



July 1, 2021

Office of the Comptroller of the Currency: Docket ID OCC-2020-0049
Board of Governors of the Federal Reserve System: Docket No. OP- 1743
Federal Deposit Insurance Corporation: RIN 3064-ZA24
Consumer Financial Protection Bureau: Docket No. CFPB-2021-0004
National Credit Union Administration: Docket No. NCUA-2021-0023

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

To the Agencies:

Please accept this comment on the use of artificial intelligence including machine learning.

The National Community Reinvestment Coalition (NCRC) consists of more than 600 community-based organizations, fighting for economic justice for almost 30 years. Our mission is to create opportunities for people and communities to build and maintain wealth. NCRC members include community reinvestment organizations, community development corporations, local and state government agencies, faith-based institutions, fair housing and civil rights groups, minority and women-owned business associations, and housing counselors from across the nation. NCRC and its members work to create wealth-building opportunities by eliminating discriminatory lending practices, which have historically contributed to economic inequality.

NCRC applauds the decision by the above-listed agencies (the Agencies) to address the use of these tools. We hold the view that such action should occur shortly, as the use of artificial intelligence (AI) and machine learning (ML) is increasing, and these practices pose both potential and peril for the ability of consumers to access credit in non-discriminatory, transparent, and inclusive ways.

NCRC urges the Agencies to prioritize activities to ensure that the use of AI and ML develops in ways that are non-discriminatory. Adopting new technologies will not occur in a vacuum but in markets whose histories include extensive evidence of unfair and exclusionary practices.

NCRC believes that future supervision, rulemaking, and enforcement of AI and ML should focus on the principles of equity, transparency, and accountability.

A commitment to *equity* includes, but is not limited to, explicitly identifying discriminatory practices as a type of risk by providing guidance on the disparate impact standard by updating the language to state that the creditor practice must meet a "substantial, legitimate, non-discriminatory interest, and by emphasizing diversity in hiring.

A commitment to *transparency* includes, but is not limited to, requiring lenders to prioritize explainability, engaging with community groups, addressing how consumers can determine if incorrect data was used to evaluate their creditworthiness, housing a testing data set at the FFIEC, updating adverse-action notice requirements, and publishing research on the use of algorithmic underwriting.

A commitment to *accountability* includes, but is not limited to, holding lenders accountable to use inclusive training data sets and not to use variables that are proxies for protected class status,¹ investing in staff and resources to facilitate robust supervision and enforcement, stating that lenders are accountable for their models even if they contract with third-party vendors, and providing guidance on how and when lenders can find less discriminatory alternatives.

The NCRC also helped draft a separate set of comments collectively submitted by a group of civil rights, consumer, technology, and other organizations ("Joint Comment"), which addresses these principles and many of the questions raised by the Agencies.² This separate comment provides additional information and requests for guidance sought by the Agencies that NCRC believes will foster the non-discriminatory use of AI, machine learning, and alternative data.

The NCRC regularly convenes discussions with its financial industry councils comprised of national banks, community banks, mortgage lenders, and financial technology firms to consider important issues related to the financial markets and these discussions helped inform this comment. The NCRC's Innovation Council for Financial Inclusion has focused on artificial intelligence, machine learning, algorithms, and alternative data as its members are fintech companies.³ The statement of the Innovation Council on Disparate Impact, which focuses on artificial intelligence, is discussed in more detail below in response to Questions 4 and 5.

1. ANSWERS to INDIVIDUAL QUESTIONS

Question 3) For which uses of AI is lack of explainability more of a challenge? Please describe those challenges in detail. How do financial institutions account for and manage the varied challenges and risks posed by different uses?

To create explainable and interpretable models, the Agencies should:

- a) require lenders to use explainable models or explainability techniques that can accurately describe the reasons for a decision made by an AI model,
- b) To fulfill the purpose of adverse-action notices, the Consumer Financial Protection Bureau (the "CFPB" or "the Bureau") should reconsider the content and form of adverse-action notices,

¹ 12 C.F.R. pt. 1002, Supp. I, ¶ 1002.2(p)-4

² The group includes the National Fair Housing Alliance, Fair Play AI, ACLU, Lawyer's Committee, Upturn, AI Blindspot, FinRegLab, Tech Equity Collaborative, BLDS, and Relman Colfax.

