Center for Responsible Lending

Comment to:

Board of Governors of the Federal Reserve System, Docket No. OP–1743

Consumer Financial Protection Bureau, Docket No. CFPB–2021– 0004

Federal Deposit Insurance Corporation, RIN 3064–ZA24

National Credit Union Administration, Docket No. NCUA –2021– 0023

Office of the Comptroller of the Currency, Docket ID OCC– 2020–0049

Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, including Machine Learning

July 1, 2021

I. Introduction

Thank you for the opportunity to comment on the Interagency Request for Information on Financial Institutions' Use of Artificial Intelligence, including Machine Learning. The Center for Responsible Lending appreciates the Federal Reserve Board, Consumer Financial Protection Bureau, Federal Deposit Insurance Corporation, National Credit Union Administration, and Office of the Comptroller of the Currency (collectively "the agencies") are focusing on the benefits and risks of artificial intelligence (AI) and machine learning (ML).

The Center for Responsible Lending (CRL) is a nonprofit, non-partisan research and policy organization dedicated to protecting homeownership and family wealth by working to eliminate abusive financial practices. CRL is an affiliate of Self-Help, one of the nation's largest nonprofit community development financial institutions. For 40 years, Self-Help has created asset-building opportunities for low-income individuals, rural communities, women, and families of color. In total, Self-Help has provided over $9 billion in financing to 172,000 homebuyers, small businesses, and nonprofit organizations and serves more than 160,000 mostly low-income families through 72 credit union branches in North Carolina, California, Florida, Illinois, South Carolina, Virginia, Washington, and Wisconsin.

While AI technologies are used for multiple purposes, CRL's comment focuses on financial institutions use of AI/ML in making credit decisions. Our comment focuses on the consumer protection and fair lending risk associated with AI/ML and recommendations on how to create a more equitable financial market. CRL is a signatory to and concurs with the detailed comments of the National Fair Housing Alliance *et al* and reiterates many of their recommendations.

Today, many credit decisions are reached through automated systems – algorithmic risk assessment models that have been shown to produce discriminatory outcomes despite industry assurances that they are free of bias.[1] The agencies should be vigilant about financial institutions' use of AI/ML models and consider the ways in which they can create unjustified outcomes in credit decisions, sometimes without detection. It is critical that the agencies incentivize lenders to remain vigilant to ensure their models are nondiscriminatory. One of the best methods is for lenders to run rigorous fair lending analysis, including a disparate impact analysis, to ensure that risk assessment models do not drive discriminatory outcomes. However, without proper oversight and enforcement from the agencies, many industry players will not conduct these rigorous tests and discrimination will go undetected and unresolved. Moreover, the agencies should conduct their own testing in examinations to ensure lender compliance with the law and to identify patterns of discrimination.

[1] Robert Bartlett, Adair Morse, Richard Stanton, and Nancy Wallace, *Consumer-Lending Discrimination in the FinTech Era*, Haas School of Business UC Berkeley (May 2019), http://faculty.haas.berkeley.edu/morse/research/papers/discrim.pdf.

II. Algorithms Are Not Immune from Discriminating or Creating Clearly Discriminatory Outcomes

As ample research demonstrates, algorithms are not objective or free of potential bias.[2] They are only as good as the data that biased humans program into them. And even when the data itself is not biased, the interactions between the data may produce biased outcomes.

Artificial intelligence and algorithms have been exposed as problematic in various sectors. These examples should serve as a cautionary tale for the use of AI/ML in the financial services space. In 2018, the New York Times published a study finding artificial intelligence – in particular, facial recognition technology – was much less effective when the subject of the analysis was not a white male.[3] While the software was correct 99 percent of the time when the subject in the photo was a white man, when the subject was a darker skinned female, the software was wrong 35 percent of the time.[4] This is because the data set used in artificial intelligence is often reflective of those creating it, who are disproportionately white and male.[5] As Joy Buolamwini, MIT professor, stated "[y]ou can't have ethical A.I. that's not inclusive" and "[w]hoever is creating the technology is setting the standards."[6] This is a fundamental issue with algorithms.

