forbes.com/ sites/ robertadams/ 2017/ 01/ 10/10-powerful-examples-of-artificial-intelligence-in-use-today/2/#55fee0603c8b [https:// perma.cc/VP72-ZFP2]. Artificial intelligence technology is present in virtual assistants, like Siri and Alexa, in video games, smart cars, targeting ads and purchase prediction, detection of fraud, online customer service, news presentation and generation and enhanced security monitoring.

49. Rainie & Anderson, *supra* note 23.

50. CECILIA MUÑOZ ET AL., EXEC. OFFICE OF THE PRESIDENT, BIG DATA: A REPORT ON ALGORITHMIC SYSTEMS, OPPORTUNITY, AND CIVIL RIGHTS 7 (2016).

51. *See id.* at 9.

52. Data mining and the use of artificial intelligence raises a number of important concerns about privacy. However, these concerns are outside the scope of this article.

53. Michal Kosinski et al., *Private Traits and Attributes are Predictable from Digital Records of Human Behavior,* 110 PNAS 5802, 5802 (2013).

orientation, ethnicity, religious and political views, gender, relationship status, personality traits, intelligence, happiness, use of addictive substances and the marital status of parental figures.[54] Of all of these characteristics, gender and ethnicity achieved the highest level of predictive accuracy.[55] African American and Caucasian users were identified with 95% accuracy.[56] Male and female users were identified with 93% accuracy.[57] This strongly suggests that "patterns of online behavior as expressed by Likes significantly differ between those groups allowing for nearly perfect classification."[58] This is particularly alarming given the ease in which employers could use this kind of data to create proxies for protected characteristics, both intentionally or inadvertently.[59]

Access to technology is also a significant issue. There are concerns that without a holistic look at the ways in which data collection and algorithms impact communities, artificial intelligence technology could reinforce socio-economic divides and inhibit social mobility.[60] The City of Boston implemented a program utilizing an app called Street Bump[61] that detected pot holes with the sensors in the smart phones of Bostonians who downloaded the Street Bump app.[62] As a result, the app directed repair resources to wealthier communities where Bostonians were more likely to own a smartphone.[63]

---

54. *Id.* at 5802-03.

55. *Id.* at 5803.

56. *Id.*

57. *Id.*

58. Kosinski et al., *supra* note 53, at 5803.

59. *See supra* Part II.A.

60. Alexis Stephens, *Big Data Has Potential to Both Hurt and Help Disadvantaged Communities*, NEXT CITY (Sept. 24, 2014), https://nextcity.org/daily/entry/big-data-good-bad-help-disadvantaged-communities [https://perma.cc/L4JE-LN8G].

61. The Street Bump app works by allowing drivers with smart phones to automatically report the presence of potholes to the city. *See About Street Bump*, STREET BUMP, http://www.streetbump.org/about [https://perma.cc/TU3K-WWUN]. In order for the app to work, drivers who are using the app start the app and set their phone on the dashboard or in a cup holder. *Id.* The app then uses the phone's accelerometer and motion detector to sense when a pothole is hit. *Id.* From there, GPS records the location and the app transmits the information to a server which the city uses to determine where to send repair resources. *Id.*

62. Barocas & Selbst, *supra* note 32, at 685.

63. *White House Looks at How 'Big Data' Can Discriminate*, REUTERS (Apr. 26, 2014, 8:02 PM), https://www.reuters.com/article/us-usa-obama-privacy/white-house-looks-

Then, there is the overarching societal problem of prejudice. Artificial intelligence and algorithms rely on training data. When these data sets are skewed as a result of bias or carelessness, the results can be discriminatory.[64] Many experts argue that the use of algorithms and artificial intelligence perpetuate socio-economic divides and promulgate existing inequalities[65] and "[t]o paraphrase Immanuel Kant, out of the crooked timber of these datasets no straight thing was ever made."[66]

C. Artificial Intelligence in Hiring

There are a number of companies that have developed artificial intelligence technology specifically for employers to use in hiring.

First, there are companies using bots,[67] like Mya. Mya "automates the process from resume to hire."[68] Mya uses bots to chat with applicants using natural language processing[69] to interact with candidates and screen them for open job positions.[70] The experience of chatting with Mya is similar to a text message exchange.[71] Mya is the first AI system to

---

at-how-big-data-can-discriminate-idUSBREA3Q00M20140427 [https://perma.cc/88MB-WGWA].

64. Barocas & Selbst, *supra* note 32, at 683-84.

65. Rainie & Anderson, *supra* note 24.

66. *Id.*

67. "[A] bot is "an application that performs an automated task." Sarah Mitroff, *What is a Bot? Here's Everything You Need to Know*, CNET (May 5, 2016, 3:23 PM), https://www.cnet.com/how-to/what-is-a-bot/ [https://perma.cc/BG2Q-9D88]. Bots are often programmed to communicate like humans by way of natural language processing. *Id.* They are found in a variety of both positive and negative contexts, from helping order pizza to spreading viruses online. *Id.'*

68. MYA, https://hiremya.com/ [https://perma.cc/ZTP6-FFA6].

69. Natural language processing is a component of machine learning that allows the bot to understand and "interpret input and produce output in the form of human language." Henk Pelk, *Machine Learning, Neural Networks and Algorithms*, CHATBOTS MAG. (Feb. 16, 2017), https:// chatbotsmagazine.com/ machine- learning- neural- networks- and-algorithms-5c0711eb8f9a [https://perma.cc/S5YQ-UX9T].

70. MYA, *supra* note 68.

71. Ryan Prior, *Your Next Job Interview Could Be with a Recruiter Bot*, CNN TECH. (May 16, 2017, 9:19 AM), http://money.cnn.com/2017/05/16/technology/ai-recruiter-mya-systems/index.html [https://perma.cc/ZS56-A54D].

interview job candidates,[72] but there are other bots that focus on personal resume marketing,[73] and product management.[74]

Other companies like ARYA use algorithms, machine learning, big data and behavioral pattern recognition to identify and isolate ideal candidates.[75] ARYA pushes personalized, suggested messages to recruiters to help them leverage their impact on quality candidates.[76] Unlike Mya, ARYA does not interview candidates, but sorts profiles to identify candidates that, in theory, will be most successful.[77]

HireVue uses artificial intelligence to analyze candidates' diction, tone, and facial movement in video job interviews.[78] HireVue uses voice recognition and facial recognition software in conjunction with a ranking algorithm to determine which candidates resemble "the ideal candidate."[79] After the algorithm informs the recruiter which candidates have excelled in the video interview, the recruiter can focus on those candidates.[80]

Finally, there are companies like Pymetrics. Founded by Dr. Frida Polli and Dr. Julie Yoo,[81] Pymetrics uses brain games

---

72. *Id.*

73. Esther Crawford turned her resume into an interactive bot that potential employers could communicate with. Esther Crawford, *How I Turned My Resume into a Bot. (And How You Can Too!)*, MEDIUM (Apr. 17, 2016), https://medium.com/the-mission/how-i-turned-my-resume-into-a-bot-and-how-you-can-too-f03847352baa [https://perma.cc/QGU6-Y6UZ].

74. TARA, https://tara.ai/ [https://perma.cc/36FE-CBHZ].

75. ARYA, https://goarya.com/solutions/ [https://perma.cc/NGR7-F948].

76. *Id.*

77. *See id.*

78. Joe Avella & Richard Feloni, *We Tried the AI Software Companies Like Goldman Sachs and Unilever Use to Analyze Job Applicants*, BUS. INSIDER (Aug. 29, 2017, 5:39 AM), http://www.businessinsider.com/hirevue-uses-ai-for-job-interview-applicants-goldman-sachs-unilever-2017-8 [https://perma.cc/GJ54-GVSJ].

79. Richard Feloni, *I Tried the Software that Uses AI to Scan Job Applicants for Companies Like Goldman Sachs and Unilever Before Meeting Them – and It's Not As Creepy As It Sounds'*, BUS. INSIDER (Aug. 23, 2017, 12:00 PM), http:// www.businessinsider.com/ hirevue- ai- powered- job- interview- platform- 2017- 8 [https://perma.cc/M2DT-2WZ4].