³ The members of NCRC's Innovation Council are Affirm, Lending Club, PayPal, Square, Oportun, and Varo.

- c) require lenders only to use data elements that would give turned-down applicants the agency to improve their creditworthiness,
- d) shorten the time between the decision and the delivery of the notice,
- e) staff and invest in the ability to test explainability, including tests for explainability among different demographic groups, and
- f) issue a statement that models that are not explainable have heightened risks for discriminatory impacts in AA notices, and
- g) provide guidance on how consumers can resolve instances where incorrect data has been used to evaluate their creditworthiness.
 - a. *Require lenders to use explainable models or to use explainability techniques that can accurately describe the reasons for a decision made by an AI model.*

In a 2021 survey, 65 percent of 100 C-suite executives at AI-focused corporations felt their companies would not be able to explain how their AI models worked, and only 20 percent said their companies were actively monitoring their models for fairness.⁴

To ensure that credit markets are non-discriminatory, lenders must understand how their models work. If a lender cannot explain its basis for a decision to an applicant, then it risks losing the trust it may have had with that individual. If the public ceases to expect financial institutions to provide explanations for their decision-making, then its lack of confidence will undermine the overall integrity of financial markets.

Lenders can achieve explainability in one of two ways: by selecting model types that are inherently explainable or using “post-hoc” tools that explain the results of models that were not fully explainable at the outset of their development.⁵

For lenders that choose the latter, the Agencies should ask lenders to provide weights for the most significant elements of the decision. For example, lenders could use Shapley values, a tool developed in game theory, to determine the level of the “contribution” made by each variable to the outcome, or contract with vendors to avail their organizations of “bolt-on” explainability solutions.

However, post-hoc methods resolve problems that would otherwise have not existed if the models were originally constructed with explainable information and interpretable models. Explainable models, which can incorporate subject-matter expertise and other aspects of human reasoning, generally present a better option. Explainable models can establish safeguards against the kinds of nonsensical conclusions that may occur when ML observes patterns in training data sets that contradict common sense, such as by making models with monotonic constraints⁶ or other types of rule sets.

Information lenders must report to show that they are providing accurate explanations of their underwriting models include:

⁴ FICO. “The State of Responsible AI.” FICO, May 28, 2021. <https://mobileidworld.com/fico-urges-businesses-prioritize-ethics-ai-development-052805/>.

⁵ FICO. “The State of Responsible AI.” FICO, May 28, 2021. <https://mobileidworld.com/fico-urges-businesses-prioritize-ethics-ai-development-052805/>.

⁶ Ajay Tiwari. “Application of Monotonic Constraints in Machine Learning Models: A Tutorial on Enforcing Monotonic Constraints in XGBoost and LightGBM Models.” *Medium*, May 1, 2020. <https://medium.com/analytics-vidhya/application-of-monotonic-constraints-in-machine-learning-models-334564bea616>.

- Showing they have compliance systems that can explain reasons for all decisions and provide evidence to support each decision. Compliance systems test the accuracy of explanations.
- Verifying that these systems provide explanations that are understandable to individual users. Given that a one-size-fits-all approach is not always possible, systems could use loan-level data to verify that the explanations are equally accurate for protected classes and white applicants.
- Building a system that includes a check to ensure that applications are made only when the algorithm has enough data to generate a decision with confidence. A knowledge-limit constraint, where a decision is not made if an underwriting system lacks enough data for training, becomes particularly important if a training data set does not include a demographically inclusive set of cases.⁷

At the minimum, explanations should provide a benefit to users. Certainly, explanations that meet the needs for compliance, for societal trust, or the needs of the system’s owner have value, but the priority should be to ensuring that the standard of user benefit is met.⁸

Some lenders may contend that using explainable models requires them to compromise on the predictive power of a model, and therefore, any expectation of building meaningful models implies a constraint against their business interests. However, explainable artificial intelligence (XAI) is not necessarily less accurate compared to black-box AI. Moreover, XAI brings values such as transparency and accountability to markets that black box models cannot.⁹ Safeguards to protect individuals deserve strong regulatory support. Indeed, European regulators are seeking to place AI-based credit underwriting in a special class of high-risk AI systems. The European Union’s 2021 proposed *Artificial Intelligence Act* notes:

Another area in which the use of AI systems deserves special consideration is the access to and enjoyment of certain essential private and public services and benefits necessary for people to fully participate in society or to improve one’s standard of living. In particular, AI systems used to evaluate the credit score or creditworthiness of natural persons should be classified as high-risk AI systems, since they determine those persons’ access to financial resources or essential services such as housing, electricity, and telecommunication services. AI systems used for this purpose may lead to discrimination of persons or groups and perpetuate historical patterns of discrimination, for example based on racial or ethnic origins, disabilities, age, sexual orientation, or create new forms of discriminatory impacts.¹⁰

NCRC’s 2020 policy agenda states that “every person in a community, regardless of their race, age or socioeconomic status, should have the opportunity to build wealth. Equal access to financial products and services is critical, and building community prosperity requires a long-term plan to expand and preserve access to credit and capital.”¹¹

⁷ Joy Buolamwini and Timnit Gebru. “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.” In *Proceedings of Machine Learning Research*, 81:1–15. New York, NY, 2018. <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>.

⁸ P. Jonathon Phillips, Carina A. Hahn, Peter C. Fontana, David A. Broniatowski, and Mark A. Przybocki. “Four Principles of Artificial Intelligence.” Gaithersburg, Maryland: National Institute of Standards and Technology, August 2020. <https://www.nist.gov/document/four-principles-explainable-artificial-intelligence-nistir-8312>.

⁹ Rudin, Cynthia and Radin, Joanna. “Why Are We Using Black Box Models in AI When We Don’t Need To? A Lesson From An Explainable AI Competition.” *Harvard Data Science Review* 1, no. 2 (2019). <https://doi.org/10.1162/99608f92.5a8a3a3d>.

¹⁰ European Union. Proposal for a Regulation laying down harmonized rules on artificial intelligence (n.d.). <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence>.

¹¹ National Community Reinvestment Coalition. “2020 Policy Agenda for the 116th Session of Congress,” May 2020. <https://ncrc.org/2020-policy-agenda-for-the-116th-session-of-congress/>.

The Agencies should issue a statement that unexplainable models do not comply with Regulation B.

- b. *To fulfill the purpose of adverse-action notices, the CFPB should reconsider the content and form of adverse-action notices.*

In its 2019 Fair Lending Report, the CFPB opined that the “existing regulatory framework has built-in flexibility that can be compatible with AI algorithms. For example, although a creditor must provide the specific reasons for an adverse action, the Official Interpretation to ECOA’s implementing regulation, Regulation B, provides that a creditor need not describe how or why a disclosed factor adversely affected an application, or, for credit scoring systems, how the factor relates to creditworthiness.”¹² 12 CFR § 1002, comment 9(b)(2)-3.

Such an approach preferences the conveniences of a lender ahead of the needs of a consumer or small business loan applicant.

In Appendix C to Part 1002, the Bureau provides its Sample Notification Forms¹³ (SNFs) to help lenders comply with the notification requirements of 12 CFR § 1002.9(a)(2)(i). Regulation B asks creditors to disclose up to four specific reasons for an adverse decision. 12 C.F.R. §1002.9(b)(2) As if to underscore the unanimity in model construction and the close ties between Regulation B and existing industry practices, “*Part 1 Principal Reason(s) for Credit Denial, Termination, or Other Action Taken Concerning Credit*” provides lenders with a list of 23 reasons, plus one line for “other,” to explain all possible adverse decisions. The explanations are already indicated, the rationale for each already accepted, and the scope of model reasoning apparently pre-determined. These strictures conflict with the complexity and non-linearity of algorithmic lending. While the Bureau notes that Regulation B does mandate the use of any particular reasons,¹⁴ lenders are likely to use the SNFs rather than take the risk of using one that has not been approved ahead of time.

The utility of those SNFs stems at least in part from their close alignment to the structure of the linear regression models used by the major credit scoring services. A similar fit does not exist for algorithmic underwriting. The SNFs are designed to work with relationship lending and systems that use linear regression. However, they are not prepared to explain algorithmic underwriting, where data sets may contain hundreds or thousands of variables.

The Bureau should build on the momentum from the 2020 Tech Sprint¹⁵ to improve the structure of adverse action notice for decisions made with algorithmic underwriting. The prospect that the Bureau could give its blessing to a message that may leave the applicant without an understanding of the basis for a lender’s decision must be prevented.