Additionally, in the employment discrimination context, new developments – such as automated hiring systems – have ushered in novel mechanisms for discrimination.[7] "The high bar of proof to demonstrate a disparate impact cause of action under Title VII of the Civil Rights coupled with the "black box" nature of many automated hiring systems, render the detection and redress of bias in such algorithmic systems difficult" and "the automation of hiring both facilitates and obfuscates employment discrimination."[8] Potential discrimination claims are shielded due to the black-box nature of algorithms plus the fact that companies claim the algorithm is a trade secret. This creates an insurmountable and unjust obstacle for disparate impact claimants. Federal Reserve Bank Governor Lael Brainard gives a disturbing example taken from a hiring firm's AI algorithm: "the AI developed a bias against female applicants, going so far as to exclude resumes of graduates from two women's colleges."[9] Brookings' Aaron Klein expanded on this example by stating "[o]ne can imagine a lender being aghast at finding out their AI was making

---

[2] Nicol Turner Lee, Paul Resnick, and Genie Barton, *Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms*, Brookings Institute, May 22, 2019, https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policiestoreduce-consumer-harms/; Claire Cain Miller, *Algorithms and Bias: Q. and A. With Cynthia Dwork*, NY Times, Aug. 10, 2015, https://www.nytimes.com/2015/08/11/upshot/algorithms-and-bias-q-and-a-with-cynthia-dwork.html.

[3] Steve Lohr, *Facial Recognition Is Accurate, if You're a White Guy*, NY Times, February 9, 2018, https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html.

[4] *Id.*

[5] *Id.*

[6] *Id.*

[7] Ifeoma Ajunwa, *Automated Employment Discrimination* (March 15, 2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3437631.

[8] *Id.*

[9] Aaron Klein, Credit Denial in the Age of AI, Brookings Institute, April 11, 2019, https://www.brookings.edu/research/credit-denial-in-the-age-of-ai/.

credit decisions on a similar basis, simply rejecting everyone from a woman's college or a historically black college or university."[10]

In the criminal justice context, COMPAS is an algorithm widely used in the United States to guide sentencing by predicting the likelihood of a criminal reoffending.[11] This system was reported in May 2016 as racially biased. According to the analysis, the system predicts that black defendants pose a higher risk of recidivism than they do, and the reverse for white defendants.[12] Also, predictive policing algorithms have been shown to lead to unjustified over-policing in communities of color.[13] Predictive policing moves police to places where large amounts of crime occurred, which the algorithm views as places where large amounts of arrests occurred. Most of the arrests used by the algorithm are for nonviolent crimes because they are more widespread and predictable, and more nonviolent crime arrests are for black individuals. Thus, the algorithm causes over policing for black neighborhoods, not because there is more crime there than in areas with large white populations, but because those neighborhoods have more arrests, often for discriminatory reasons. As has been demonstrated time and time again, there is enormous racial disparity and bias in the criminal justice system.[14] Algorithms can both build in bias and reinforce bias in a systemic way. Our nation's current reckoning on racial injustice, driven by injustices in policing but extending to every facet of life, must include calls for reform in the use of AI and ML. These technologies must not become a tool to perpetuate systemic inequity and racism.

Moreover, algorithms have been at the center of Medicaid litigation. For example, *K.W. v. Armstrong* was a class action lawsuit representing approximately 4,000 Idahoans with development and intellectual disabilities who receive assistance from the state's Medicaid program.[15] The State of Idaho had used an in-house formula to determine the dollar value of the disability services available to qualifying individuals.[16] A significant number of peoples' "dollar-figure numbers" decreased dramatically.[17] When pressed, the state said that a formula had caused the numbers to drop, but the state considered the formula a trade secret.[18] In litigation the court ordered the state to disclose its formula.[19] The court found that the formula was unconstitutionally arbitrary and ordered the state to fix the formula so it

---

[10] *Id.*

[11] Julia Angwin et. al., *Machine Bias*, Pro Publica, May 23, 2016, https://www.propublica.org/article/machine-biasrisk-assessments-in-criminal-sentencing.

[12] *Id.*

[13] Andrew Guthrie Ferguson, *The Police Are Using Computer Algorithms to Tell If You're a Threat*, TIME Magazine, October 3, 2017, https://time.com/4966125/police-departments-algorithms-chicago/.

[14] Report to the United Nations on Racial Disparities in the U.S. Criminal Justice System, Sentencing Project, April 19, 2018, https://www.sentencingproject.org/publications/un-report-on-racial-disparities/.

