



---

GW Law Faculty Publications & Other Works

Faculty Scholarship

---

2023

## A New Approach to Measuring AI Bias in Human Resources Functions: Model Risk Management

Keith E. Sonderling

*The George Washington University Law School*

Aram Gavoor

*The George Washington University Law School*

Follow this and additional works at: [https://scholarship.law.gwu.edu/faculty\\_publications](https://scholarship.law.gwu.edu/faculty_publications)



Part of the [Law Commons](#)

---

### Recommended Citation

A New Approach to Measuring AI Bias in Human Resources Functions: Model Risk Management, 54 The Seton Hall Law Review (with Keith E. Sonderling) (forthcoming 2024).

This Article is brought to you for free and open access by the Faculty Scholarship at Scholarly Commons. It has been accepted for inclusion in GW Law Faculty Publications & Other Works by an authorized administrator of Scholarly Commons. For more information, please contact [spagel@law.gwu.edu](mailto:spagel@law.gwu.edu).

**A NEW APPROACH TO MEASURING AI BIAS IN  
HUMAN RESOURCES FUNCTIONS: MODEL RISK MANAGEMENT**

© 2023 Keith E. Sonderling\* & Aram A. Gavoor\*\*

**INTRODUCTION**

The Artificial Intelligence (“AI”) revolution is here, and when it comes to AI’s use in employment decision-making there is no exception.<sup>1</sup> As thousands of AI-enabled human resources technologies are on the market, technology companies offering AI services in the human resources (“HR”) space are already valued in the billions of dollars, with this value increasing dramatically each year.<sup>2</sup> Unsurprisingly, businesses are eager to use AI in their HR processes because the cost savings can be enormous. According to a recent survey, AI can reduce the cost of screening resumes by 75%, improve the revenue per employee by 4%, and decrease employee turnover by 35%.<sup>3</sup>

Aside from the financial incentives for businesses, the use of AI in HR can benefit employees and employers alike by advancing equality of opportunity, diversity, inclusion, and accessibility.<sup>4</sup> Indeed, by eliminating some human decision-making and replacing it with carefully designed algorithms, AI holds the potential to substantially reduce the kind of bias that has been unlawful in the United States since the civil rights movement of the mid-twentieth century.<sup>5</sup>

---

\* Commissioner, U.S. Equal Employment Opportunity Commission; Professorial Lecturer in Law, the George Washington University Law School.

\*\* Associate Dean for Academic Affairs and Professorial Lecturer in Law, the George Washington University Law School.

<sup>1</sup> One recent study found that 83% of human resources leaders rely in some form of automated technology in employment decision-making. Dinah Wisenberg Brin, *Employers Embrace Artificial Intelligence for HR*, SHRM (Mar. 22, 2019), <https://www.shrm.org/resourcesandtools/hr-topics/global-hr/pages/employers-embrace-artificial-intelligence-for-hr.aspx>. A recent Mercer survey said that 79% of HR leaders already use or plan to use AI to identify job candidates. <https://www.mercer.com/content/dam/mercerc/attachments/private/global-talent-trends/2021/gl-2021-gtt-global-eng-mercerc.pdf>. See also SHRM, *Automation & AI in HR*, <https://advocacy.shrm.org/SHRM-2022-Automation-AI-Research.pdf> (“Nearly 1 in 4 organizations use automation and/or AI to support HR-related activities.”); SHRM, *Regulations Ahead on AI*, SHRM, <https://www.shrm.org/hr-today/news/all-things-work/pages/regulations-ahead-on-artificial-intelligence.aspx> (“[N]early 92% of executives said they were increasing investments in data and artificial intelligence systems.”).

<sup>2</sup> <https://www.globenewswire.com/en/news-release/2022/10/21/2538959/0/en/With-7-5-CAGR-Human-Resource-Technology-Market-Size-Worth-USD-39-90-Billion-In-2029.html>. Andy Charlwood and Nigel Guenole, *Can HR adapt to the paradoxes of artificial intelligence?* Human Resource Management Journal, 07 January 2022 <https://onlinelibrary.wiley.com/doi/full/10.1111/1748-8583.12433> (“Recently, eightfold, an AI based talent intelligence platform that helps attract, develop and retain top talent, has raised \$220 million and is now valued at over \$2 billion. Phenom, an HR technology firm that automates job tasks and improves the job search experience, was reported by Forbes as having ‘quietly become a billion dollar unicorn.’”) (citations omitted).

<sup>3</sup> Anna Verasai, *AI Evolves the Hiring Process*, THE HR DIGEST (Aug. 24, 2021), <https://www.thehrdigest.com/ai-evolves-the-hiring-process/>.

<sup>4</sup> Haiyan Zhang, et al., *The Role of AI in Mitigating Bias to Enhance Diversity and Inclusion*, IBM SMARTER WORKFORCE INSTITUTE (Mar. 2019), <https://www.ibm.com/downloads/cas/2DZELQ4O>.

<sup>5</sup> Frida Polli, *Using AI to Eliminate Bias from Hiring*, HARVARD BUSINESS REVIEW (Oct. 29, 2019), <https://hbr.org/2019/10/using-ai-to-eliminate-bias-from-hiring>.

At the same time, if poorly designed and carelessly implemented, AI can discriminate on a larger scale than that of any individual. In milliseconds, AI can eliminate qualified candidates from consideration for jobs based on characteristics that federal law generally prohibits employers from factoring into consideration, such as sex, race, national origin, age, or disability.<sup>6</sup> This potential misuse of AI can result in the loss of exemplary candidates and employees, stunted growth, lowered company morale, reputational damage, and large fines and penalties. It also can deprive individuals of a fundamental civil right: an equal opportunity to enter and thrive in the workplace.

Although AI is helping companies make decisions more efficiently, effectively, and economically, many simply do not know how the products work, or, even worse, whether they are violating laws in the process of using this technology. Sixty-five percent of surveyed C-suite executives who have implemented AI in their corporations felt their companies would not be able to explain how their AI models worked. Only 22% have an AI ethics board to monitor their algorithms.<sup>7</sup>

As a result, companies using or seeking to use AI in HR technology face a difficult question: How can they benefit from this technology without violating federal or state laws, including the federal laws prohibiting discrimination in employment? Unfortunately, companies are currently confronted with a general lack of guidance, disjointed federal approach, and patchwork of legislative proposals as well as confusing state, local, federal, and international regulatory regimes.<sup>8</sup>

Because legislative change and federal regulatory solutions remain elusive, companies using AI for their HR functions have little choice but to pursue voluntary compliance as a function of risk mitigation or principle. This practice has always played a critical role in the regulation of employment in the United States. The financial services industry has utilized a form of voluntary regulation, the model risk management (“MRM”) framework, for over a decade to help improve compliance under regulatory authority so that it can meet the challenges of self-regulating this ever-changing technology.

This Essay makes the case that MRM concepts, which were issued in response to the global financial crisis in 2007-2008 and have since governed US financial institutions, are applicable to the unique technology challenges that AI-powered HR tools present. Thus, these concepts can and should be appropriately calibrated to manage discrimination risk as well as other business risks when applied to the use of these tools. As a result, employers can have confidence using AI tools to gain the many benefits this technology offers, while significantly lessening their risk of violating federal, state, and local employment discrimination laws.

---

<sup>6</sup> See, e.g., Jeffrey Dastin, *Amazon scraps secret AI recruiting tool that showed bias against women*, REUTERS (Oct 10, 2018), <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>.