80. *Id.*

81. Dr. Frida Polli is a Harvard and MIT trained neuroscientist. *Founding Story*, PYMETRICS, https://www.pymetrics.com/about/ [https://perma.cc/N6RU-NGLR]. Dr. Julie Yoo is a former postdoctoral neuroscientist at MIT and the Department of Defense. *Julie Yoo: Founder & Chief Data Scientist*, WOMEN DATA SCI. CONF. (2018), http://www.widsconference.org/julie-yoo.html [https://perma.cc/WZ6C-E4E9]. Dr. Yoo

and artificial intelligence to remove biases such as classism, racism, sexism, and ageism.[82] Pymetrics primarily works with large companies because the software needs a substantial amount of data to develop algorithms with a higher degree of accuracy.[83] To create Pymetrics's algorithm, a company's top 100-150 performers play twelve neuroscience games.[84] For example, the game that assesses risk aversion "gives users three minutes to collect as much 'money' as possible using the following system[:] Clicking 'pump' inflates a balloon by 5 cents; at any point, the user can click 'collect money.' If the balloon pops, the user receives no money."[85]

After the top performers finish the games, Pymetrics creates a customized algorithm that generates a portrait of an ideal candidate.[86] Candidates play the same games, and recruiters compare the results.[87] According to Dr. Polli, "[t]he resume is the most biased piece of information used in the hiring process."[88] The system created by Pymetrics bypasses resume review, and does not account for the candidate's ethnicity, educational background, referrals, or gender.[89] The goal is to create a process that "doesn't preference white guys from elite schools who were on the sailing team just like the recruiter."[90] Recognizing that computers are likely to reflect the same gender and ethnic biases present in society, Pymetrics adjusts its algorithms for each company and creates a reference list of 10,000 people who have used Pymetrics.[91] Pymetrics knows the

---

specialized in using machine learning to "predict optimal learning time based on real-time neuroimaging data, and building automatic speech recognition systems." *Id.*

82. Leanna Garfield, *A Startup Claims to Have Finally Figured Out How to Get Rid of Bias in Hiring with Artificial Intelligence*, BUS. INSIDER (Sept. 25, 2017, 11:20 AM), http://www.businessinsider.com/hiring-diversity-brain-games-artificial-intelligence-automation-2017-9 [https://perma.cc/VML5-SVTK].

83. *Id.*

84. *Id.*

85. *Id.*

86. *Id.*

87. Josh Constine, *Pymetrics Attacks Discrimination in Hiring with AI and Recruiting Games*, TECHCRUNCH (Sept. 20, 2017), https:// techcrunch.com/ 2017/ 09/ 20/ unbiased- hiring/ [https://perma.cc/L2GG-4VYN].

88. *Id.*

89. Garfield, *supra* note 82.

90. Constine, *supra* note 87.

91. *Id.*

gender and ethnicities of the reference group, and if they notice "for example, that men are receiving higher scores than women on a given trait," they can adjust the model to correct the disproportionate result.[92]

D. Lack of Meaningful Diversity in Silicon Valley

Perhaps the most significant, overarching problem is the severe lack of diversity in tech. Although business leaders in the United States have tended to focus on remedying this issue by pushing for more relaxed standards on H-1B visas to produce a more diverse workplace, there is plenty of information to suggest that the issue may not be a lack of talented, capable American female and minority candidates.[93] Instead, the problem appears to be the inability of technology companies to attract and retain talented women and minority candidates.[94]

Around nine percent of graduates from highly regarded computer science programs are from "under-represented minority groups."[95] Meanwhile, only five percent of the workforce in the technology industry are from one of the underrepresented groups.[96] Forty-one percent of highly qualified scientists and engineers are women.[97] However, what is perhaps the most alarming is that, over time, fifty-two percent of these women leave their jobs, usually in their mid-thirties.[98]

There seem to be a number of key reasons for this mass exodus. First, around two-thirds of women report having to prove themselves "over and over again," having their successes undercut and abilities consistently questioned.[99] Black women

---

92. Garfield, *supra* note 82.

93. Sylvia Ann Hewlett et al., *Stopping the Exodus of Women in Science*, HARV. BUS. REV., June 2008, at 22, 22.

94. *Id.* at 22-23.

95. EEOC, DIVERSITY IN HIGH TECH 7 (2016), https:// www.eeoc.gov/ eeoc/ statistics/ reports/hightech/upload/diversity-in-high-tech-report.pdf [https://perma.cc/F29Z-4AZJ].

96. *Id.*

97. Hewlett et al., *supra* note 93, at 22.

98. *Id.* at 23.

99. Joan C. Williams, *The 5 Biases Pushing Women Out of STEM*, HARV. BUS. REV. (Mar. 24, 2015), https://hbr.org/2015/03/the-5-biases-pushing-women-out-of-stem [https://perma.cc/L76M-4P3G].

experience this at a higher rate than Latina, Asian, and White women.[100] Second, women "need to behave in masculine ways in order to be seen as competent—but women are expected to be feminine."[101] Black and Latina women are more acutely at risk of being seen as "angry" for failing to conform to traditional gender roles.[102] Third, two-thirds of female scientists face questions about their commitment to their careers after starting a family.[103] Specifically, these female scientists faced an assumption "that your career is more of a hobby than a career, and you're only going to do it until you find a husband and/or have a family."[104] Fourth, although not completely unavoidable, gender bias tends to cause conflict between women.[105] While three-fourths of women reported a supportive work environment among their female colleagues, a fifth of female scientists surveyed felt that they were competing for the "woman" spot.[106] Finally, women feel the need to isolate themselves from other colleagues to be taken seriously.[107]

Timnit Gebru, a Ph.D. candidate at Stanford, attended an important artificial intelligence conference. She looked around the room and counted six black people in the entire audience, and realized she was the only black woman in attendance.[108]

The workforce in tech is predominantly white, Asian, and male.[109] Many in the industry argue that "gender and racial bias

100. *Id.* 77% of Black women reported having to provide "more evidence of competence than others to prove themselves," while 65% of Latina women, 64% of Asian women and 63% of white women reported this. *Id.*

101. *Id.*

102. *Id.*

103. *Id.*

104. Williams, *supra* note 99.

105. *Id.*

106. *Id.*

107. *Id.*

108. Mariya Yao, *Fighting Algorithmic Bias and Homogenous Thinking in A.I.*, FORBES (May 1, 2017, 12:02 PM), https:// www.forbes.com/ sites/ mariyayao/ 2017/ 05/ 01/ dangers- algorithmic- bias- homogenous- thinking- ai/#7c086b8070b3 [https://perma.cc/TTJ4-M2U6].

109. Bonnie Marcus, *The Lack of Diversity in Tech is a Cultural Issue*, FORBES (Aug. 12, 2015, 8:48 AM), https://www.forbes.com/sites/bonniemarcus/2015/08/12/the-lack-of-diversity-in-tech-is-a-cultural-issue/#961969c79a21 [https:// perma.cc/6C9N-KXKT]. While many in the technology industry point to a "pipeline problem," there is a growing mass of research suggesting that the dominance of white and Asian men in the technology industry is the result of stereotyping and bias which results in abysmal

is so ubiquitous in the technology industry that it forces talented female and minority employees to leave."[110] Women in the United States are considerably more likely than men to state that gender discrimination exists in the technology industry.[111] Additionally, Black and Hispanic people are much more likely than white people to say that there is more discrimination in the tech industry than other industries.[112]

Those involved in the research and development of artificial intelligence "pride themselves on being rational and data-driven, but can be blind to issues such as racial or gender bias that aren't always easy to capture with numbers."[113] Unfortunately, while people in the tech industry pay lip service to the importance of diversity, there is little meaningful change.[114]

Algorithms are in large part "our opinions embedded in code."[115] Research shows that applying machine learning to everyday human language reproduces existing societal bias.[116] This is a particularly significant problem considering the alarming lack of diversity in the tech industry, the authors of these algorithms.

For example, word embeddings in computer programming "know" that flowers are pleasant, and insects and weapons are not, based only on exposure to human language.[117] By this same

---

participation of African Americans, women and other minority groups. For a general discussion of this research, see generally EEOC, *supra* note 95.

110. Marcus, *supra* note 109. Research suggests that about 50% of women in the technology industry are abandoning their careers as scientists, engineers, programmers and technologists. Hewlett et al., *supra* note 93. In fact, there seems to be "a key moment in women's lives—in their mid to late thirties—when most head for the door." *Id.* Recent research suggests that "bias, not pipeline issues or personal choices, pushes women out of science—and that bias plays out differently depending on a woman's race or ethnicity." Williams, *supra* note 99.

111. Kim Parker & Cary Funk, *Women are More Concerned Than Men About Gender Discrimination in Tech Industry*, PEW RES. CTR. (Oct. 10, 2017), http://www.pewresearch.org/fact-tank/2017/10/10/women-are-more-concerned-than-men-about-gender-discrimination-in-tech-industry/ [https://perma.cc/RL2C-X2R4].