[15] Rashida Richardson, Jason M. Schultz, and Vincent M. Southerland, *Litigating Algorithms 2019 US Report: New Challenges to Government use of Algorithmic Decision Systems*, AI Now Institute, New York University, https://ainowinstitute.org/litigatingalgorithms-2019-us.pdf; Jay Stanley, *Pitfalls of Artificial Intelligence Decision making Highlighted in Idaho ACLU Case*, ACLU, June 2, 2017, https://www.aclu.org/blog/privacytechnology/pitfalls-artificial-intelligence-decisionmaking-highlighted-idahoaclu-case.

[16] *Id.*

[17] *Id.*

[18] *Id.*

[19] *Id.*

allocated funds fairly to recipients.[20] In addition, the court ordered the state to test the formula regularly.[21]

These examples provide stark warnings against allowing AI/ML models to bypass robust fair lending scrutiny in the financial services sector. Existing civil rights laws and supervisory policies, including the Fair Housing Act and the Equal Credit Opportunity Act, provide a framework for the agencies to analyze fair lending risk in AI and to engage in supervisory or enforcement actions as appropriate. The agencies must take a leadership role to ensure that financial companies utilize AI/ML technology properly and equitably.

III. Algorithmic Models are Black Boxes

Devising a model's intent is challenging and often impossible. The complex interactions that AI engages in to form a decision can be so opaque that they prevent any party from being able to devise the intent of the machine's creator.[22] For this reason, AI models are referred to as black boxes. When AI programs are black boxes, they are able to form predictions and decisions in the same way as humans, but they are not able to communicate their reasons for making these conclusions.[23] This situation has been analogized to a human attempting to communicate with another highly intelligent species, with both species able to reason and understand but not able to communicate with each other.[24] Scholars have stated that this difficulty in communication "means that little can be inferred about the intent or conduct of the humans that created or deployed the AI, since even they may not be able to foresee what solutions the AI will reach or what decisions it will make."[25] Indeed, a recent paper argues that artificial intelligence is inherently structured in a manner that makes "proxy discrimination" a likely possibility.[26] Consumers have no way of knowing what data is fed into the models, which factors the algorithm used in making the determination, whether there are proxies for protected classes, or whether the algorithm denied credit based on erroneous or biased data.

A. Black-Box AI Models: Neural Networks & Support Vector Models (SVMs)

Neural networks are among the most commonly used models, but these networks are considered black boxes because of their complexity. The structure of a neural network is made up of input nodes, hidden nodes, and output nodes.[27] The complexity arises with the interactions between hidden nodes, which process data from the input nodes to form the output nodes.[28] This is because no node is responsible for

---

[20] *Id.*
[21] *Id.*
[22] Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 Harv. J.L. & Tech. 890 (2018), at 892, 897, 907, https://jolt.law.harvard.edu/assets/articlePDFs/v31/The-ArtificialIntelligence-BlackBox-and-the-Failure-of-Intent-and-Causation-Yavar-Bathaee.pdf.
[23] *Id.* at 907.
[24] *Id.* at 893.
[25] *Id.*
[26] Anya Prince and Daniel B. Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, Iowa Law Review (August 5, 2019), https://ssrn.com/abstract=3347959.
[27] Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 Harv. J.L. & Tech. 890 (2018), at 901, https://jolt.law.harvard.edu/assets/articlePDFs/v31/The-Artificial-IntelligenceBlack-Box-andthe-Failure-of-Intent-and-Causation-Yavar-Bathaee.pdf.
[28] *Id.* at 902.

a distinct function; thousands of nodes overlap each other to form a decision.[29] Humans are able to extract and examine one of these groups of nodes.[30] But because of the different language of AI black boxes, this will likely appear as visual noise to humans.[31] This means that neural networks are often highly unintelligible to humans.

Support Vector Models (SVMs) are also widely used and considered black boxes. Unlike neural networks, which have a lack of transparency that arises from complexity, SVMs are black boxes because they possess geometric relationships that humans cannot visualize.[32]

IV. AI May Use Biased Data to Form Biased Conclusions and the Use of Non-Traditional Variables Places Algorithmic Models at Risk of Not Distinguishing Correlation from Causation

Non-traditional variables increases the likelihood that conclusions will be biased as well as increase the likelihood that AI will draw a conclusion that there is causation where there is only correlation.[33] Nontraditional variables include data obtained from internet search histories, shopping patterns, social media activity, and various other consumer-related inputs.[34] This non-traditional information can be fed into machines, which can draw conclusions based on the patterns it observes in the dataset.[35] This is a major concern because financial technology companies are using nontraditional data more and more to make consumer credit decisions. As one article put it: "If there are data out there on you, there is probably a way to integrate it into a credit model. But just because there is a statistical relationship does not mean that it is predictive, or even that it is legally allowable to be incorporated into a credit decision."[36] The following is an example of the use of non-traditional variables in a manner that causes bias from the article *Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms*:

> Latanya Sweeney, Harvard researcher and former chief technology officer at the Federal Trade Commission (FTC), found that online search queries for African-American names were more likely to return ads to that person from a service that renders arrest records, as compared to the ad results for white names. Her research also found that the same differential treatment occurred in the micro-targeting of higher-interest credit cards and other financial products when the computer inferred that the subjects were African-Americans, despite having similar backgrounds to whites. During a public presentation at a FTC hearing on big data, Sweeney demonstrated how a web site, which marketed the centennial celebration of an all-black

---

29 *Id.*
30 *Id.*
31 *Id.* at 903.
32 *Id.*
33 White & Case, *Algorithms and Bias: What Lenders Need to Know*, January 20, 2017, https://www.whitecase.com/publications/insight/algorithms-and-bias-what-lenders-need-know. *See also* Ian Ayres, *Testing for Discrimination and the Problem of Included Variable Bias* at p. 6 (2010), https://www.law.upenn.edu/live/files/1138-ayresincludedvariablebiaspdf.
34 *Id.*
35 *Id.*
36 Aaron Klein, *Credit Denial*, Brookings Institute (April 11, 2019), https://www.brookings.edu/research/creditdenial-in-the-age-of-ai/.

fraternity, received continuous ad suggestions for purchasing "arrest records" or accepting high interest credit card offerings.[37]

The *National Fair Housing Alliance v. Facebook* lawsuit serves as an additional example of biased data used in an algorithm. The main allegation in the lawsuit was that Facebook's advertising platform contained pre-populated lists that allowed advertisers to place housing, employment, and credit ads that could exclude certain protected groups, such as African-Americans, Hispanics, and Asian Americans.57 The plaintiffs also challenged that Facebook permitted advertisers to include or exclude Facebook users from receiving ads based on their sex or age, or based on demographics, behaviors and interests that were associated with protected classes.[38] Plaintiffs alleged that Facebook "extracts data from its users' online behavior, both on Facebook and off, and uses algorithms designed to sort that data, process it, and repackage it to group potential customers into new and salient categories for advertisers to choose from when targeting their ads."[39] Therefore, data sets were allegedly being crafted to increase the likelihood of particular outcomes with groups that were the equivalent of protected classes.

Facebook's inclusion of certain groups and exclusion of others resulted in groups being disproportionately targeted by predatory lenders or excluded from reasonable and beneficial loans. This shows the risk of discrimination that can come with AI, and that past bias (in this case, Facebook's selection of particular categories) can result in current bias (the discriminatory outcomes).

Furthermore, algorithms do not distinguish causation from correlation or know when it is necessary to gather additional data to form a sound conclusion. One notable example is social media. This is particularly relevant in the lending context, as some fintech lenders may use social media data as a predictor of default. But using this information might interfere with other more important and relevant indicators.[40]

Additionally, although consumers can check their credit reports for false information, "consumers cannot easily verify the myriad forms of nontraditional data that could be fed into a credit assessment algorithm. Consumers may not know whether an algorithm has denied them credit based on erroneous data from sources not even included in their credit reports."[41]

While some argue that the usage of non-traditional variables is beneficial in providing targeted information to different groups, it can lead to "unfair or discriminatory lending decisions if not

---

[37] Nicol Turner Lee, Paul Resnick, and Genie Barton, *Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms*, Brookings Institute, May 22, 2019, available at https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policiestoreduce-consumer-harms/.

[38] *National Fair Housing Alliance v. Facebook, Inc.*, No. 1:18-cv-02689 (S.D.N.Y), https://nationalfairhousing.org/facebook-settlement/.

[39] *Id.*

[40] First Amended Complaint, *National Fair Housing Alliance v. Facebook, Inc.*, No. 1:18-cv-02689, ¶ 52 (S.D.N.Y. June 25, 2018).

[41] White & Case, *Algorithms and Bias: What Lenders Need to Know*, https://www.whitecase.com/publications/insight/algorithms-and-bias-what-lenders-need-know.

appropriately implemented and monitored."[42] It can lead to decisions where patterns of discrimination are perpetuated from the initial entry of data to the conclusion.