<sup>7</sup> [https://www.fico.com/en/resource-access/download/36776?access\\_token\\_9f4fd=23e4b9be5ae628820105166a999405d0140066fb956f1ab61459deb8f3aac290](https://www.fico.com/en/resource-access/download/36776?access_token_9f4fd=23e4b9be5ae628820105166a999405d0140066fb956f1ab61459deb8f3aac290).

<sup>8</sup> See Roy Maurer, *Use of AI in the Workplace Raises Legal Concerns*, SHRM (Apr. 7, 2021), <https://www.shrm.org/resourcesandtools/hr-topics/technology/pages/use-of-ai-in-the-workplace-raises-legal-concerns.aspx>.

The voluntary compliance measures that are embodied by MRM as laid out below fit neatly into federal administrative law and Title VII of the Civil Rights Act of 1964 (“Title VII”). In the absence of new legislative policymaking and timely and practical AI guidance from federal agencies, the MRM approach to AI-powered HR tools squarely offers an intentional, traceable mechanism of business risk mitigation for companies and the HR industry. The approach is intuitive, modular, and subject to best-practices evolution that could find its way with the support of the regulatory agencies once its adoption is settled and its outcomes are proven to reduce unlawful uses of AI. Short of earning EEOC approval, sound MRM implementation could serve as a reliable shield against Commission substantive enforcement at multiple stages of Title VII enforcement.

Part I of this Essay describes AI and discusses the ways businesses are using AI-powered HR tools for employment decision-making. Part II explains how an employer’s use of AI tools is subject to federal employment discrimination law. Part III highlights the importance of employer voluntary compliance initiatives, explains the key concepts of MRM as implemented by the financial industry, and applies those concepts to an employer’s use of AI tools. Part IV lays out the administrative law legitimacy for this commonsense approach to voluntary compliance.

## I. ARTIFICIAL INTELLIGENCE AND HR TECHNOLOGY

Although it is challenging to define, AI generally examines data, identifies patterns, and makes probabilistic inferences from these patterns.<sup>9</sup> AI can be thought of as “a large class of software-based systems that receive signals from the environment and take actions that affect that environment by generating outputs such as content, predictions, recommendations, classifications, or decisions influencing the environments they interact with, among other outputs.”<sup>10</sup> Oftentimes, commentators reference AI through similar terms such as machine learning, algorithmic decision-making, and automated decision-making.<sup>11</sup>

The global workforce is already using AI in HR. Over 80% of HR leaders rely on some form of automated technology in making their hiring decisions.<sup>12</sup> HR professionals will increasingly use this technology for social media tools like LinkedIn,<sup>13</sup> job-aggregator sites like Indeed or Zip

---

<sup>9</sup> Pauline T. Kim & Matthew T. Bodie, *Artificial Intelligence and the Challenges of Workplace Discrimination and Privacy*, 35 ABA J. LAB. & EMP. L. 289, 290 (2021).

<sup>10</sup> NIST SP-1270 at 3.

<sup>11</sup> Pauline T. Kim & Matthew T. Bodie, *Artificial Intelligence and the Challenges of Workplace Discrimination and Privacy*, 35 ABA J. Lab. & Emp. L. 289, 290 (2021).

<sup>12</sup> Lisbeth Perez, *EEOC Commissioner: Companies Must Mitigate the Use of AI for Employment Decisions*, MERITALK, (Oct. 19, 2021 9:00 AM), <https://www.meritalk.com/articles/eec-commissioner-companies-must-mitigate-the-use-of-ai-for-employment-decisions/>.

<sup>13</sup> By 2015, 92% of recruiters reported using social media as part of their candidate job search. Sharone, O. (2017), *LinkedIn or LinkedOut? How Social Networking Sites are Reshaping the Labor Market*, Emerging Conceptions of Work, Management and the Labor Market (Research in the Sociology of Work, Vol. 30), Emerald Publishing Limited, Bingley, 2, <https://doi.org/10.1108/S0277-283320170000030001>. Use of some of the search and matching features of social media services like LinkedIn necessarily implicates AI and machine learning. See, e.g., Qi Guo, *The AI Behind LinkedIn Recruiter search recommendation systems*, LinkedIn Engineering, Apr. 22, 2019 <https://engineering.linkedin.com/blog/2019/04/ai-behind-linkedin-recruiter-search-and-recommendation-systems>

Recruiter<sup>14</sup> or more-sophisticated AI products that use machine learning to match candidates based upon actual, adjacent, or perceived skills.<sup>15</sup>

Companies are already using AI-powered tools to write job descriptions, screen resumes, chat with applicants, conduct job interviews, and even predict whether a candidate will accept an offer. Other AI-powered tools seek to predict how a candidate might interact with their new co-workers, and try to identify current employees' existing and potential skills, track productivity, assess workers, and pick who might receive valuable and career-changing upskilling and reskilling opportunities.<sup>16</sup> Employers are also using AI tools to help decide whether to terminate an employee—or even allow the AI to decide for them.<sup>17</sup> In short, AI is now making nearly all types of decisions once made by HR personnel.

Most of these AI-driven HR tools are sold with the laudable goals of increasing diversity and equality of opportunity by making objectively fair decisions untarnished by intentionally or implicitly biased human decision-making.<sup>18</sup> However, many AI tools might “not operate as intended because they are making inferences from patterns observed in data rather than a true understanding of what causes those patterns.”<sup>19</sup> AI is also subject to the biases of its human creators, who may “unwittingly ‘bake’ existing biases into systems by training them on biased data or with ‘rules’ created by experts with implicit biases.”<sup>20</sup> Unlawful biases in the existing (or “training”) data can lead to biased algorithms because, if left unremedied, AI tools simply replicate

---

<sup>14</sup> In a 2021 survey, 22% of talent acquisition and HR leaders were beginning to explore programmatic job advertising or automation. A further 32% used a combination of automation and traditional advertising. Roy Maurer, *Programmatic Technology Upgrades Job Advertising*, SHRM (Mar. 21, 2021), <https://www.shrm.org/resourcesandtools/hr-topics/talent-acquisition/pages/programmatic-technology-upgrades-job-advertising.aspx>.

<sup>15</sup> Jona Babi, *LinkedIn – AI helps companies and professionals find each other faster*, HARVARD BUSINESS SCHOOL DIGITAL INNOVATION AND TRANSFORMATION: MBA STUDENT PERSPECTIVES, Apr. 18, 2020, <https://digital.hbs.edu/platform-digit/submission/linkedin-ai-helps-companies-and-professionals-find-each-other-faster/> (explaining that LinkedIn “uses data analytics and AI extensively to improve their members’ and customers’ experiences.”).

<sup>16</sup> See, e.g., Nick Otto, *HR Tech QuickHits: The Power of AI in Employee Skills Identification*, HR Executive (Sept. 14, 2022), <https://hrexecutive.com/hr-tech-quick-hits-the-power-of-ai-in-employee-skills-identification/>.

<sup>17</sup> Jacob Zinkula, *AI is Helping Your Company Decide Who to Lay Off*, Business Insider (Feb. 23, 2023), <https://www.businessinsider.com/layoffs-today-trends-ai-data-companies-jobs-algorithms-chatgpt-2023-2>.

<sup>18</sup> See, e.g., Eightfold, *Strong DE&I Programs are Critical to Attract and Retain Your Talent*, <https://eightfold.ai/why-eightfold/diversity-and-inclusion/>.