112. *Id.*

113. Yao, *supra* note 108.

114. *Id.*

115. Mann & O'Neil, *supra* note 25.

116. Aylin Caliskan et al., *Semantics Derived Automatically from Language Corpora Contain Human-Like Biases* 356 SCIENCE 183, 183 (2014).

117. *Id.*

method, researchers discovered "extreme effects of race as indicated simply by name."[118] A group of names typically associated with being European-American were found to be "significantly more easily associated with pleasant than unpleasant terms, compared with a bundle of African-American names."[119]

Data training and data labeling have already demonstrated discriminatory impact outside of the hiring sphere. In online searches of names "racially associated" with the Black community, it was twenty-five percent more likely that ads would appear suggesting the person had a criminal record.[120] When Latayna Sweeney, the first Black woman to receive a Ph.D. in computer science at MIT and current Harvard Professor, googles her own name, she comes across ads like: "Latanya Sweeney, Arrested? 1) Enter name and state 2) Access full background. Checks instantly."[121] This presents a serious problem for minority candidates applying for jobs or competing for promotions. When a minority candidate is googled, it is more likely that, alongside her list of accomplishments, there will be an advertisement suggesting that she has a criminal record, whether or not she actually has one.[122] Even more damaging, the ads might not appear on searches conducted of competitors' names.[123]

E. Public Perception

As a whole, Americans do not have a positive perception of the use of artificial intelligence in hiring.[124] The vast majority of

118. *Id.*

119. *Id.*

120. Yao, *supra* note 108.

121. *Id.*

122. Latanya Sweeney, *Discrimination in Online Ad Delivery*, COMMC'NS ACM, May 2013, at 44, 46-47.

123. *Id.* at 44.

124. Only three percent of Americans are "very" enthusiastic about the implementation of artificial intelligence technology in hiring. Aaron Smith & Monica Anderson, *Automation in Everyday Life'*, PEW RES. CTR. (Oct. 4, 2017), http:// www.pewinternet.org/ 2017/ 10/ 04/ americans- attitudes- toward- hiring-algorithms/ [https://perma.cc/RQC2-96KN]. Meanwhile, twenty-one percent of Americans are very "worried" about artificial intelligence technology's use in hiring. *Id.*

Americans are completely unfamiliar with the "concept of using computer-generated algorithms to analyze job candidates."[125] Most Americans would not want to apply for a job knowing that a computer program or an algorithm would be evaluating candidates.[126] The American public generally feels that algorithms would do a worse job than humans in several areas of the hiring process.[127]

III. ARTIFICIAL INTELLIGENCE & TITLE VII DISPARATE IMPACT LIABILITY

Title VII liability is divided into two types of claims: disparate treatment claims and disparate impact claims.[128] Disparate treatment claims arise when an employer intentionally discriminates on the basis of a protected characteristic.[129] Disparate impact claims occur when an employer utilizes a practice that appears to be neutral, but in reality, has a discriminatory effect on the basis of a protected characteristic.[130] Disparate impact claims originated in *Griggs v. Duke Power Co.*, where the Supreme Court held that Title VII "proscribes not only overt discrimination but also practices that are fair in form, but discriminatory in operation."[131] The Court's analysis of Title VII was entrenched in a desire to carry out the intent of Congress in enacting Title VII, which was to eradicate "preference for any group, minority or majority."[132] Disparate

---

125. *Id.*

126. *Id.*

127. Seventy-six percent of Americans stated that they "would not want to apply for a job knowing that a computer program would be" utilized to make a hiring decision. *Id.* Of this group, forty-one percent stated that this was because "[c]omputers can't capture everything about an applicant," twenty percent stated that the process was "[t]oo impersonal," four percent stated that "[a]pplicants can game [the] system," and two percent stated that the process was more biased than current hiring practices. *Id.*

128. Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. 857, 902-03 (2017).

129. *Id.* at 903.

130. *Id.*

131. 401 U.S. 424, 431 (1971) (holding that the employer was not permitted to require applicants to hold a high school diploma or specific score on a written test because neither was shown to be significantly related to job performance).

132. *Id.*

impact claims were refined over the next twenty years, eventually codified in the Civil Rights Act of 1991.[133]

A. Making a Prima Facie Case

Disparate impact liability exists when a plaintiff demonstrates that an employer has implemented a practice that produces an adverse effect on the basis of a protected characteristic, such as race or gender.[134] To establish causation, a plaintiff compares "the selection rates of majority and minority applicants for a position and then showing that the disparity is statistically significant or that it violates the four-fifths rule."[135]

Statistical significance seems to be the most common, widely accepted method of proof.[136] Tests that determine statistical significance indicate the level of mathematical certainty that can lead to a conclusion that the practice causes a disparate impact.[137] In practice,

> Researchers most commonly use the ninety-five percent confidence level, which is also termed the five percent (0.05) level of significance. . . . At the ninety-five percent level, we can be ninety-five percent certain that the observed disparity in the applicant pool reflects a real disparity in the relevant labor market with respect to the challenged practice. There is still, however, a one in twenty possibility that there is no disparity in the overall population.[138]

While these numbers can be enlightening, they ultimately indicate that it is not likely the disparity in the labor market is the product of mere chance.[139] It does not provide evidence that the challenged employment practice is the cause of the disparity in question.[140] In other words, a statistically significant

133. Kim, *supra* note 128, at 905.

134. 42 U.S.C. § 2000e-2(k)(1)(A)(i) (2012).

135. Jennifer L. Peresie, *Toward a Coherent Test for Disparate Impact Discrimination*, 84 IND. L.J. 773, 777 (2009).

136. *See id.* at 777, 785.

137. *Id.* at 785.

138. *Id.* at 785-86.

139. *Id.* at 786.

140. Peresie, *supra* note 135, at 786.

difference "strongly indicates *some* influence on the results other than the operation of pure chance."[141]

There is no clearly established level of statistical significance, and this has resulted in criticism and confusion.[142] Statistical significance is also sensitive to sample size, with smaller sample sizes substantially more likely to return a finding than larger sample sizes.[143]

A plaintiff may also demonstrate causation by satisfying the four-fifths rule. The Equal Employment Opportunity Commission adopted this standard in the immediate aftermath of *Griggs*.[144] To bring a disparate impact claim,

> A selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded by Federal enforcement agencies as evidence of adverse impact.[145]

If the lesser represented group has a selection rate of less than eighty percent (or four-fifths) of selection rate for the most represented group, it is evidence of adverse impact.[146] The Supreme Court has noted that the four-fifths rule is a "rule of thumb" but deserves respect from courts.[147]

This approach has a number of advantages, namely its simplicity and the fact that it puts employers on notice about the balance they must maintain to avoid litigation.[148] However, this potentially points to the existence of an acceptable level of

---

141. Carpenter v. Boeing Co., 456 F.3d 1183, 1202 (10th Cir. 2006).

142. The statistical significance test has faced criticism from scholars because of a lack of clearly established level of statistical significance, and because the statistical significance test is incredibly sensitive to sample size. Peresie, *supra* note 135, at 786. In other words, the "larger the number of applicants, the smaller the magnitude of difference that will be statistically significant (at whatever level is selected)." *Id.* at 787.

143. *Id.* at 787.

144. *Id.* at 781.

145. 29 C.F.R. § 1607.4(D) (2017).

146. Peresie, *supra* note 135, at 781.

147. *See, e.g,* Fed. Express Corp. v. Holowecki, 552 U.S. 389, 399 (2008).

148. Peresie, *supra* note 135, at 783.

discrimination, and disproportionally burdens small employers.[149]

To establish a claim involving artificial intelligence, a plaintiff would likely still need to utilize either the statistical significance test or the four-fifths rule.[150] Given the data-driven nature of artificial intelligence, it would not seem that this would be a difficult threshold to meet. Biased data labeling and poor selection of target variables can result in the kind of statistical evidence that would make it relatively easy for a plaintiff to establish a disparate impact claim.

St. George's Hospital, a medical school in the United Kingdom, developed a computer program in the 1980's to simplify the process of sorting through applicants.[151] The computer program was created after carefully reviewing previous admission decisions.[152] Interestingly, the program did not introduce new biases into the system, but simply replicated and reflected existing bias at St. George's by setting these biases as target variables.[153] Although some of the racial and gender related issues were fairly obvious in the training data, some were less apparent.[154] For example, "[a] good number of the applications with foreign names, or from foreign addresses, came from people who clearly had not mastered the English language. Instead of considering the possibility that great doctors could learn English, which is obvious today, the tendency was simply to reject them."[155] This program eventually resulted in around sixty applicants per year being refused an interview for admission on the basis of their gender or ethnicity.[156]

However, establishing a prima facie case for disparate impact becomes incredibly challenging when the discrimination

---

149. *Id.*

150. *See supra* Part II.A.