¹² Consumer Financial Protection Bureau. “Fair Lending Report of the Bureau of Consumer Financial Protection.” Washington, DC: Consumer Financial Protection Bureau, April 2020. https://files.consumerfinance.gov/f/documents/cfpb_2019-fair-lending_report.pdf.

¹³ Consumer Financial Protection Bureau. “Appendix C to Part 1002 — Sample Notification Forms.” Interactive Bureau Regulations 12 CFR Part 1002 (Regulation B), n.d. <https://www.consumerfinance.gov/rules-policy/regulations/1002/c/>.

¹⁴ Consumer Financial Protection Bureau. “Fair Lending Report of the Bureau of Consumer Financial Protection.” Washington, DC: Consumer Financial Protection Bureau, April 2020. https://files.consumerfinance.gov/f/documents/cfpb_2019-fair-lending_report.pdf.

¹⁵ Consumer Financial Protection Bureau. “Tech Sprint on Electronic Disclosures of Adverse Action Notices.” Innovation at the Bureau, October 5, 2020. <https://www.consumerfinance.gov/rules-policy/innovation/cfpb-tech-sprints/electronic-disclosures-tech-sprint/>.

The Bureau should state that a lack of explainability is not grounds to not provide interpretable adverse-action notices.

Relatedly, it should provide guidance to lenders on ways that they can use graphics and interactive tools to enhance model explainability.

- c. The Agencies should require lenders to only use data elements that would give turned-down applicants the agency to improve their creditworthiness.*

Explainability includes actionability - but to reach actionability after a lending decision, a lender must start with an explainable system or at least one that uses post-hoc methods to accurately explain the basis for a decision. Applicants cannot understand a “black-box” model, but they are equally disempowered by a “locked box” to which they can never enter. Locked boxes are models that evaluate applicants in ways that will leave some permanently excluded from credit.

The Agencies should insist that lenders never use models constructed from data points that are unchangeable unless the creditor can show that the use of the variable serves a “substantial, legitimate, and non-discriminatory interest.”¹⁶ For example, if a notice reveals that the main determinant of an adverse decision was the average SAT score of the applicant’s undergraduate institution,¹⁷ then the applicant is left without a means to improve his chance to qualify for credit going forward. To affect the goal of creating actionable notices, a lender must exclude from its choice of data points that those that do not leave a borrower with the agency to improve on his or her creditworthiness. This example highlights a broader principle in explainability: it is much easier to realize the goal of explainability if models are initially built with that criterion in mind.¹⁸ Creating post-hoc XAI, using interpretability tools or counterfactual explanations, is inherently more complicated, fundamentally less transparent, and prone to more risk.

- d. The Agencies should shorten the time between the decision and the delivery of adverse action notices.*

Currently, lenders must notify consumers and business within 30 days after it has obtained all of the information it has considered to make a credit decision. 12 C.F.R. §1002.9(a)(1) The Bureau should revisit the length of time given to lenders to complete notification. The value of the information in an adverse action notice may degrade over time. The Bureau should provide guidance on how lenders could expedite the response, including if a lender could use a digital message in addition to the currently required written or oral statements to explain an adverse decision. 12 C.F.R. §1002.9(a)(2)

- e. The Agencies should staff and invest in the ability to test explainability, including tests for explainability among different demographic groups.*

While the level of diversity staffing in the federal government is somewhat consistent with the overall population, enormous discrepancies exist in the makeup of senior staff. In 2020, people of color held 46

¹⁶ Department of Housing and Urban Development. “Implementation of the Fair Housing Act’s Discriminatory Effects Standard 24 CFR 100.” Federal Register, February 15, 2013. Pg. 11,460
<https://www.hud.gov/sites/documents/DISCRIMINATORYEFFECTRULE.PDF>.

¹⁷ Katherine Welbeck and Ben Kaufman. “Domino: A Blog about Student Debt.” Student Borrower Protection Center. *Fintech Lenders’ Responses to Senate Probe Heighten Fears of Educational Redlining* (blog), July 31, 2020.
<https://protectborrowers.org/fintech-lenders-response-to-senate-probe-heightens-fears-of-educational-redlining/>.