<sup>19</sup> NIST AI Risk Management Framework: (AI RMF 1.0) January 26, 2023 at 1.

<sup>20</sup> See, e.g., Karen Hao, *AI is sending people to jail – and getting it wrong*, MIT Technology Review, January 21, 2019, <https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/> (“[P]opulations that have historically been disproportionately targeted by law enforcement—especially low-income and minority communities—are at risk of being slapped with high recidivism scores. As a result, the algorithm could amplify and perpetuate embedded biases and generate even more bias-tainted data to feed a vicious cycle.”).

those embedded unlawful biases.<sup>21</sup> The result can be a self-fulfilling prophecy where AI continues to perpetuate and institutionalize the biases of its creators.<sup>22</sup>

A widely reported alleged example of AI bias is an algorithmic system Amazon used to evaluate job candidates that “taught itself that male candidates were preferable” to women.<sup>23</sup> Other reports abound, including an example of a resume-screening tool, marketed to diversify a company’s workforce, that ultimately considered the name “Jared” and playing high school lacrosse as most-indicative of job performance.<sup>24</sup> Examples such as these should capture the attention of employers using AI for employment decision-making, as they have to regulators from cities, to states, to governments worldwide.

Faced with this intersecting set of benefits and risks, employers should recognize the extent to which algorithmic bias in employment decision-making may violate federal, state, and local laws.

## II. EXISTING FEDERAL EQUAL EMPLOYMENT OPPORTUNITY FRAMEWORK WITHIN WHICH AI-ASSISTED HR PRACTICES ARE BOUND.

An employer’s use of AI in its HR processes is subject to the laws enforced by the U.S. Equal Employment Opportunity Commission (“EEOC”). These laws prohibit discrimination in employment on the basis of protected characteristics such as race, color, religion, sex, national origin, age, disability, and genetic information.<sup>25</sup> Although almost all of these statutes predate even the concept of AI, they unquestionably apply to an employer’s use of AI to make or assist in employment decisions.<sup>26</sup>

---

<sup>21</sup> See, e.g., Matthew Gault, *AI Trained on 4Chan Becomes ‘Hate Speech Machine’*, VICE, June 7, 2022, <https://www.vice.com/en/article/7k8zwx/ai-trained-on-4chan-becomes-hate-speech-machine>. See also Brian Uzzi, *A Simple Tactic That Could Help Reduce Bias in AI*, Harvard Business Review (Nov. 4, 2020), <https://hbr.org/2020/11/a-simple-tactic-that-could-help-reduce-bias-in-ai>; Aram A. Gavoor & Raffi Teperdjian, *A Structural Solution to Mitigating Artificial Intelligence Bias in Administrative Agencies*, 89 Geo. Wash. L. Rev. Arguendo, 71, 78 (2021) (“The prevailing technique to Narrow AI is known as ‘machine learning,’ defined as ‘an automated process of discovering correlations . . . between variables in a dataset, often to make predictions or estimates of some outcome.’ During this machine learning algorithm development process, which is commonly referred to as its ‘training process,’ computer systems use large amounts of data to draw these new correlations. Many of the bias issues in machine learning arise from the methodology employed to train the dataset.”); *c.f.* Talia B. Gillis, *The Input Fallacy*, 106 MINN. L. REV. 1175 (2022).

<sup>22</sup> Brian Uzzi, *A Simple Tactic That Could Help Reduce Bias in AI*, Harvard Business Review (Nov. 4, 2020), <https://hbr.org/2020/11/a-simple-tactic-that-could-help-reduce-bias-in-ai>; see also Miranda Bogen & Aaron Rieke, *Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias*, UPTURN, 26–39 (Dec. 10, 2018), <https://www.upturn.org/reports/2018/hiring-algorithms/> (“We conclude that without active measures to mitigate them, bias will arise in predictive hiring tools by default.”).

<sup>23</sup> Jeffrey Dastin, *Amazon scraps secret AI recruiting tool that showed bias against women*, REUTERS (October 10, 2018), <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>.

<sup>24</sup> See Miranda Bogen, *All the ways hiring algorithms can introduce bias*, Harvard Business Review (May 6, 2019), <https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias> citing Dave Gershgorin, *Companies are on the hook if their hiring algorithms are biased*, QUARTZ (October 22, 2018), <https://qz.com/1427621/companies-are-on-the-hook-if-their-hiring-algorithms-are-biased/>

<sup>25</sup> <https://www.eeoc.gov/discrimination-type>.

<sup>26</sup> See, e.g. Keith Sonderling, *Artificial Intelligence is Changing How HR is Handled at Companies. But Do Robots Care About Your Civil Rights?*, CHICAGO TRIBUNE (Sept. 20, 2021),

For example, under Title VII, there are two established theories of discrimination: disparate treatment and disparate impact. Under disparate treatment, an employer deliberately treats an individual less favorably than others because of the individual's membership in a protected class—for example, the person's race, sex, or religion. If an employer uses an AI tool to eliminate certain individuals from their resume database, for instance older workers based upon college graduation dates, the employer would be using technology to engage in intentional discrimination.

Alternatively, disparate impact discrimination exists when an employment practice that is neutral on its face has an adverse impact on members of a protected class.<sup>27</sup> Intent to discriminate is not necessary for disparate impact liability. Rather, in a disparate impact case, the focus is on the consequences of the employer's actions, not on their intent. So, if an employer carelessly implements the recommendations of an AI-powered resume-screening tool that considers male candidates to be preferable to female candidates, the employer would likely be liable under a disparate impact theory, even though the employer did not intend to discriminate against female candidates.<sup>28</sup>

Thus, employers seeking to comply with their obligations under Title VII must take care that the algorithmic tools they are using do not result in impermissible adverse impact against a protected group. However, the EEOC last issued formal guidance on testing related to such an adverse impact nearly 45 years ago.<sup>29</sup> Thus, the current guidance on discriminatory impact testing predates not only the development and use of AI, but also the enactment of other anti-discrimination laws that apply to an employer's use of AI in HR technology, such as the Americans with Disabilities Act.<sup>30</sup>

While the longstanding employment antidiscrimination laws apply with full force to the use of AI as they do to an employment decision made by a manager, employers continue to look for much-needed guidance on their modern-day application. Instead of focusing on the current laws, companies are forced to redirect their efforts to comprehend numerous proposals and new laws purporting to add tech-specific limitations and burdens to AI's use in particular contexts.<sup>31</sup> These new laws and proposals are creating additional challenges and confusion.

While some of these domestic and international efforts may be appreciable, the pursuit of meaningful legal compliance and ethical use of AI tools remains problematic. The absence of clear, consistent federal guidance along with the overlapping and evolving patchwork of state and local

---

<https://www.chicagotribune.com/opinion/commentary/ct-opinion-robots-ai-civil-rights-amazon-20210920-tef7m7az3rgjtacauazvw3u224-story.html>.

<sup>27</sup> *Griggs v. Duke Power Co.*, 401 U.S. 424, 425–36 (1971).

<sup>28</sup> *Id.*; 42 U.S.C. §§ 2000e - 2000e17 (as amended)

<sup>29</sup> See 29 C.F.R. § 1607. In 1978, the EEOC joined other agencies in issuing the Uniform Guidelines on Employee Selection Procedures (“UGESP” or the “Uniform Guidelines”).