151. Stella Lowry & Gordon Macpherson, *A Blot on the Profession,* 296 BRITISH MED. J. 657, 657 (1988).

152. *Id.*

153. *Id.*

154. CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 116 (2016).

155. *Id.*

156. *See id.* at 117.

is the result of incomplete, incorrect, or non-representative data.[157] Even if a data set is relatively complete, it may be problematic because it fails to represent groups in accurate proportions.[158] As an example, data from Twitter suggests that people are happier when they are away from home, and they are saddest on Thursday nights.[159] However, this is actually not as conclusive as it seems because, as of April 4, 2016, only twenty-one percent of American adults use Twitter.[160]

"[N]ot all data is "created or even collected equally," and as a result, "there are 'signal problems' in big-data sets—dark zones or shadows where some citizens and communities are overlooked or underrepresented."[161] It is incredibly difficult to use statistics to demonstrate the existence of discrimination when a protected class is not even represented in the data set to begin with. In an employment context, segments of protected classes could be excluded from employment opportunities because of a lack of access to the required technology to participate in the hiring practices that use artificial intelligence, similar to what was seen in Boston with the Street Bump app.[162]

B. Business Necessity

From this point, the burden shifts to the employer to rebut the plaintiff's *prima facie* case by showing that the practice is job-related and connected to a business necessity.[163] In *Griggs*, the Supreme Court articulated that in enacting Title VII, Congress "placed on the employer the burden of showing that any given requirement must have a manifest relationship to the

157. Barocas & Selbst, *supra* note 32, at 684.

158. *Id.*

159. Kate Crawford, *Think Again: Big Data*, FOREIGN POL'Y (May 10, 2013, 12:40 AM), http://foreignpolicy.com/2013/05/10/think-again-big-data/ [https://perma.cc/6PPX-C4MK].

160. *Social Media Fact Sheet*, PEW RES. CTR. (Feb. 5, 2018), http://www.pewinternet.org/fact-sheet/social-media/ [https://perma.cc/A3NR-LBQE].

161. Crawford, *supra* note 159.

162. *See supra* Part II.B.

163. Albemarle Paper Co. v. Moody, 422 U.S. 405, 425 (1975) (holding that the burden of showing that the given requirement has a relationship to employment "arises, of course, only after the complaining party or class has made a prima facie case of discrimination.").

employment in question."[164] This seems simple, but there does not appear to be a settled definition of what actually suffices as a business necessity.[165] In *Griggs*, the court focused on the requirement set out by employers and its relationship to performance of the job.[166]

Initially intended as a narrow defense, the Court expanded its meaning,[167] eventually expanding it so far in *Wards Cove Packing Co. v. Atonio* that Congress responded with the Civil Rights Act of 1991.[168] At this point, it seems that a business necessity "lies somewhere in the middle of two extremes" seen in *Griggs* and *Wards Cove*, but different courts have applied the standard in different ways.[169]

In the context of artificial intelligence, the heart of the issue seems to be "whether the sought-after trait—the target variable—is job related, regardless of the machinery used to predict it."[170] If the prioritized trait is not related to the job, then the business necessity defense will fail regardless.[171] However, if the target variable is related to job performance, then it must be determined if the model is actually predictive of the job related trait.[172] Artificial intelligence algorithms are prognostic

---

164. Griggs v. Duke Power Co., 401 U.S. 424, 432 (1971).

165. Barocas & Selbst, *supra* note 32, at 705.

166. *See Griggs*, 401 U.S. at 432-33 (noting that "Congress has placed on the employer the burden of showing that any given requirement must have a manifest relationship to the employment in question. . . . Diplomas and tests are useful servants, but Congress has mandated the commonsense proposition that they are not to become masters of reality").

167. In *New York City Transit Authority v. Beazer*, the Court expanded the business necessity doctrine by allowing the implementation of a "narcotics rule" which was connected with the safety of maintaining the transit system, despite the fact that 25% of jobs in the transit system were related to safety. 440 U.S. 568, 587-90 (1979); *see also* Barocas & Selbst, *supra* note 32, at 703.

168. In *Wards Cove*, the Court essentially "reallocated the burden to plaintiffs to prove that business necessity was lacking and even referred to the defense as a 'business justification' rather than a business necessity." Barocas & Selbst, *supra* note 32, at 703; *see* Wards Cove Packing Co., Inc. v. Atonio, 490 U.S. 642, 659 (1989) (noting that "there is no requirement that the challenged practice be 'essential' or 'indispensable' to the employer's business for it to pass muster: this degree of scrutiny would be almost impossible for most employers to meet, and would result in a host of evils we have identified above.").

169. Barocas & Selbst, *supra* note 32, at 704.

170. *Id.* at 706.

171. *Id.*

172. *Id.* at 706-07.

by nature, making their selections necessarily job related.[173] However, given the complex nature of data mining and algorithmic construction, it is difficult to even determine which characteristics are being legitimately targeted as job related, and which ones are being used as proxies[174] for protected characteristics.[175]

If the algorithm is complex, as most are, a claimant would have to be able to examine the training data and model, and determine how the data was collected.[176] This is unreasonable and unattainable, given the extensive resources this would require, not to mention potential issues with intellectual property contained in the algorithms.[177] This, coupled with the innately predictive nature of hiring algorithms, make it highly likely that an employer would succeed on a business necessity defense.[178]

This has led some to argue that, in the context of hiring, "employers should bear the burden of establishing the model's validity."[179] The employer's business necessity defense is at its core an assertion that the algorithm it uses accurately predicts future job performance.[180] If an employer is going to assert that

173. *Id.* at 706.

174. This raises disparate treatment concerns. However, given the complex nature of algorithms and data collection, it would likely be difficult to demonstrate intent. This difficulty points to disparate impact as the most realistic vehicle to pursue a discrimination claim.

175. Kim, *supra* note 128, at 920.

176. *Id.*

177. There are a number of potential issues tied up in the tension between the potential demands for more transparency in the development of algorithms, intellectual property concerns, and privacy concerns. However, they are outside the scope of this Comment.

178. Barocas & Selbst, *supra* note 32, at 709.

179. The argument that an employer should have to establish the validity of the model is based on the lack of transparency and the complexity of the algorithms used by employers. Essentially,

> . . . because the employer's justification for using an algorithm amounts to a claim that it actually predicts something relevant to the job, the employer should carry the burden of demonstrating that statistical bias does not plague the underlying model. In other words, the employer should have to defend the accuracy of the correlations it relies on by showing that no problems exist with the data or model construction . . . and not simply by showing a statistical correlation in the existing data.

Kim, *supra* note 128, at 921.

180. *Id.*

the targeted characteristic is job related, "the employer should have to defend the accuracy of the correlations it relies on by showing that no problems exist with the data or model construction that are biasing the results, and not simply by showing a statistical correlation in the existing data."[181] Given all of this, if an employer can properly connect its employment algorithm to a legitimate job-related purpose, it is highly likely that it will meet the burden required by the business necessity defense.[182]

C. Alternative Employment Practice

Even if the employer demonstrates that the practice is related to a business necessity, the plaintiff can still prevail if she can demonstrate the existence of a less discriminatory alternative.[183] In the Civil Rights Act of 1991, Congress codified this as the "alternative employment practice"[184] requirement. However, Congress did not give the phrase a clear meaning.[185] The first case to use the phrase "alternative employment practice" was *Wards Cove Packing Co. v. Antonio*.[186] This is curious because Congress expressly rejected the Supreme Court's holding in *Wards Cove* when enacting the Civil Rights Act of 1991.[187] Despite mimicking the language from *Wards Cove,* the instruction from Congress seems to point to the standard articulated by *Albemarle Paper Co. v. Moody*, the case that originally established the three step, burden-shifting framework.[188]

---

181. *Id.*

182. Barocas & Selbst, *supra* note 32, at 709.

183. *Id.*

184. 42 U.S.C. § 2000e-2(k)(1)(A)(ii) (2012).

185. Barcoas & Selbst, *supra* note 32, at 705-06.

186. Michael J. Zimmer, *Individual Disparate Impact Law: On the Plain Meaning of the 1991 Civil Rights Act*, 30 LOY. U. CHI. L.J. 473, 484 (1999).