¹⁸ Cynthia Rudin. “Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead.” *Nature Machine Intelligence* 1, no. 5 (May 2019): 206–15.

percent of entry-level positions, but only 32 percent of senior-level roles and only 22 percent of career Senior Executive Service civil servants were people of color.¹⁹

Reviewing models for explainability is one of the critical aspects of artificial intelligence governance, as black-box models are by definition opaque and therein at risk of having potentially discriminatory effects.

Each Agency should commit to hiring adequate staff to supervise artificial intelligence and machine learning systems. Given that the use of these techniques is still emerging, often only for marketing purposes and only among a subset of financial institutions, it stands to reason that Agencies may still need to invest in additional expertise. Relatedly, the Agencies should devote resources to provide staff with the tools they need to supervise financial institutions that use AI and ML.

Each Agency should include staff diversity in its long-term strategic plan for supervision and enforcement of the use of artificial intelligence.

Each Agency should also commit to hiring a demographically representative staff. Agencies should strive to have staff whose demographic composition mirrors the makeup of the US population. The Agencies should commit to having a demographically diverse makeup of senior staff.

f. The Agencies should issue a statement that models that are not explainable have heightened risks for discriminatory impacts.

The Agencies should clarify that fair lending risk is a primary element of any overall evaluation of a financial institution's safety and soundness by defining "model risk" to include the risk of discriminatory or inequitable outcomes for consumers. By adding this element to each Agency's compliance manuals, the Agencies would link the risks of black-box models to discriminatory treatments. The use of unexplainable black box models invites disparate treatment, and the Agencies should send strong signals that lenders cannot claim a lack of understanding to defend such practices.

g. The Agencies should address how consumers can resolve instances where incorrect data was used to evaluate their creditworthiness.

In a 2016 study from the Federal Trade Commission ("FTC"), errors were found to exist on credit reports for 21 percent of surveyed consumers.²⁰

The Fair Credit Reporting Act calls for consumer protections on record accuracy and identity theft.²¹

Unfortunately, the use of alternative data in algorithmic underwriting presupposes the use of more information. For example, a large participant in consumer non-bank online installment lending claims to have more than 10,000 data points for each adult in the United States.²²

¹⁹ Brandon Lardy. "A Revealing Look at Racial Diversity in the Federal Government: Diversity, Equity, and Inclusion." Partnership for Public Service, July 14, 2020. <https://ourpublicservice.org/blog/a-revealing-look-at-racial-diversity-in-the-federal-government/>.

²⁰ Federal Trade Commission. "Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues." Washington, DC, January 6, 2016. <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf>.

²¹ Federal Trade Commission. "Fair Credit Reporting Act 15 U.S.C § 1681," September 2018. https://www.ftc.gov/system/files/documents/statutes/fair-credit-reporting-act/545a_fair-credit-reporting-act-0918.pdf.

²² Elevate Credit, Inc. "Elevate: Leading the Path to Progress." Annual Report for 2018, March 2019. https://www.annualreports.com/HostedData/AnnualReportArchive/e/NYSE_ELVT_2018.pdf.

A consumer who wants to check the data that was used to determine their creditworthiness would find it challenging to know the extent of information held by an underwriter, and would face insurmountable obstacles to attain and review the information. The task would be even more daunting when lenders use “black box” models.

The Agencies should provide guidance on how lenders should meet the needs of consumers who want to review the information used by lenders to evaluate applications.

Question 4) How do financial institutions using AI manage risks related to data quality and data processing? How many, if at all, have control processes or automated data quality routines charged to address the data quality needs of AI? How does risk management for alternative data compare to that of traditional data? Are there any barriers or challenges that data quality and data processing pose for developing, adopting, and managing AI? If so, please provide details on those barriers or challenges.

- AND -

Question 5) Are there specific uses of AI for which alternative data are particularly effective?

Alternative data presents a challenge for risk management. Financial institutions seeking to expand access to credit and services have good reason to consider sources of data other than the traditional measures of creditworthiness, like established credit scores, which include their own biases.²³ Adding to the toolbox of available data can aid a financial institution in reaching out to underserved communities while at the same time diversifying its customer base and improving its long-term financial stability.