<sup>30</sup> See, e.g., Americans with Disabilities Act, 42 U.S.C. 12111 *et seq.*

<sup>31</sup> Keith E. Sonderling, Bradford J. Kelley, Lance Casimir, *The Promise and the Peril: Artificial Intelligence and Employment Discrimination*, 77 U. MIAMI L. REV. 1, 37–50 (2022).

laws is likely to persist. Nevertheless, the private sector should do what it does best – take initiative and pursue voluntary compliance.<sup>32</sup>

### III. VOLUNTARY COMPLIANCE AND MODEL RISK MANAGEMENT

Unbeknownst to many, the regulation of employment in the United States is historically rooted in voluntary compliance.<sup>33</sup> Title VII codified this emphasis when it stated that the EEOC “shall endeavor to eliminate any such alleged unlawful employment practice by informal methods of conference, conciliation, and persuasion.”<sup>34</sup> The U.S. Supreme Court has also observed that voluntary compliance is Congress’s intended means of “achieving the objectives of Title VII”<sup>35</sup> and “essential to the statutory scheme.”<sup>36</sup> While some of the EEOC’s voluntary compliance efforts focus on resolving charges of unlawful discrimination after they are filed,<sup>37</sup> such efforts by employers appropriately focus on the prevention of unlawful discrimination.

Consistent with this approach, companies using or seeking to use AI for employment decision-making should also strive for internal compliance. Companies have the benefit of looking at key concepts from the MRM framework that corporations in the financial sector have implemented for over a decade. MRM encompasses a framework of principles and best practices that institutions apply to improve compliance with regulatory guidelines.<sup>38</sup> By implementing a robust MRM framework as a form of voluntary compliance, employers can have confidence that they will be able to use AI tools to fulfill and support HR functions. These key MRM concepts can help employers use AI throughout all stages of the employee lifecycle while mitigating the risk of unlawful employment discrimination and advancing diversity, equality of opportunity, and inclusion in the workplace.

#### A. Background on Financial Industry Regulation and Model Risk Management

In the United States, large financial institutions are generally governed by the Federal Reserve Board (the “Federal Reserve”), the Office of the Comptroller of the Currency (“OCC”), the Federal Depository Insurance Commission (“FDIC”), and other agencies. The Federal Reserve aims its

---

<sup>32</sup> Keith E. Sonderling and Bradford J. Kelley, FILLING THE VOID: ARTIFICIAL INTELLIGENCE AND PRIVATE INITIATIVES

<sup>33</sup> Kelly Trindel et al., *Fairness in Algorithmic Employment Selection: How to Comply with Title VII*, 35 A.B.A. Journal of Labor Employment Law 2 (2021), 241, 244,  
[https://www.americanbar.org/content/dam/aba/publications/aba\\_journal\\_of\\_labor\\_employment\\_law/v35/no-2/fairness-algorithmic-employment-selection.pdf](https://www.americanbar.org/content/dam/aba/publications/aba_journal_of_labor_employment_law/v35/no-2/fairness-algorithmic-employment-selection.pdf).

<sup>34</sup> *Id.* at 244–45.

<sup>35</sup> *Ricci v. DeStefano*, 557 U.S. 557, 581 (2009), quoting *Int’l Ass’n of Firefighters v. City of Cleveland Local No. 93*, 478 U.S. 501, 515 (1986).

<sup>36</sup> *Id.* at 573.

<sup>37</sup> The EEOC has administered pilot programs “dedicated to increasing ‘voluntary resolutions’ through mediation.” *EEOC Announces Pilot Programs to Increase Voluntary Resolutions*, (July 7, 2020),

<https://www.eeoc.gov/newsroom/eeoc-announces-pilot-programs-increase-voluntary-resolutions>

(“The U.S. Equal Employment Opportunity Commission (EEOC), today announced two six-month pilot programs that will expand opportunities for parties to voluntarily resolve charges through mediation and increase the effectiveness of the conciliation process.”).

<sup>38</sup> CFI Team, *Model Risk*, CFI (Jan. 13, 2023),

<https://corporatefinanceinstitute.com/resources/knowledge/modeling/model-risk/>.



oversight of large financial institutions at increasing resiliency and mitigating the impact of possible weaknesses on the overall domestic economy.<sup>39</sup> The intent of this regulatory approach is to serve the dual goals of preventing large-scale institutional failures, while simultaneously mitigating the potential impact these failures would have on the financial market more broadly.<sup>40</sup>

Financial institutions have been using complex mathematical models for decades. During the 2007-2008 global financial crisis, the American economy narrowly avoided large-scale failures of financial institutions. While the causes and outcomes of that crisis are beyond the scope of this Essay, financial economists have explained that the failure of certain mathematical models was a key contributor to the crisis.<sup>41</sup> A mathematical model is a system with three essential components: (1) an input; (2) an analysis; and (3) an output.<sup>42</sup> This system takes a set of data or assumptions, analyzes that data to identify patterns and commonalities, and reports estimates, recommendations, or decisions.<sup>43</sup> Model risk refers to the potential negative consequences that an institution may face because errors in the development, implementation, or use of mathematical models contributed to flawed and inaccurate decisions.<sup>44</sup>

## B. Model Risk Management Principles Applied to HR Technology

In 2011, as part of remedial regulatory efforts in the wake of the financial crisis, the Federal Reserve and OCC issued joint supervisory guidance regarding model risk management in the form of SR 11-7.<sup>45</sup> Financial regulators acknowledge this guidance as the definitive statement regarding the MRM framework.<sup>46</sup>

SR 11-7 does not explicitly include artificial intelligence under the model risk management framework.<sup>47</sup> However, in 2021, the OCC issued a Model Risk Management Handbook which

---

<sup>39</sup> See Board of Governors of the Federal Reserve System, *Large Financial Institutions*, BD. OF GOVERNORS OF THE FED. RSRV. SYS., <https://www.federalreserve.gov/supervisionreg/large-financial-institutions.htm>.

<sup>40</sup> See Board of Governors of the Federal Reserve System, *Large Financial Institutions*, BD. OF GOVERNORS OF THE FED. RSRV. SYS., <https://www.federalreserve.gov/supervisionreg/large-financial-institutions.htm>.

<sup>41</sup> See Felix Salmon, *Recipe for Disaster: the Formula that Killed Wall Street*, WIRED MAGAZINE (Feb. 23, 2009), <https://www.wired.com/2009/02/wp-quant/> (“[the] Gaussian copula formula will go down in history as instrumental in causing the unfathomable losses that brought the world financial system to its knees.”).

<sup>42</sup> See Rob Trippe, *SR 11-7 and Corporate Finance Modelling: Managing Risk and Promoting Success*, FP&A Trends (May 11, 2017), [https://fpa-trends.com/article/sr-11-7-and-corporate-finance-modelling-managing-risk-and-promoting-](https://fpa-trends.com/article/sr-11-7-and-corporate-finance-modelling-managing-risk-and-promoting-success#:~:text=Since%20the%20great%20recession%2C%20the%20Federal%20Reserve%20has,are%20financial%20institution%E2%80%99s%20model%20developers%2C%20owners%20and%20users.)

[success#:~:text=Since%20the%20great%20recession%2C%20the%20Federal%20Reserve%20has,are%20financial%20institution%E2%80%99s%20model%20developers%2C%20owners%20and%20users.](https://fpa-trends.com/article/sr-11-7-and-corporate-finance-modelling-managing-risk-and-promoting-success#:~:text=Since%20the%20great%20recession%2C%20the%20Federal%20Reserve%20has,are%20financial%20institution%E2%80%99s%20model%20developers%2C%20owners%20and%20users.) See also CFI Team, *Model Risk*, CFI (Jan. 13, 2023), <https://corporatefinanceinstitute.com/resources/knowledge/modeling/model-risk/>.