187. *Wards Cove* was decided on June 5, 1989. In the Civil Rights Act of 1991, Congress explicitly stated that "The demonstration referred to by subparagraph (A)(ii) shall be in accordance with the law as it existed on June 4, 1989, with respect to the concept of 'alternative employment practice.'" 42 U.S.C. § 2000e-2(k)(1)(C) (2012).

188. *See* Barocas & Selbst, *supra* note 32, at 706; Zimmer, *supra* note 186, at 477-78.

This interpretation would allow "the complaining party to "show that other tests or selection devices . . . would also serve the employer's legitimate interest in 'efficient and trustworthy workmanship.'"[189] This opportunity for the plaintiff to rebut business necessity seems to make up for the weakness resulting from the ambiguity of business necessity.[190]

Even if an employer is able to demonstrate the existence of a business necessity and the target variable is job related, the plaintiff can still demonstrate the existence of a less discriminatory alternative.[191] Operating under the *Albemarle* standard, the plaintiff must demonstrate that a less discriminatory alternative exists that would serve the employer's legitimate, job related business interest.[192] In dealing with artificial intelligence analytics, this becomes challenging.

First, because of the enigmatic nature of artificial intelligence, it can be difficult to determine exactly what is being targeted in an algorithm.[193] If an employer fails to effectively disclose or defend the validity of its algorithm and data collection, the alternative employment practices loses all of its "teeth"[194] and the plaintiff is hamstrung.[195]

Second, there is no clear definition of what it means to "refuse" to adopt an alternative employment practice.[196] This is particularly problematic given the expense required to produce and implement artificial intelligence programs.[197] While larger companies might have sufficient resources to correct an algorithm that produces discriminatory results, the burden on small employers may be severe.[198] Given that the statistical significance test and the four-fifths rule tend to more harshly affect smaller employers, this is likely a group that will be impacted by these suits. If a smaller employer cannot afford to

189. Albemarle Paper Co. v. Moody, 422 U.S. 405, 425 (1975) (quoting McDonnell Douglas Corp. v. Green, 411 U.S. 792, 801 (1973)).

190. *See* Barocas & Selbst, *supra* note 32, at 706.

191. *Infra* Part III.A.

192. Zimmer, *supra* note 186, at 477-78.

193. *See* Barocas & Selbst, *supra* note 32, at 706.

194. *Id.* at 706

195. *See* Kim, *supra* note 128, at 921.

196. Barocas & Selbst, *supra* note 32, at 710.

197. *Id.*

198. *Id.*

retool an algorithm, does this mean that it refused to adopt a less discriminatory alternative?[199] It is unclear the role cost might play in what constitutes a refusal to adopt a less discriminatory alternative.[200]

D. Ricci v. Destefano

Avoiding disparate impact liability became more complicated for employers when the Supreme Court decided *Ricci v. DeStefano.*[201] In *Ricci*, white and Hispanic firefighters in New Haven, Connecticut sued the City of New Haven regarding the city's decision not to certify a test needed for promotion to Lieutenant and Captain.[202]

The City's civil service board did not certify the test results because the results would have resulted in the promotion of a disproportionate number of white candidates compared to minority candidates.[203] The disparate treatment suit brought by the white and Hispanic firefighters alleged the city of New Haven discriminated against them on the basis of race by disregarding the test results that would have resulted in their promotion.[204]

---

199. *Id.*

200. *Id.* at 710-11. Some have argued that the Supreme Court's use of "efficient" in *Albemarle* "strongly supports using costs as a factor in analyzing a proposed alternative employment practice" and that the use of cost as a consideration in both business necessity and alternative employment practice is supported by lower court precedent. Ernest F. Lidge III, *Financial Costs as a Defense to an Employment Discrimination Claim*, 58 ARK. L. REV. 1, 32-39 (2005). Ultimately, a consideration of cost may entail an analysis of undue burden under the Americans with Disabilities Act of 1990 (ADA). The ADA prohibits employers from discriminating on the basis of disability and requires them to provide reasonable accommodations to employees with disabilities as long as those accommodations do not constitute an undue burden. 42 U.S.C. § 12112(5)(A) (2012). An undue burden is statutorily defined as "an action requiring significant difficulty or expense" in light of a number of factors, including "the nature and cost of the accommodation needed under this chapter." 42 U.S.C. § 12111(10) (2012). For a discussion of the role of undue hardship and cost, see generally Mark C. Weber, *Unreasonable Accommodation and Due Hardship,* 62 FLA. L. REV. 1119 (2010).

201. 557 U.S. 557 (2009).

202. *See id.* at 562.

203. *See id.*

204. *Id.* at 562-63.

The Supreme Court held that a "race-based action like the City's in this case is impermissible under Title VII unless the employer can demonstrate a strong basis in evidence that, had it not taken the action, it would have been liable under the disparate-impact statute."[205] The Court rejected the City's argument that its good faith belief that using the exams would result in disparate impact liability justified disregarding the test results.[206]

In the dissent, Justice Ginsburg argued that this decision would likely "not have staying power"[207] and that the majority "sets at odds [with] the statute's core directives" because the "characterization of an employer's compliance-directed action shows little attention to Congress's design or to the *Griggs* line of cases Congress recognized as pathmarking."[208]

In response to Justice Ginsburg's dissent, Justice Alito authored a concurring opinion, emphasizing the personal sacrifices made by the white firefighters.[209] Some have argued that a reading of *Ricci* suggests disparate treatment occurred because of the presence of victims.[210] This interpretation, it is argued, appears most consistent with the language of the statute and opinions.[211] In this line of thinking, the text of Title VII "does not forbid any employer decision just because it is made with an awareness of race. Instead, it forbids 'adverse employment actions' taken 'because of an individual's race.'"[212]

There is a current debate among scholars about how to deal with the Supreme Court's holding in *Ricci* and concerns about predictive algorithms.[213] In *Accountable Algorithms*, Kroll et al. contend that *Ricci* ultimately demonstrates the existing tensions between the disparate impact and disparate treatment

205. *Id.* at 563.
206. *Ricci*, 557 U.S. at 606.
207. *Id.* at 609 (Ginsburg, J., dissenting).
208. *Id.* at 624-25 (Ginsburg, J., dissenting).
209. *Id.* at 607 (Alito J., concurring).
210. Kim, *supra* note 128, at 930.
211. *Id*.
212. *Id*.
213. *See* Pauline T. Kim, *Auditing Algorithms for Discrimination* 166 U. PA. L. REV. ONLINE 189, 190 (2017); Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 694 (2017).

doctrines.[214] Kroll et al. claim that under *Ricci*, auditing a predictive algorithm for discriminatory outcomes and making changes based on findings necessarily opens an employer up to disparate treatment liability.[215] In essence, if an employer utilizes an algorithm that creates a disparate impact, an audit and subsequent corrections "will trigger the same kind of analysis as New Haven's rejection of its firefighter test results."[216] As a result, Kroll et al. argue that the legal challenges presented by the holding in *Ricci* necessitate greater emphasis on the design and construction of predictive algorithms.[217] The safest way to proceed is to incorporate "nondiscrimination in the initial design of algorithms."[218]

An audit is the most pervasive social science method to uncover discriminatory practices.[219] Audit studies are normally "field experiments in which researchers or their confederates participate in a social process that they suspect to be corrupt in order to diagnose harmful discrimination."[220] Despite the fact that auditing carries a financial connotation, audits were originally developed as a way to identify racial discrimination in housing, and "[a]lthough the word 'audit' has a similar dictionary meaning in both cases, the 'audit study' as it evolved in social science is distinct from financial auditing."[221] Typically, social science audits either take the form of an audit study or a correspondence study.[222] However, scholars argue that audits of algorithms may require methods that differ from traditional social science audit methods.[223]

---

214. Kroll et al., *supra* note 213, at 694.

215. *See id.* at 694-95.

216. *Id.*

217. *Id.*

218. *Id.* at 695.

219. Christian Sandvig et al., Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms 5 (May 22, 2014) (unpublished paper), http://www-personal.umich.edu/ ~csandvig/ research/ Auditing% 20Algorithms% 20— %20Sandvig%20— %20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf [https://perma.cc/JK29-5G8F].

220. *Id.*

221. *Id.* at 6.

222. *Id.*

223. For a more complete discussion of potential methods of auditing an algorithm, see generally Sandvig et al., *supra* note 219.