Using alternative data also presents unique risks for financial institutions. The data may demonstrate promise at one snapshot in time. However, the impacts over a more extended period may not have been measured, or the data may have been successfully implemented in another country with a more homogenous population but without needed safety and soundness oversight. Thus, some alternative data might not be viable here or present different risks given our diverse population, segregation patterns, and regulatory systems. The alternative data may also introduce its own yet unrecognized bias into the system.²⁴

However, alternative data can be very beneficial for supporting the financial inclusion of individuals who would otherwise fall outside of coverage by national credit reporting institutions. According to a 2015 study from the CFPB, 26 million consumers were credit invisible in 2010, and another 19 million either lacked enough information to be fully scored or did not have a recent credit history.²⁵

a) The Agencies should supervise and take enforcement action where lenders are unable to demonstrate they have effectively performed disparate impact analysis of their own use of AI and alternative data, including that provided by third-parties.

²³ J. Jatgiana and C. Lemiux. “The Role of Alternative Data and Machine Learning in Fintech: Evidence from the Lending Club Platform.” Philadelphia, Pennsylvania: Federal Reserve Bank of Philadelphia, January 2019.

²⁴ Preston Gralla. “Computerworld: Opinion” *Amazon Prime and the Racist Algorithms* (blog), May 11, 2016. <https://www.computerworld.com/article/3068622/amazon-prime-and-the-racist-algorithms.html>.

²⁵ Kenneth P. Brevoort, Philipp Grimm, and Michelle Kambara. “Data Point: Credit Invisibles.” Washington, DC: Consumer Financial Protection Bureau, May 2015. https://files.consumerfinance.gov/f/201505_cfpb_data-point-credit-invisibles.pdf.

A financial institution can address these risks if its compliance management system (CMS) is rigorous and demonstrates a commitment to diversity not only in the way it evaluates possible uses of alternative data, including disparate impact analysis but also in its oversight of vendors and data providers.

Until recently, many of the largest financial institutions advocated for only limited accountability for their use of AI and alternative data, particularly when lenders used models and data purchased from third parties. The debate over HUD's 2020 Disparate Impact Rule revealed a growing rift between those financial institutions that embraced the importance of disparate impact analysis of all models and data used, including information provided by a vendor, and those that did not. Several of the most prominent financial services trade associations successfully advocated that the HUD disparate impact rule include provisions excluding any accountability for third-party models and data.²⁶ Only recently, as racial inequality has been in the spotlight, have many of those same financial institutions and trade associations reevaluated the extent of accountability that their constituents should bear as a result of their practices.²⁷

b) The Agencies should make it clear that all financial institutions are accountable for ensuring they have a strong CMS that includes rigorous evaluations of their use of AI and alternative data.

The Agencies can aid the financial industry in promoting accountability by providing guidance on what is practically significant disparate impact, what data, including demographic data, can be used in self-testing, providing examples of what constitutes a substantial legitimate business justification for disparate impact and how and when financial institutions should implement alternatives analyses.

Some financial institutions have already embraced a disparate impact framework for the use of AI and alternative data and asked for regulatory guidance to help successfully implement that framework. The Innovation Council's Disparate Impact Statement (DI Statement) unequivocally affirms the signatories' commitment to using a disparate impact framework to review their use of AI, models, and alternative data. (See Appendix) The DI Statement also asks the Bureau to provide additional guidance on appropriate disparate impact tolerances, self-testing methods, and use of alternatives testing, as well as clarifying that financial industry compliance with the Equal Credit Opportunity Act (ECOA) requires as part of the disparate impact framework that a financial institution demonstrates a substantial legitimate business interest.²⁸ The DI Statement is entirely consistent with not only the CFPB's Responsible Business Conduct Bulletin,²⁹ but also the Office of Comptroller of the Currency's third-party oversight

²⁶ "Comments of the American Bankers Association, Consumer Bankers Association, and Housing Policy Council in Support of Proposed Amendments to the Fair Housing Act's Disparate Impact Standard to Reflect United States Supreme Court Precedent," October 18, 2019. <https://www.aba.com/-/media/documents/comment-letter/joint-hud-disparate-impact-101819.pdf>

²⁷ Emily Flitter. "Big Banks' 'Revolutionary Request: Please Don't Weaken This Rule.'" New York Times. July 16, 2020. <https://www.nytimes.com/2020/07/16/business/banks-housing-racial-discrimination.html>

²⁸ National Community Reinvestment Coalition, Lending Club, Affirm, Varo Bank, Oportun, PayPal, and Square. "NCRC, Fintechs Call On CFPB To Clarify Applying Fair Lending Rules To Artificial Intelligence." National Community Reinvestment Coalition, June 29, 2021. <https://www.ncrc.org/ncrc-fintechs-call-on-cfpb-to-clarify-applying-fair-lending-rules-to-artificial-intelligence/>.