<sup>43</sup> See Mark Carey and Jo Paisley, *Model Risk Management at the Crossroads: Meeting New Demands with Limited Resources* (Sept. 6, 2018), [https://www.garp.org/garp-risk-institute/model\\_risk\\_management](https://www.garp.org/garp-risk-institute/model_risk_management).

<sup>44</sup> CFI Team, *Model Risk*, CFI (Jan. 13, 2023), <https://corporatefinanceinstitute.com/resources/knowledge/modeling/model-risk/>.

<sup>45</sup> See *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>.

<sup>46</sup> See *Adoption of Supervisory Guidance on Model Risk Management*, FDIC (June 07, 2017), <https://www.fdic.gov/news/financial-institution-letters/2017/fil17022.html>; *Sound Practices for Model Risk Management: Supervisory Guidance on Model Risk Management*, OCC (Apr. 4, 2011), <https://www.occ.gov/news-issuances/bulletins/2011/bulletin-2011-12.html>. See also OCC 2021-39.

<sup>47</sup> See *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>.

explained that some AI might meet the guidance’s definition of a model.<sup>48</sup> The handbook clarified that AI is usually based on complex mathematical methods even though its outputs may not always be quantitative.<sup>49</sup> Financial institutions have embraced regulatory guidance such as SR 11-7 and have implemented MRM governance structures that assess the risks associated with their use of AI.<sup>50</sup>

Considering the way financial institutions have applied MRM principles to their use of mathematical models, employers should adopt these same MRM principles to AI tools used for employment decision-making. The following sections introduce the core MRM principles from the financial industry and explain how these concepts are transferable to managing the risks associated with HR technology. The first section will discuss model validation generally and explain the importance of ongoing monitoring and proportionality for validation activities. The second section will explain the guiding principle of “effective challenge” and outline its core elements. The final section will clarify whether these MRM principles apply to AI tools that employers purchase from vendors.

### 1. Model Validation Generally

In the MRM context, SR 11-7 describes model validation as “the set of processes and activities intended to verify that models are performing as expected, in line with their design objectives and business uses.”<sup>51</sup> Notably, model validation activities cannot be limited to solely ensuring legal compliance. This idea is consistent with the EEOC’s mission and vision. While the EEOC is responsible for enforcing federal laws that prohibit employment discrimination, its vision is to promote “[r]espectful and inclusive workplaces with equal employment opportunity for all.”<sup>52</sup> Thus, while employers should strive to ensure legal compliance with the use of AI for employment decision-making, they should also consider how their use of AI can promote equal employment opportunity for all in the workplace.

The following general principles underlying model validation similarly apply to HR technology. First, model validation activities should involve evaluating the model for conceptual soundness to ensure the model is designed and constructed according to industry practice.<sup>53</sup> Similar to AI in the financial industry, employers would benefit from evaluating their AI tools to confirm the quality of the tool’s design. This evaluation may include reviewing empirical evidence supporting the AI tool and determining whether it is consistent with published research.<sup>54</sup> The published research

---

<sup>48</sup> <https://occ.gov/publications-and-resources/publications/comptrollers-handbook/files/model-risk-management/index-model-risk-management.html> (Page 4 in the PDF).

<sup>49</sup> <https://occ.gov/publications-and-resources/publications/comptrollers-handbook/files/model-risk-management/index-model-risk-management.html> (Page 4 in the PDF).

<sup>50</sup> See, e.g., PWC, Financial risk analytics and modeling, <https://www.pwc.com/us/en/industries/financial-services/regulatory-services/financial-risk-analytics.html> (describing the 20-year development of PWC’s MRM products and expertise since the release of the OCC 2000-16 guidance).

<sup>51</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 9).

<sup>52</sup> <https://www.eeoc.gov/overview>.

<sup>53</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 11).

<sup>54</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 11).

may involve the explanations of regulators, such as fact sheets, best practices, and formal regulations. For instance, the EEOC's May 2022 ADA guidance provides an assessment for companies and vendors to ensure that AI tools do not discriminate against individuals with disabilities or collect certain medical information prohibited by law.<sup>55</sup> Further, lawyers, HR professionals, and AI vendors are constantly providing guidance, and best practices while raising awareness of the risks associated with AI through various means.<sup>56</sup>

Model validation should also involve ongoing monitoring to recognize the model's limitations.<sup>57</sup> SR 11-7 encourages banks to annually review their models to determine whether their validation activities are adequately addressing the flaws and risks that arise.<sup>58</sup> For employment decision-making, lawyers and HR professionals have similarly recommended an annual audit of AI tools to catch any unlawful biases.<sup>59</sup> Now some city and state legislative proposals are even demanding such audits.<sup>60</sup> Based on this, employers should continually monitor their AI tools to determine whether they need any adjustment, redevelopment, or replacement.

Under SR 11-7, validation activities should also incorporate an outcome analysis to compare the model's initial predictions to the actual outcomes.<sup>61</sup> Employers can adopt this recommendation by comparing the data forming the traits of successful employees with the AI tool's recommendations for which candidates to hire. By comparing this data to the ultimate applicant recommendations, employers can once again increase awareness of unlawful biases before making a series of employment decisions.

Lastly, proportionality is another general MRM principle that employers should implement. In this context, proportionality refers to the application of MRM principles to financial institutions of varying size and complexity.<sup>62</sup> Specifically, the Federal Reserve and the OCC emphasize that SR 11-7 "should be applied as appropriate to all banking organizations supervised by the Federal Reserve, taking into account each organization's size, nature, and complexity, as well as the extent and sophistication of its use of models."<sup>63</sup> Therefore, the MRM framework as outlined in SR 11-

---

<sup>55</sup> <https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence>.

<sup>56</sup> See, e.g., Littler Mendelson P.C., "What Are the Key Areas Our Organization Should Keep Top of Mind When Rolling Out an AI-Based Hiring Tool?," Littler.com (January 25, 2021), <https://www.littler.com/publication-press/publication/what-are-key-areas-our-organization-should-keep-top-mind-when>; See also Gibson, Dunn & Crutcher LLP, "10 Evolving AI Compliance Considerations for Employers," Gibsondunn.com (May 6, 2021), <https://www.gibsondunn.com/10-evolving-ai-compliance-considerations-for-employers/>.

<sup>57</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 12).

<sup>58</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 10).

<sup>59</sup> *EEOC Focuses on Hiring Bias Claims: A 5-Step Compliance Plan for Staffing Firms*, FISHER PHILLIPS (Nov. 14, 2022), <https://www.fisherphillips.com/print/v2/content/30627/eeoc-focuses-on-hiring-bias-claims%3A-a-5-step-compliance-plan-for-staffing-firms.pdf>.

<sup>60</sup> New York Local Law 144; California State proposal

<sup>61</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 13).

<sup>62</sup> Federal Reserve Board, "SR 11-7: Guidance on Model Risk Management," (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm>.

<sup>63</sup> Federal Reserve Board, "SR 11-7: Guidance on Model Risk Management," (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm>.