In response, Pauline Kim argues that Kroll's reading of *Ricci* is incorrect, and, in reality, auditing predictive algorithms is both permissible and beneficial for employers.[224] Kim articulates that nothing in *Ricci* prohibits employers from performing audits of their predictive algorithms and implementing necessary changes in those algorithms to correct discriminatory issues.[225] Kim claims that auditing and taking corrective action "is not only legally permitted, but is precisely the type of compliance effort that the law encourages."[226] In this context, corrective action could entail both a technical and nontechnical action, because "the causes of bias often lie not in the code, but in broader social processes."[227] Auditing provides a vehicle to identify problems with automated decision making.[228] According to Kim, the facts in *Ricci* are distinct from a scenario that involves an employer seeking to "change an algorithm prospectively to remove bias."[229] Kim's argument is rooted in the contention that the "strong basis in evidence" requirement only applies if the employer actually participates in intentional discrimination and "seeks to defend its actions as necessary to avoid disparate impact liability."[230] Kim distinguishes a scenario in which an employer would correct a discriminatory algorithm after an audit from *Ricci* because, in *Ricci*, the rejection of the test results negatively affected specific innocent parties.[231] Unlike *Ricci*,

> . . . applicants . . . have not suffered an adverse action because of their race merely because the employer decided to change its hiring algorithm. Applicants would have no legitimate expectations that the company's hiring criteria would never change, and could not credibly claim to have acted in reliance on a particular version of a complex and opaque algorithm.[232]

---

224. *See* Kim, *supra* note 213, at 190.
225. *Id.* at 191.
226. *Id.* at 197.
227. *Id.* at 191.
228. *See id.*
229. Kim, *supra* note 213, at 197.
230. *Id.* at 198.
231. *Id.* at 198-99.
232. *Id.* at 199.

Working from this line of thinking, an employer would not disrupt legitimate expectations, and there are no real victims from taking corrective action.[233] Therefore, nothing would prevent an employer from auditing its predictive hiring algorithm and making adjustments accordingly.[234]

The Supreme Court's holding in *Ricci* is a justifiable cause of concern for employers. While ultimately coming to different conclusions regarding the usefulness and permissiveness of auditing, the contention of both Kroll et al. and Kim offer valuable takeaways for employers seeking to utilize artificial intelligence technology to hire employees in a post *Ricci* world. There is still uncertainty surrounding how the Supreme Court will refine the relationship between disparate treatment and disparate impact and how broadly the ruling from *Ricci* will be applied moving forward.[235] This uncertainty points to the importance of responsible behavior both before and after implementing a predictive algorithm to hire new employees.[236] Kroll's emphasis on responsible algorithm design and implementation should be taken seriously because responsible design is certainly an important component of reducing discriminatory outcomes.[237] Kim is also correct in pointing out that, ultimately, a completely technological solution is unworkable because "[d]esigning a system to be accountable for a substantive goal like nondiscrimination is difficult because it requires specifying the policy goals in terms precise enough to be reduced to code."[238] However, this should not discourage programmers from continuing to strive to create algorithms that do not produce discriminatory outcomes. Conscientious behavior is both beneficial and completely necessary from both programmers and employers seeking to implement artificial intelligence to streamline hiring processes. Failure on either end increases the likelihood of outcomes that would bring rise to a disparate impact claim.

---

233. *Id.*
234. Kim, *supra* note 213, at 199.
235. *Id.* at 202.
236. *Id.*
237. *See* Yao, *supra* note 108.
238. Kroll et al., *supra* note 213, at 192.

Kim's argument in favor of auditing algorithms is certainly well-reasoned and supported by recent case law at lower courts.[239] However, it is worth pointing out that artificial intelligence technology that is used in hiring has taken a number of different forms. For example, as discussed earlier, some technology assesses performance on games, others review resumes and others analyze video interviews.[240] It is conceivable that each of these generalized types of artificial intelligence technology could produce different analyses under *Ricci*. Given the fast pace at which artificial intelligence technology evolves,[241] it is simply not possible to make a blanket assertion that adjustments based on audits of AI systems would not produce a disparate treatment violation similar to what was seen in *Ricci.* However, it is reasonable to conclude that adjustments to target variables, like Kim specifically articulates,[242] would survive the standard set by *Ricci*.

IV. PROPOSED SOLUTION: A BALANCED APPROACH

The rapid advance of artificial intelligence technology and its proliferation into modern society makes it clear that this technology is largely here to stay.[243] Part of what makes artificial intelligence so attractive is its potential benefits, both in its ability to increase efficiency and tackle daunting social challenges.[244] Given the uncertainty in the law and the rapid

239. *See generally* Maraschiello v. City of Buffalo Police Dep't, 709 F.3d 87, 95 (2d Cir. 2013) (holding that there was no violation under *Ricci* because "[u]nlike in Ricci, where the results of a specific test were simply discarded based on the racial statistics reflected in the results, here the City replaced the 2006 list with the 2008 list after spending more than a year preparing to revise its assessment methods. Its problem was with the test itself, rather than with a particular set of results" (footnote omitted)); Carroll v. City of Mount Vernon, 707 F. Supp. 2d 449 (S.D.N.Y. 2010).

240. *See supra* Part II.C.

241. *The Evolution of Artificial Intelligence: AI's Coming of Age*, UBS, https://www.ubs.com/microsites/artificial-intelligence/en/ai-coming-age.html [https://perma.cc/WYB4-PGWY].

242. Kim, *supra* note 213, at 194.

243. *See generally* Rodney Brooks, *The Seven Deadly Sins of AI Predictions*, MIT TECH REV. (Oct. 6, 2017), https://www.technologyreview.com/s/609048/the-seven-deadly-sins-of-ai-predictions/ [https://perma.cc/KG9A-SYAX].

244. *See supra* Part II.

advance of artificial intelligence technology, employers looking to avoid disparate impact liability should utilize a balanced approach that combines the use of algorithms with human decision making. On a long-term basis, employers can advocate for measures that will ultimately work to reduce the overarching biases that are most problematic.

A. A Balanced Approach

Implementing a balance between predictive analytics algorithms and human insight is perhaps the most promising short-term solution for employers who desire to implement artificial intelligence technology into their hiring process while safeguarding their disparate impact liability. This section will discuss practical examples of successes associated with a balanced approach between human decisions and artificial intelligence and how these successes may be recreated in a hiring context.

1. Practical Successes Associated with Balance

Although the challenges associated with implementing predictive algorithms have been well-documented, the development and implementation of an algorithm designed to identify child abuse allegations that warrant investigation provides a framework to which employers should look to for guidance.[245]

Nationally, forty-two percent of child abuse allegations were screened out in 2015 "often based on sound legal reasoning but also because of judgement calls, opinions, biases and beliefs."[246] Despite this, thousands of children died in 2015 as a result of child abuse.[247] After a series of heartbreaking tragedies,[248] Allegheny County, Pennsylvania (which includes Pittsburgh) turned to two scientists, Emily Putnam-Hornstein

245. Hurley, *supra* note 7.

246. *Id.*

247. *Id.*

248. Children had died as a result of their families being "screened out," the worst of which involved two children dying in a fire while their mother was out working as an exotic dancer. *Id.* The Department of Children, Youth and Families had received multiple calls about the family, but screened them out. *Id.*

and Rhema Vaithianathan, to help develop a system of predictive analytics that could improve how Allegheny County handled the call-screening process.[249]

When a call is placed to the Pittsburgh hotline for child abuse and neglect, a screener searches the Department of Children, Youth and Families ("C.Y.F.") database for other allegations that might have been made against the family.[250] For example, a preschool teacher called the hotline and relayed to the screener that a three-year-old child had described that a friend of her mother's "hurt their head and was bleeding and shaking on the floor and the bathtub."[251] The teacher later saw on the news that the mother's boyfriend had died of a drug overdose.[252] The screener saw previous allegations that were ultimately unsubstantiated, and while the current claim was startling, it fell short of the minimum legal requirement to send an investigator, and the screener indicated that the child faced no safety threat.[253]

However, before moving on, the screener's last step is to click on the Allegheny Family Screening Tool.[254] In seconds, the screen displays a "vertical color bar, running from a green 1 (lowest risk) at the bottom to a red 20 (highest risk) on top."[255] For the three-year-old's family, "the score came back as 19 out of a possible 20."[256] A review of the child's mother revealed the mother was in treatment for drug addiction, she had a history of arrests, and the three fathers of the three-year-old child and her siblings had a history of drug addiction and violence.[257] The next morning, a caseworker was sent to investigate the family of the three-year-old to see "what a score of 19 looks like in flesh and blood."[258] All of this information was available to the

---

249. *Id.*
250. *See* Hurley, *supra* note 7.
251. *Id.*
252. *Id.*
253. *Id.*
254. *Id.*
255. Hurley, *supra* note 7.
256. *Id.*
257. *Id.*
258. *Id.*

screener, but navigating the "county's maze of databases" would have taken him hours he simply did not have.[259]