²⁹ Consumer Financial Protection Bureau. "Responsible Business Conduct: Self-Policing, Self-Reporting, Remediation, and Cooperation," June 25, 2013. https://files.consumerfinance.gov/f/201306_cfpb_bulletin_responsible-conduct.pdf.

requirements,³⁰ but goes a step further to ask for specific guidance from the regulatory community, particularly the CFPB, to provide more granularity in methods for maintaining an effective disparate impact framework.

Question 11) What techniques are available to facilitate or evaluate the compliance of AI-based credit determination approaches with fair lending laws or mitigate risks of noncompliance? Please explain these techniques and their objectives, limitations of those techniques, and how those techniques relate to fair lending legal requirements.

- and -

Question 12) What are the risks that AI can be biased and/or result in discrimination on prohibited bases? Are there effective ways to reduce the risk of discrimination, whether during development, validation, revision, and/or use? What are some of the barriers to or limitations of those methods?

The Agencies must apply their supervisory and enforcement powers to require lenders to refine their models and hold them accountable for the disparate impacts posed by models.

The Agencies should:

a) Build and maintain an independent testing data set, to be housed at the Federal Financial Institutions Examination Council, that can be a resource for reviewing models and potentially for identifying errors in training data sets used by lenders.

A testing data set should be expansive, updated regularly, and sample a demographically representative of the US population. By making the testing data set extensive and inclusive of a diverse sample of applicants, the risk of overfitting will be minimized. The testing data set should not underrepresent or overrepresent the share of protected class applicants.

The Agencies should also review training data sets used to develop algorithmic models. The Agencies should include insufficient training data as a model risk factor and as a risk factor for fair lending when a data set does not have a representative cross-section of applicants from all demographic groups.

Precedent exists to support the construction and maintenance of a training data set. The Bureau's Consumer Credit Panel (CCP) contains a large sample of de-identified credit bureau records. With the CCP, the Bureau reviews the state of play in credit markets. The database contributes to the Bureau's ability to supervise markets, may hasten the power of the Bureau to address problems in real-time, and enhances the work of its Office of Research. For example, the Bureau used the CCP for its April 2020 Special Issue Brief on the early effects of the COVID-19 pandemic on credit applications.³¹

b) Provide guidance on the disparate impact standard by updating the language to state that the creditor practice must meet a "substantial, legitimate, nondiscriminatory interest."