7 does not propose a “one size fits all” approach to model validation. An effective management of model risk requires understanding the proportional benefits sought to be achieved using the model as well as the potential damage that the financial institution will incur when the model is wrong.<sup>64</sup>

This call for proportionality based on an organization’s size and complexity equally applies to businesses deploying HR technology. Fortune 500 companies as well as small businesses can find value in AI for their HR functions.<sup>65</sup> Accordingly, their approaches to model validation may differ because of their differing size. Proportionality would also apply to the range of employment decisions upon which employers rely on AI. For example, if an employer only uses AI for recruiting, screening resumes, and interviewing, that employer likely exposes itself to fewer risks compared to an employer that uses AI for performance management decisions as well. Thus, employers should tailor their model validation activities to the extent that they deploy AI in their business.

## 2. Model Validation through Effective Challenge

According to SR 11-7, effective challenge is a “guiding principle” for managing model risk.<sup>66</sup> Effective challenge is the “critical analysis [performed] by objective, informed parties who can identify model limitations and assumptions and produce appropriate changes.”<sup>67</sup> Simply put, effective challenge is objective analysis of a model to prevent harm, such as employment discrimination, and address defects as they arise. Effective challenge requires (1) competence, (2) independence and influence, and (3) incentives.<sup>68</sup>

### a) Competence

SR 11-7 highlights the need for “informed parties” who objectively analyze a model.<sup>69</sup> This refers to the first key component of effective challenge, namely competence.<sup>70</sup> The financial industry’s experience in the wake of the 2008 global financial crisis and the subsequent adoption of the Federal Reserve’s supervisory guidance underscores the necessity of competence for objective analysis. The financial industry realized that model validation cannot be competently performed, and should not be performed, by lay businesspeople with no technical background.

---

<sup>64</sup> Federal Reserve Board, “SR 11-7: Guidance on Model Risk Management,” (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm>.

<sup>65</sup> <https://www.uschamber.com/co/run/technology/artificial-intelligence-small-business-applications>.

<sup>66</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 4).

<sup>67</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 4).

<sup>68</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Pages 4, 7, 9).

<sup>69</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 4).

<sup>70</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 4).

The staff performing validation activities should have the necessary knowledge, skills, and experience to objectively analyze the models.<sup>71</sup> In practice, these people need the important combination of technical skills and experience along with significant familiarity with the business uses of the model. Technical experience is necessary because the models may vary based on the complexity of the model's design and application.<sup>72</sup>

For employers, this means they cannot rely solely on HR professionals as that discipline is currently defined or solely on software developers, because each would typically lack the competence necessary to be the "informed party" who can objectively analyze AI tools. For these purposes, HR professionals armed only with an issue-spotting "checklist"<sup>73</sup> would not likely possess the knowledge and skills to understand the uses of the employment data, and they would not understand the technical complexities of AI. Similarly, an AI software developer would not fully understand the HR processes and decision-making without training.

Based on this discrepancy, employers should recognize the need to hire or train employees who have sufficient knowledge and experience with HR management and AI development. Employers could consider upskilling and reskilling programs to address this competency gap.<sup>74</sup> Employers could train HR professionals to have the requisite technical skills to understand the complexities of AI, especially concerning employment decision-making. Similarly, employers could train their in-house software developers to have the necessary background in HR management. Companies may also consider training their in-house employment lawyers to adequately understand the development, implementation, and use of AI tools for employment decision-making. These options may help employers assign competent staff who can objectively analyze AI tools to ensure compliance with the law and the fulfillment of design objectives and business uses.<sup>75</sup>

## **b) Independence and Influence**

Objective analysis requires the people who perform the analysis to have sufficient independence from model creators and influence within the organization to ensure that employment decision-makers can properly address and promptly correct the issues they discover during the validation

---

<sup>71</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 9-10).

<sup>72</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 9-10).

<sup>73</sup> See, e.g., World Economic Forum, *Human-centred artificial intelligence for human resources* (retrieved April 20, 2022), from <https://www.weforum.org/projects/human-centred-artificial-intelligence-for-human-resources/>; Rieke, A., & Bogen, M. *Help wanted: An examination of hiring algorithms, equity, and bias* (retrieved April 20, 2022), <https://upturn.org/work/help-wanted/>; Lohr, Steve. "Group backed by top companies moves to combat A.I. bias in hiring." <https://www.nytimes.com/2021/12/08/technology/data-trust-alliance-ai-hiring-bias.html>; Secretariat, Treasury Board of Canada. *Algorithmic impact assessment tool* (2021), (retrieved Apr 20, 2022), from <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html>.

<sup>74</sup> Joseph Romsey, *5 Tips to Help Workers Upskill and Adapt to Artificial Intelligence*, SHRM (Nov. 30, 2020), <https://www.shrm.org/ResourcesAndTools/hr-topics/technology/Pages/How-HR-Can-Help-Workers-Upskill-and-Adapt-to-Artificial-Intelligence-5-Tips.aspx>.

<sup>75</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 9).

process.<sup>76</sup> Organizations should also create an internal governance structure that supports and enhances the independence and influence of those engaged in model validation.<sup>77</sup>

SR 11-7 explains that model validation involves independence from those who create, implement, and use the model.<sup>78</sup> Thus, employees who are analyzing the model ideally should not be involved in the development and deployment of the model.<sup>79</sup> If model developers and users are involved in validation activities, an independent party should thoroughly examine their work to ensure its objectivity.<sup>80</sup>

As validation activities progress, this competent and independent staff should be appropriately influential inside the organization, so the senior leaders trust the results of the analysis and recognize its objectivity.<sup>81</sup> These leaders must trust the validation personnel to convey the model's successes, failures, and areas requiring immediate correction.<sup>82</sup> Influence is also necessary to implement the recommended changes with the full support of those governing the institution.<sup>83</sup>

Businesses that adopt MRM for AI tools in HR should establish a proper governance structure to protect the integrity of the validation process.<sup>84</sup> Promoting independence and influence should be the aim of this structure.<sup>85</sup> For example, some large financial institutions with sophisticated MRM practices have formed departments reporting to a C-suite executive, such as a Chief Model Risk Officer.<sup>86</sup> For institutions with in-house model developers, the model risk departments should be independent from those that develop and use the models they assess.<sup>87</sup> Supervision by a C-suite executive enhances the influence of the MRM team and mission, because these employees report directly to one of the institution's most senior executives.<sup>88</sup> This structure allows a Chief Model

---

<sup>76</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 4).

<sup>77</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 16).

<sup>78</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 9).

<sup>79</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 9).

<sup>80</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 9).

<sup>81</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 9-10).

<sup>82</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 9-10).

<sup>83</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 9-10).

<sup>84</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 16).

<sup>85</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Pages 4, 18).

<sup>86</sup> See, e.g., <https://www.usbank.com/about-us-bank/leadership.html>. The Vice Chair and Chief Risk Officer of U.S. Bank "oversees all aspects of [U.S. Bank's] risk management activities, including operational risk, credit risk, market risk, model risk, global compliance, independent risk review and regulatory services."

<sup>87</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 9).

<sup>88</sup> See *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 4, 17).