Screeners in Allegheny County are faced with massive amounts of data, and it is difficult to navigate multiple children, parents, and other adults that might be present in the home.[260] The reality is that each of these potential abusers might be in the system, and the person screening the call can take time to investigate, but "the human brain is not that deft at harnessing and making sense of all that data."[261] Unfortunately, further complicating things, when dealing with a problem like deciding where to devote investigative resources, the problem "is not one of finding a needle in a haystack but of finding the right needle in a pile of needles."[262]

The algorithm used by C.Y.S in Allegheny County is rather unique because it is owned by the county, and, as a result, its workings are public information.[263] Before the algorithm was implemented, it was put through a ringer of lawyers, child advocates, former foster children and an independent ethics committee, who "asked hard questions not only of the academics but also of the county administrators who invited them."[264] While other predictive algorithms used by police departments and cities have faced sharp criticism, the system put in place by Allegheny County has received cautious praise because of the care that has been taken in its implementation, the transparency in its creation, and because the program only calls for investigation, not removal of a child from a family.[265]

In dealing with potential biases present in their system, directors of C.Y.F. acknowledge that bias might be inherently present in their work, with or without the use of an algorithm.[266] As discussed at length in this paper, predictive algorithms are associated with entrenched bias against African-Americans and

259. *Id.*
260. Hurley, *supra* note 7.
261. *Id.*
262. *Id.*
263. *Id.*
264. *Id.*
265. Hurley, *supra* note 7.
266. *See id.*

other minority groups.[267] Despite this, "the Allegheny experience suggests that its screening tool is *less* bad at weighing biases than human screeners have been, at least when it comes to predicting which children are most at risk of serious harm."[268]

Walter Smith, a deputy director of C.Y.F., acknowledged that

> We know there are racially biased decisions made . . . . There are all kinds of biases. If I'm a screener and I grew up in an alcoholic family, I might weigh a parent using alcohol more heavily. If I had a parent who was violent, I might care more about that. What predictive analytics provides is an opportunity to more uniformly and evenly look at all those variables.[269]

It should be apparent that identifying child abuse and making hiring decisions are different worlds. The stakes in identifying which families to investigate are much more substantial than a hiring decision. Allegheny County is a government organization, and a large number of employers who might use this technology to make hiring decisions are private entities. Indeed, the process of reporting child abuse is far different from the process of applying for a job.

What seems to set the Allegheny County algorithm apart from situations like what happened at St. George's Hospital is the scrutiny that it has faced, both before its implementation and after.[270] While the St. George's program was certainly well intended—the administration sought to streamline the admissions process—it does not appear that there was any party asking the difficult questions that might have exposed the serious flaws that now make it a cautionary tale.[271] Allegheny County's ability to navigate the myriad of challenges associated with artificial intelligence to create a successful program that

---

267. *See supra* Part II.

268. Hurley, *supra* note 7.

269. *Id.*

270. *See id.*

271. *See* O'NEIL, *supra* note 154, at 117 (writing that "a bit of creative thinking at St. George's could have addressed the challenges facing women and foreigners. The British Medical Journal Report said as much").

creates positive societal outcomes is one of which private employers seeking to minimize disparate impact liability should take notice.[272] While St. George's used its artificial intelligence technology to compare candidates with one another, Allegheny County used the strengths of artificial intelligence to view each individual uniquely and holistically.[273] In a scenario where artificial intelligence is still deeply flawed, Allegheny County has created a system which seems to capitalize on the incredible strength of artificial intelligence to process mass quantities of data, and balancing it with human ability to recognize more intangible realities of what that data might mean in a practical setting.

2. Partnership with People

The success of the C.Y.F. program points to the ways in which algorithms, when used in partnership with human decision making, can be used to create positive outcomes.[274] The American Civil Liberties Union in Pennsylvania offered praise for the C.Y.F. program, suggesting that the program's strength was that it did not decide to actually remove children, and instead was used to help screeners decide where to direct resources.[275] In an employment context, this points to the importance of balancing the incredible information processing capabilities of algorithms with human eyes.

The failures and shortcomings of algorithms in making balanced decisions have been well documented in this article.[276] However, these shortcomings are ultimately the product of human decisions and biases.[277] Some have argued that flawed

272. *See* Hurley, *supra* note 7.

273. *See* O'NEIL, *supra* note 154, at 117-18 (stating that "[t]he key is to analyze the skills each candidate brings to the school, not to judge him or her by comparison with people who seem similar. . . . [W]e've seen time and again that mathematical models can sift through data to locate people who are likely to face great challenges, whether from crime, poverty, or education. It's up to society whether to use that intelligence to reject and punish them—or to reach out to them with the resources they need. We can use the scale and efficiency that make WMDs so pernicious in order to help people. It all depends on the objective we choose").

274. *See* Hurley, *supra* note 7.

275. *Id.*

276. *See generally supra* Part II.

277. *See* Hurley, *supra* note 7.

algorithms are "anecdotal reflections of society's deep-rooted biases and a lingering digital divide."[278] As a result, simply changing the algorithm is only a temporary fix that does not deal with long-term social consequences.[279]

Blindly leaning on a nominal diversity task force to keep an algorithm accountable is unlikely to be a viable source of balance.[280] Research has shown that diversity initiatives are not always effective in promoting a diverse workplace, and can ultimately be harmful to creating an inclusive, diverse workplace.[281] Studies have shown that poorly implemented diversity programs and messaging in work environments can signal to white male candidates and employees that "they might be undervalued and discriminated against. These concerns interfered with their interview performance and caused their bodies to respond as if they were under threat."[282] Interestingly, this outcome seemed to occur regardless of the male's "political ideology, attitudes toward minority groups, . . . or beliefs about the fairness of the world."[283] This ultimately points to how deeply a negative response to poorly positioned diversity messaging is rooted.[284] Therefore, for a partnership between algorithms and humans to produce a truly effective balance that reduces the likelihood of disparate impact liability, company initiatives need to be intentionally implemented and robust.[285]

On the other side of the token, the algorithm itself must be responsibly produced and subject to an appropriate level of accountability. The C.Y.S. algorithm was put through a ringer of lawyers, parents, advocates, former foster children and an

278. Omer Tene & Jules Polonetsky, *Taming the Golem: Challenges of Ethical Algorithmic Decision-Making*, 19 N.C. J.L. & TECH. 125, 135 (2017).

279. *Id.*

280. *See* Tessa L. Dover et al., *Diversity Policies Rarely Make Companies Fairer, and They Feel Threatening to White Men*, HARV. BUS. REV. (Jan. 4, 2016), https://hbr.org/2016/01/diversity-policies-dont-help-women-or-minorities-and-they-make-white-men-feel-threatened [https://perma.cc/U7N4-WASS].

281. *Id.*

282. *Id.*

283. *Id.*

284. *Id.*

285. *See generally* Evan M. Roberts, *Creating Stronger Diversity Initiatives in Employment Settings*, CORNELL HR REV. (Nov. 4, 2011), https:// digitalcommons.ilr.cornell.edu/ cgi/ viewcontent.cgi?referer=http:/ /scholar. google.com/ &httpsredir=1&article=1026&context=chrr [https://perma.cc/X5ZK-H4R2].

independent ethics committee, where it was subject to a range of difficult questions about its potential impact.[286] The transparency of C.Y.S and its willingness to listen to difficult criticisms allowed for the production of a system that is more effective and promising than its predecessors.[287] The result is an algorithm that, while imperfect, is able to "more uniformly and evenly look at all . . . variables."[288]

The government ownership of the C.Y.S. algorithm allowed for levels of transparency that largely do not exist with private algorithms.[289] However, some developers of predictive hiring algorithms argue that if developed in a responsible, comprehensive way, they have the capability to "increase diversity, advance the interests of minorities, and fight discrimination."[290] Ultimately, while acknowledging the negative impact of poorly designed and defined predictive algorithms, the argument is that technology at its most basic level, is neutral.[291] It is not the technology itself that perpetuates bias; the positive or negative impact of the technology depends on the design and implementation.[292] In fact, Frida Polli[293] goes as far as asserting that it is possible to "make sure that your algorithms are not biased even if your training set is. [W]e personally believe that no algorithm should be released unless it has been tested to be bias-free (which we do!)."[294] Whether this is actually true remains to be seen. However, it points to the importance of selecting an algorithm that is created with an eye towards reducing bias and discriminatory effects. In selecting a hiring assistance service that uses artificial intelligence

---

286. Hurley, *supra* note 7.

287. *See id.*

288. *Id.*

289. There are a number of intellectual property issues surrounding private algorithms, however they are outside the scope of this article.