While artificial intelligence holds great promise, we must not assume it is objective or bias-free. Rather than shield algorithms, we must find ways to test and audit algorithms to ensure they do not perpetuate bias or cause unjustified disparate outcomes. As such, it is critical for there to be a regulatory framework in place that allows for the regular examination of the data used, analysis of the algorithmic model and calculations, and evaluations of any disparate impacts. Such review cannot be left to financial companies themselves; scrutiny is essential to ensure accountability and to prevent widespread harm.

V. Recommendations

Because of the pervasive and hidden discrimination algorithmic systems can produce, the agencies must take action to ensure non-discriminatory and equitable outcomes for all who participate in the financial services market. The following recommendations – and much further detail – is included in the National Fair Housing Alliance, *et al* comment letter, to which CRL is a signatory.

1) As a threshold issue, the agencies should define "model risk" to include the risk of discriminatory or inequitable outcomes for consumers, rather than just the risk of financial loss to a financial institution.[43] Fair lending risk and equity must be considered in every aspect of a model.

2) After reviewing stakeholder responses to the RFI, the agencies should issue a detailed action plan that outlines next steps and offers opportunities for further stakeholder engagement.

3) The agencies should conduct in-depth supervisory reviews of financial institutions' use of AI, including evaluating compliance with fair lending laws.

4) The agencies should encourage and support public research that analyzes specific uses of AI in financial services and the impact of AI in financial services for consumers of color and other protected classes. For example, the agencies should encourage the CFPB and the Federal Housing Finance Agency to release more de-personalized loan-level data from the National Survey of Mortgage Originations and the National Mortgage Database so researchers, advocacy groups, and the public can study potential discriminatory and inequitable outcomes in the financial services market, especially as they relate to the use of AI.

5) The agencies should expeditiously hire staff with specialized knowledge of AI and fair lending risks and integrate those staff members into all work related to modeling, including supervisory and enforcement issues, policy statements, and rulemakings.

---

[42] *Id.*

[43] *See* Federal Reserve Board and OCC, *Supervisory Guidance on Model Risk Management*, SR 11-7 at 3 (Apr. 4, 2011), https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf ("Model Risk Management Guidance") (defining "model risk" to focus on the financial institution rather than the consumer by stating that "[m]odel risk can lead to financial loss, poor business and strategic decision making, or damage to a bank's reputation").

6) The agencies should ensure that all AI stakeholders—including regulators, financial institutions, and tech companies—receive regular fair lending and racial equity training. Given the ever-evolving nature of AI, the training should be updated and provided on a periodic basis.

7) The agencies should ensure agency staff working on AI issues reflect diversity, including diversity based on race and national origin. In addition, the agencies should encourage financial institutions to engage diverse staff for the AI development and design teams. Research has shown that diverse teams are more innovative and productive[44] and that companies with more diversity are more profitable.[45] Moreover, people with diverse backgrounds and experiences bring unique and important perspectives to understanding how data impacts different segments of the market.[46]

8) The agencies should prioritize transparency. They should: 1) strive to share their data, models, decisions, and proposed solutions so that all of the key stakeholders can stay informed of and comment on the potential impact of proposed agency actions; and 2) require financial institutions to share information with the public regarding their AI systems to enable researchers and those impacted to evaluate the efficacy and impact of the systems.

9) The agencies should engage with a diverse group of key stakeholders, including civil rights organizations, consumer advocates, and impacted communities in order to receive ongoing input and feedback on these important decisions.

Thank you for considering our views on this critical issue.

Sincerely,

Center for Responsible Lending

---

[44] *See, e.g.*, John Rampton, *Why You Need Diversity on Your Team, and 8 Ways to Build It*, Entrepreneur (Sept. 6, 2019), https://www.entrepreneur.com/article/338663.

[45] *See, e.g.*, David Rock and Heidi Grant, *Why Diverse Teams Are Smarter*, Harvard Business Review (Nov. 4, 2016), https://hbr.org/2016/11/why-diverse-teams-are-smarter (reporting that companies in the top quartile for ethnic and racial diversity in management were 35% more likely to have financial returns above their industry mean, and those in the top quartile for gender diversity were 15% more likely to have returns above the industry mean).

[46] *See, e.g.*, Inioluwa Deborah Raji et al., *Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing*, in Conference on Fairness, Accountability, and Transparency 33, 39 (2020), https://dl.acm.org/doi/pdf/10.1145/3351095.3372873; Model Risk Management Guidance at 4.