Risk Officer to objectively communicate the results of the model validation process as well as a plan for corrective action to fellow senior executives without direct interference from a separate C-suite executive who oversees the development and implementation of the models.<sup>89</sup>

Similarly, this proposed governance structure can equally apply to corporations that rely on AI tools for employment decision-making. The departments that develop, implement, and use the AI tools should be distinguished from the departments that evaluate the AI tools for legal compliance, ethical concerns, and consistency with business objectives such as equal employment opportunity. Essentially, the model validation department should be independent from the HR department and any in-house department creating the AI tools. Furthermore, legal practitioners have recommended that businesses establish a Chief AI Officer to oversee the company's development and use of AI.<sup>90</sup> These businesses can similarly create a Chief Model Risk Officer to supervise the model validation activities. Therefore, a company's governance structure could entail two separate C-suite executives to ensure proper independence and influence as recommended in the MRM framework. Moreover, suitable incentives may similarly encourage model validation personnel to pursue objectivity and thoroughness during validation activities.

### c) Incentives

SR 11-7 explains that incentives for model validation personnel are more effective when the validation process is independent from the model developers and users.<sup>91</sup> Incentives are also more successful when linked to compelling compensation practices and performance evaluation standards directly connected to the quality of the evaluations and the degree of unbiased review.<sup>92</sup> In implementing SR 11-7, the Federal Reserve recognized that financial incentives to approve mathematical models, including a higher quantity of those models, might incentivize validation personnel to recommend that the institution undertake needless risk.<sup>93</sup> Alternatively, institutions could tie financial incentives to the consistent depth, thoroughness, and objectivity of evaluations instead of the number of models an employee approves.<sup>94</sup> However, there is no greater incentive for corporations than avoiding government investigations, large fines, class action suits, and the correlating negative public and adverse financial impact on their businesses.

Concerning HR technology, employers should similarly establish incentives based on the quality and objectivity of an employee's assessment of the AI tools performing and assisting with HR functions. Financial incentives, such as year-end bonuses and promotion opportunities, could also be based on the accuracy and number of defects identified as well as the level of risk related to

---

<sup>89</sup> See *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 4, 17, 18).

<sup>90</sup> Andrew Ng, *Hiring Your First Chief AI Officer*, HARVARD BUSINESS REVIEW (Nov. 11, 2016), <https://hbr.org/2016/11/hiring-your-first-chief-ai-officer>.

<sup>91</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 9).

<sup>92</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 9).

<sup>93</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 9).

<sup>94</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 9).

each defect. Employers will likely see higher quality validation activities with financial incentives consistent with business goals such as legal compliance and ethical practice, and equal employment opportunity for all candidates.

Furthermore, a corporate culture that fully supports the purpose of model risk management should reinforce these incentives.<sup>95</sup> A corporate culture of transparency and accountability at all levels should be united under the business's goals. A culture of accountability can motivate MRM personnel to pose difficult questions and challenge the objective and subjective decisions during the design and application of AI tools throughout the employment lifecycle.<sup>96</sup> A culture of transparency will encourage MRM personnel to unwaveringly communicate an assessment's results to a Chief Model Risk Officer, who will in turn share the results and corrective action plan with the other senior executives.

Ultimately, a competent model risk management department with sufficient independence, influence, and incentives will allow employers using AI tools to more effectively “verify that models are performing as expected, in line with their design objectives and business uses.”<sup>97</sup>

### 3. Vendor-provided tools

Employers should recognize that the MRM principles and practices apply equally to the AI tools developed in-house and to those tools purchased from vendors. Under Title VII, employers are ultimately responsible for the employment decision, whether made by AI or a human. Under the current federal statutory legal framework, AI vendors will not share any liability for discriminatory uses or outcomes from their products.<sup>98</sup> Once again, SR 11-7 anticipates this dichotomy, explaining that financial institutions should validate vendor-created models as well as models developed in-house.<sup>99</sup> The guidance acknowledges that the validation of vendor-provided models can be challenging since some model components are proprietary and lack transparency.<sup>100</sup>

For employers considering AI tools, there are numerous vendor-provided products on the market, offering a wide range of transparency.<sup>101</sup> Greater transparency provides employers the ability to validate the AI tool consistent with MRM practices. If validation would be especially difficult due

---

<sup>95</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 9).

<sup>96</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 9).

<sup>97</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 9).

<sup>98</sup> The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees, U.S. Equal Employment Opportunity Commission (May 12, 2022), <https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence> (explaining that, in many cases, an employer is responsible under the ADA for its use of algorithmic decision-making tools even if the tools are designed or administered by another entity, such as a software vendor).

<sup>99</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 9).

<sup>100</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 15-16).

<sup>101</sup> See Lauren Smith, *8 HR Tech Companies Using AI to Enable the Future of Work*, ASCENDIFY, <https://www.ascendify.com/blog/8-hr-tech-companies-using-ai-to-enable-the-future-of-work>.



to a lack of transparency, an employer should recall the principle of proportionality and assess whether the intended benefit of using the AI tool outweighs the various risks the employer would incur. This risk assessment should involve legal compliance, ethical concerns, and business objectives.

Furthermore, the risk assessment is particularly important, because the employer is ultimately liable for the harms resulting from the AI tools.<sup>102</sup> For instance, the EEOC does not have direct jurisdiction over vendors who provide AI tools, unlike other agencies such as the FTC.<sup>103</sup> However, the EEOC does have direct jurisdiction over an employer who uses AI tools.<sup>104</sup> Therefore, while vendors sometimes validate the tools they provide, employers should manage the risks accordingly by continually monitoring and performing their own validation of the tools they purchase.<sup>105</sup>

Lastly, many small and mid-size employers may lack the necessary in-house personnel and resources to objectively analyze AI tools. For financial institutions, SR 11-7 recognizes this concern and encourages those institutions to partner with external parties who can assist with or perform the model validation.<sup>106</sup> Small and mid-size employers should similarly consider hiring external consultants to assess the AI tools they use. These external consultants can provide an additional layer of objective analysis that these employers may find beneficial.<sup>107</sup>

#### **IV. AN MRM MITIGATION APPROACH AI-POWERED HR IS CONSONANT WITH ADMINISTRATIVE LAW PRINCIPLES**

In the absence of legislative or regulatory change—neither of which appear to be coming in the coming years—the introduction of MRM as a countermeasure to unintentional non-compliance with federal civil rights laws is undergirded by federal administrative law principles. The voluntary compliance measures that are embodied by MRM as laid out above fit neatly into federal administrative law and Title VII.<sup>108</sup> In operation an MRM undergirded AI-powered HR function

---

<sup>102</sup> The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees, U.S. Equal Employment Opportunity Commission (May 12, 2022), <https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence>.

<sup>103</sup> See K.C. Halm and Nancy Libin, *FTC Warns of Greater Scrutiny Over Biased AI, Offers Best Practices to Mitigate Potential Harm*, DAVIS WRIGHT TREMAINE (Apr. 26, 2021), <https://www.dwt.com/blogs/artificial-intelligence-law-advisor/2021/04/ftc-ai-bias-best-practices-guidance#:~:text=To%20that%20end%2C%20the%20FTC%20is%20moving%20quickly,FTC%20significant%20authority%20to%20act%20in%20this%20area>.

<sup>104</sup> See *EEOC Launches Initiative on Artificial Intelligence and Algorithmic Fairness*, U.S. EQUAL EMP'T OPPORTUNITY COMM'N (Oct. 28, 2021), <https://www.eeoc.gov/newsroom/eeoc-launches-initiative-artificial-intelligence-and-algorithmic-fairness>.

<sup>105</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 15-16).

<sup>106</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 20).

<sup>107</sup> *Supervisory Guidance on Model Risk Management*, Federal Reserve (Apr. 4, 2011), <https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf>. (Page 20).