290. Frida Polli, *Algorithms: Friend or Foe of Diversity?*, LINKEDIN (April 20, 2015), https:// www.linkedin.com/ pulse/ algorithms- friend- foe- diversity- frida- polli/ [http://perma.cc/ZRU7-PCRR].

291. *Id.*

292. *Id.*

293. Frida Polli is one of the founders of Pymetrics. *See supra* text accompanying note 81.

294. Frida Polli, Comment to Polli, *supra* note 290, https:// www.linkedin.com/ pulse/ algorithms- friend-foe-diversity-frida-polli/ [http://perma.cc/ZRU7-PCRR].

algorithms, employers should ask difficult questions about the creation of the algorithm in order to truly protect themselves from disparate impact liability.

B. Long Term Reforms

Long term, employers should push for reforms both on a regulatory level and within the technology industry as a whole. While these changes will likely take many years, employers should utilize their influence to push for overarching reforms that will improve the overall quality of predictive hiring algorithms and ultimately reduce the likelihood that their use would give rise to a disparate impact claim.

1. Government Regulation

Some have suggested the creation of a regulatory body as a comprehensive solution to the overarching issues that have emerged from predictive algorithms.[295] This argument makes an analogy between the creation of the Food and Drug Administration ("F.D.A.") in the midst of a public health crisis, and the ways in which the F.D.A. proactively deals with drug makers as a model for regulating algorithms before they are put into the market.[296] However, the reality is that actually creating an agency that oversees a complex and quickly developing industry should "merit careful scrutiny" because of legitimate concerns that a heavy-handed agency[297] could stifle the very innovations within the technology industry that could make these algorithms less harmful.[298] However, the fast developing pace of technology warrants careful consideration of the positive influence of a uniform system of accountability for algorithms.

On a more localized level, New York City Council recently passed a bill that established a task force to examine the city's

295. *See* Andrew Tutt, *An FDA for Algorithms*, 69 ADMIN. L. REV. 83, 90 (2017) (advocating for the creation of a federal agency to ensure that algorithms are safe and fair).

296. *Id.* at 120-22.

297. The administrative challenges associated with creating a new federal agency are outside the scope of this article. For a more robust discussion of what this might look like, see *id.* at 117-18.

298. *Id.* at 122.

use of automated decision systems.[299] Originally, the bill was intended to

> . . . require agencies that use algorithms or other automated processing methods that target services, impose penalties, or police persons to publish the source code used for such processing. It would also require agencies to accept user-submitted data sets that can be processed by the agencies' algorithms and provide the outputs to the user.[300]

When the bill was introduced, sponsor James Vacca stated, "If we're going to be governed by these machines and algorithms and data, well, they better be transparent."[301] However, after backlash at hearings regarding the potential complications around increasing transparency by disclosing algorithm source codes[302] the bill was clawed back to create a task force to "determine how we can evaluate the outputs of automated systems and figure out if and when there is harm done."[303] At hearings, experts testified in favor of qualified transparency, "less than total disclosure of the source code, more than nothing at all," but it was not enough to overcome concerns that publishing proprietary information breaching contracts that the city contracts with.[304] As things currently stand, the Council is unable to access "basic knowledge" that may ultimately limit its effectiveness producing comprehensive findings.[305] Additionally, the bill fails to address "how the city government,

---

299. Julia Powles, *New York City's Bold, Flawed Attempt to Make Algorithms Accountable*, NEW YORKER (Dec. 20, 2017), https:// www.newyorker.com/ tech/ elements/ new- york- citys- bold- flawed- attempt- to-make-algorithms-accountable [http://perma.cc/CQ4B-2SU3].

300. N.Y.C. COUNCIL COMM. ON TECH., REP. ON INT. NO. 1696, at 4 (Oct. 16, 2017), http://legistar.council.nyc.gov/View.ashx?M=F&ID=5505265&GUID=0BDD96E6-A15A-4A36-83BF-2FA3D5A0DCFC [https://perma.cc/DU48-84J9].

301. Powles, *supra* note 299.

302. Don Sunderland, Testimony Before the New York City Council Committee on Technology (Oct. 16, 2017), (transcript available at http:// legistar .council. nyc. gov/View.ashx?M=F&ID=5522569&GUID=DFECA4F2-E157-42AB-B598-BA3A8185E3FF [https://perma.cc/FV25-KHLS]).

303. Sidney Fussell, *New York City Wants to Audit the Powerful Algorithms That Control Our Lives*, GIZMODO (Dec. 14, 2017, 6:00 PM), https://gizmodo.com/new-york-city-wants-to-audit-the-powerful-algorithms-th-1821305284 [https://perma.cc/ZPR4-ZGPW].

304. Powles, *supra* note 299.

305. *Id.*

and those who advise it, can exercise some muscle in their dealings with the companies that create automated-decision systems."[306]

Despite the flaws and potential issues associated with the bill, it is a fascinating "experiment in the world of algorithmic accountability, sent out much like Captain Picard, from 'Star Trek,' would send out a probe to explore a wormhole."[307] It is unclear what the task force will uncover and what it will report. It is clear, however, that this bill did not take the forceful path that Vacca originally intended when it was authored.[308] As it stands now, the task force represents a relatively passive effort. The significance of this task force will likely be determined by the action or inaction that comes from its findings. However, given the size and influence of New York City, its findings, set to be released in 2019, will likely have substantial impact on the world of algorithmic accountability.

2. Diversity in Tech

As discussed earlier, the lack of diversity in the technology industry is a serious problem, and the homogenous nature of the industry allows for homogenous opinions and worldviews to creep into the algorithms that assist in hiring decisions.[309]

In order to effectuate meaningful change in the homogenous makeup of the technology industry, tech companies must move away from simply accepting "cognitive diversity" a reinterpretation of diversity "to encompass what Silicon Valley has never had a shortage of—individual white men, each with their unique thoughts and ideas."[310] Different viewpoints are not in and of themselves bad. However, cognitive diversity becomes dangerous when it is used as an excuse to sidestep racial and gender diversity in the workplace. Ultimately, the effort to increase cognitive diversity cannot "come at the

306. *Id.*
307. *Id.*
308. *Id.*
309. *See* discussion *supra* Part II.D.
310. Bärí A. Williams, Opinion, *Tech's Troubling New Trend: Diversity Is in Your Head*, N.Y. TIMES (Oct. 16, 2017), https:// www.nytimes.com/ 2017/ 10/ 16/ opinion/ diversity- tech- women- silicon- valley.html [https://perma.cc/C2KZ-Y9UC].

expense of hiring members of actual underrepresented communities."[311] In fact, studies suggest that increased racial and gender diversity "is associated with increased sales revenue, more customers, greater market share, and greater relative profits."[312]

Employers should use their influence to push technology companies to increase racial and gender diversity in the workforce. Given how pervasive and deeply-rooted the problems associated with a lack of racial and gender diversity in tech, it is unlikely that this change will come in the short term. However, given how drastically this could reduce the discriminatory effect of algorithms, the meaningful use of time and financial resources would likely be a worthwhile investment in the future of predictive hiring algorithms as a whole.

V. CONCLUSION

It is clear that artificial intelligence and predictive algorithms are not going anywhere in society at large. Their incredible potential to increase efficiency and allow companies to focus more on innovation rather than mundane tasks makes it incredibly likely that they will occupy a substantial space in the workplace, including the hiring process.[313] However, it is equally clear that with the incredible potential artificial intelligence offers, comes a variety of challenges that could substantially increase an employer's disparate impact liability.

It should be clear that artificial intelligence is not a bias-free savior for employers.[314] At the same time, shunning algorithms completely because of their bias does nothing to solve the problems that create disparate impact liability. Employers seeking to take advantage of the benefits of artificial intelligence technology to increase hiring efficiency should be prepared to ask difficult questions and ensure that this technology is implemented in a way that is responsible. A balance between human accountability and a responsibly created

311. *Id.*

312. Cedric Herring, *Does Diversity Pay?: Race, Gender, and the Business Case for Diversity*, 74 AM. SOC. REV. 208, 219 (2009).

313. *See* Tutt, *supra* note 295 at 99-100.

314. *See* discussion *supra* Part II.D.

and chosen artificial intelligence system may be the best way to deal with these core tensions.

Bias is a challenge that is rooted in human nature, and it is passed in code to predictive hiring algorithms. As a result, attention must be given to both the short-term issues with biased algorithms and long-term issues associated with regulation and increasing diversity in the technology industry.

McKenzie Raub