<sup>108</sup> *Alexander v. Gardner-Denver Co.*, 415 U.S. 36, 44 (1974) (“Cooperation and voluntary compliance were selected as the preferred means for achieving [Title VII’s] goal[s].”); see also *Ricci v. DeStefano*, 557 U.S. 557, 581

could have an effect that reduces the rate that individuals or their representatives file EEOC charges on account of a belief that individuals have been discriminated against.<sup>109</sup> An employer's MRM implementation would not only prevent discrimination, but in the event of automated discriminatory employment decisions occur, remedial action can be immediately taken, preventing the need for EEOC intervention.

However, once a charge is filed with the EEOC, an MRM framework could also favorably affect the outcome of the EEOC investigation, especially if the parties engage in early voluntary mediation.<sup>110</sup> MRM would be highly probative in an EEOC investigation that will endeavor to secure evidence that would determine whether there is a "reasonable cause to believe that the charge [alleging discrimination] is true."<sup>111</sup> Through MRM, the employer would have ample evidence to show the EEOC that it not only complies with the law, it invested in proactive steps to ensure that discrimination did not occur. This direct evidence of compliance would likely ensure the EEOC that the employer goes above and beyond to ensure good-faith compliance with federal employment law. The employer could and should argue that the agency's limited prosecutorial resources should be used against employers who do not engage in such voluntary compliance.<sup>112</sup>

If the EEOC concludes that the evidence does not support a finding of discrimination, it will issue a "not reasonable cause" (or "no cause") finding and a "Notice of Right to Sue" to the party that brought the discrimination charge.<sup>113</sup> In potential litigation that follows, there will be a more defensible final agency action from the start and an MRM framework can be especially useful in the litigation itself because it can evidence a process that the employer followed in good faith to mitigate or eliminate bias before the alleged discriminatory action took place as part of an affirmative defense.<sup>114</sup>

Accordingly, the most logical approach at this stage of AI implementation and regulatory uncertainty, is for the EEOC to urge voluntary compliance instead of leading with enforcement. The MRM approach is expressly intuitive, modular, and subject to best-practices evolution that could earn the formal support of the EEOC once its industry adoption is settled and its outcomes are proven to reduce incidents of civil rights violations. Short of earning EEOC sanction, sound MRM implementation could serve as limitation on liability and damages at all stages of EEOC enforcement and litigation under Title VII.<sup>115</sup> If MRM is adopted by employers and the EEOC indeed proceeds with enforcing Title VII violations a data set will emerge of those who can show

---

(2009) (recognizing "Congress's intent that 'voluntary compliance' be 'the preferred means of achieving the objectives of Title VII.')" (quoting *Firefighters v. Cleveland*, 478 U.S. 501, 515 (1986)).

<sup>109</sup> 42 U.S.C. § 2000e-5(b); 29 C.F.R. § 1601.7.

<sup>110</sup> 29 C.F.R. § 1601.20.

<sup>111</sup> 42 U.S.C. § 2000e-5(b); 29 C.F.R. §§ 1601.15(a), 1601.16(b).

<sup>112</sup> AI & Antidiscrimination: AI Entering the Arena of Labor & Employment Law

YOUTUBE, <https://www.youtube.com/watch?v=Mpzhq3u0hA8> (last visited May 5, 2023).

<sup>113</sup> 42 U.S.C. § 2000e-5(b); EEOC, Filing a Lawsuit, <https://www.eeoc.gov/filing-lawsuit>.

<sup>114</sup> See *Faragher v. City of Boca Raton*, 524 U.S. 775, 807-09 (1998); *Burlington Ind., Inc. v. Ellerth*, 524 U.S. 742, 764-65 (1998) (These cases conclude that an employer can raise an affirmative defense to discrimination liability by establishing by a preponderance of the evidence that (a) the employer exercised reasonable care to prevent and prevent discriminatory behavior "suitable to the employment circumstances" and (b) the employee failed to take advantage of any preventive or corrective opportunities provided by the employer.).

<sup>115</sup> See 29 C.F.R. § 1601.25; see also Stephanie Greene and Christine Neylon O'Brien, *Judicial Review of the EEOC's Duty to Conciliate*, 119 PENN. ST. L. REV. 837 (2015).

good-faith compliance with the law and those whose AI systems were unaudited and violate the law.

To be clear, MRM is not a panacea. It is one of many ways employers can implement internal governance to ensure their AI tools meet their stated purpose and do not discriminate. But, it could be highly probative on the merits for affected entities to defend against charges of disparate treatment.<sup>116</sup> Depending on the quality and traceability of the MRM challenge process, it could also be probative to defend against disparate impact theory enforcement of unintentional discrimination.<sup>117</sup>

Notwithstanding the higher public good appurtenant to laying out guidelines, the EEOC is incorrectly electing to regulate AI in HR through enforcement. In its 2023 Draft Strategic Enforcement Plan, the Chairwoman has listed the elimination of barriers in AI assisted recruitment and hiring as one of its prime “subject matter priorities.”<sup>118</sup> This enforcement priority extends to “technology-related employment discrimination” in which the “EEOC will focus on employment decisions, practices, or policies in which covered entities’ use of technology contributes to discrimination based on a protected characteristic” including “the use of . . . algorithmic decision-making or machine learning, including artificial intelligence . . . used in employment decisions.”<sup>119</sup> In the absence of new legislative policymaking and the absence of timely and specific AI use regulation from the EEOC, the MRM approach to AI-powered HR tools squarely offers an intentional, traceable mechanism of business and legal risk mitigation for companies and the HR industry.

## CONCLUSION

Companies increasingly rely on AI-powered tools to assist with decisions throughout the employment lifecycle. The use of this HR technology certainly forces employers to balance the many benefits with the numerous risks. As a result, employers are eagerly seeking guidance in the absence of clear, consistent federal regulation along with the overlapping and evolving patchwork of state and local laws.

Like all other self-governance and internal compliance mechanisms, the MRM framework from the financial sector is not a silver bullet for employers seeking to use AI for employment decision-making. Rather, it is an established process for managing the risks associated with mathematical models, now including the risks involving the use of AI. The principle of proportionality allows employers to scale their validation efforts depending on the risks, complexity, and range of employment decisions upon which they rely on AI. Employers can also adjust those efforts as the underlying technology and government regulations change.

---

<sup>116</sup> EEOC v. Abercrombie & Fitch Stores, Inc., 575 U.S. 768, 772–74 (2015).

<sup>117</sup> Griggs v. Duke Power Co., 401 U.S. 424, 430–31 (1971) (“[P]ractices, procedures, or tests neutral on their face, and even neutral in terms of intent, cannot be maintained if they operate to ‘freeze’ the status quo of prior discriminatory employment practices.”).

<sup>118</sup> Equal Employment Opportunity Comm’n, Draft Strategic Enforcement Plan, 88 Fed. Reg. 1379, 1381 (Jan. 10, 2023).

<sup>119</sup> Equal Employment Opportunity Comm’n, Draft Strategic Enforcement Plan, 88 Fed. Reg. 1379, 1382 (Jan. 10, 2023).

The key components of objective analysis are critical for managing the risks associated with AI. A competent MRM team with independence, influence, and proper incentives will help employers more successfully ensure that AI tools are performing consistently with the design objectives and business uses.<sup>120</sup> Moreover, applying MRM principles to HR technology can help improve public trust in this technology while supporting legal compliance, ethical use, and equal employment opportunity.

---

<sup>120</sup> *See supra* Part III.B.