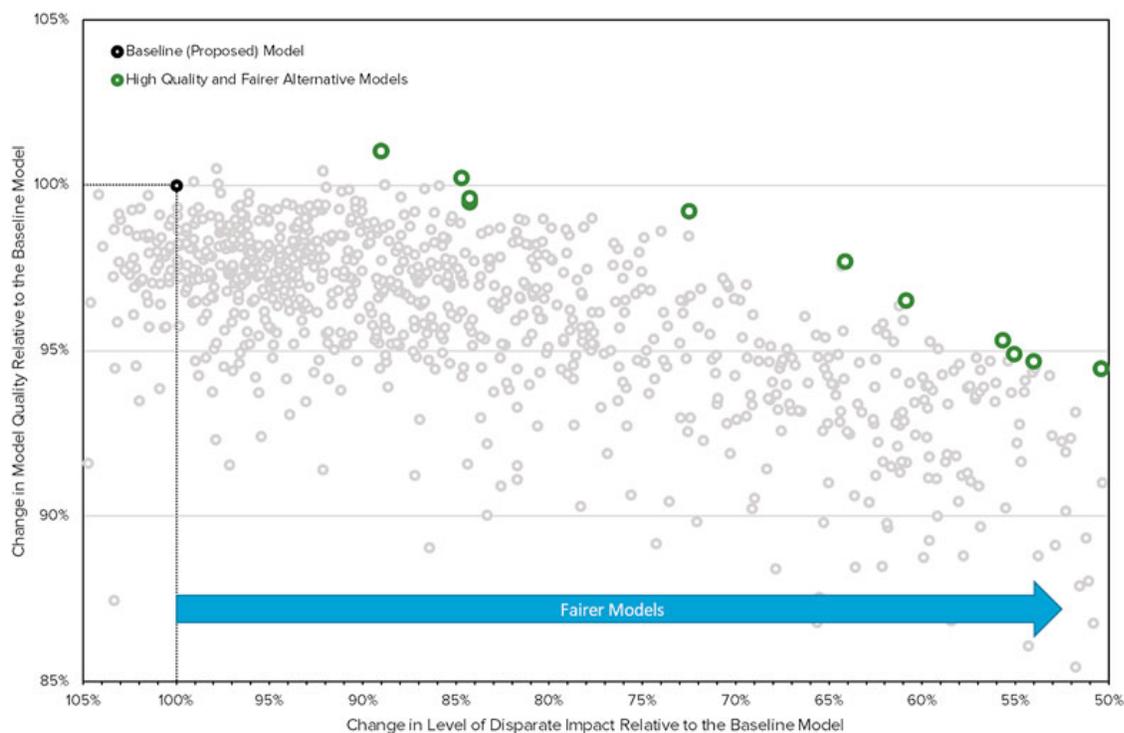
³⁰ Policy Statement on Discrimination in Lending" (59 Fed. Reg. 18266 (April 15, 1994)); OCC Bulletin 1997-24, "Credit Scoring Models: Examination Guidance"; OCC Bulletin 2011-12, "Sound Practices for Model Risk Management: Supervisory Guidance on Model Risk Management"; OCC Bulletin 2013-29, "Third-Party Relationships: Risk Management"; and OCC Bulletin 2017-43, "New, Modified, or Expanded Bank Products and Services: Risk Management Principles.

³¹ Consumer Financial Protection Bureau. "The Early Effects of the COVID-19 Pandemic on Credit Applications." Special Issue Brief. Office of Research, April 2020. https://files.consumerfinance.gov/f/documents/cfpb_issue-brief_early-effects-covid-19-credit-applications_2020-04.pdf.

The Agencies should consider how lenders can be encouraged to develop less discriminatory alternatives to biased models and clarify how those efforts can occur while still accommodating a lender’s legitimate business objectives.

Academic research demonstrates that “reducing discrimination to a reasonable extent is possible while maintaining a relatively high profit”³² and our engagement with private testing companies reveals the same perspective. Ever-increasing gains to computing power make it possible for modelers and compliance professionals to quickly search for less discriminatory models that still maintain predictive power.³³

The chart below shows how iterative improvements in modeling can lead to gains in fairness and accuracy. The y-axis shows the accuracy of a model, where a score of 100 percent is equivalent to the accuracy of the baseline model. The x-axis shows the change in the disparate impact ratio, where a score of 100% is equally disparate and scores of less than 100% have lower adverse impact ratios (AIRs).



Source: Hall, Cox, Dickerson, Kannan, Kulkarni, and Schmidt³⁴

The modeler built ten models that reduced the level of disparate impact. Of those, two increased on prior levels of model quality (predictive power of default) and only one demonstrated a reduction in model quality of more than five percentage points. One conclusion would be adopting any model that could

³² Nikita Kozodoi, Johannes Jacob, and Stefan Lessmann. “Fairness in Credit Scoring: Assessment, Implementation, and Profit Implications.” *European Journal of Operational Research* 295, no. 1 (June 2021). <https://doi.org/10.1016/j.ejor.2021.06.023>.

³³ Nicholas Schmidt and Bryce Stephens. “An Introduction to Artificial Intelligence and Solutions to the Problems of Algorithmic Discrimination.” *Algorithmic Discrimination* 73, no. 2 (2019): 130–45.

³⁴ Patrick Hall, Benjamin Cox, Steven Dickerson, Arjun Ravi Kannan, Raghu Kulkarni, and Nicholas Schmidt. “A United States Fair Lending Perspective on Machine Learning.” *Frontiers of Artificial Intelligence*, June 7, 2021. <https://doi.org/10.3389/frai.2021.695301>.

achieve an improvement in both criteria. Alternatively, regulators could lend support for improvements in fairness that resulted in only minimal concessions to model quality.

We contend that the Agencies should commit to the latter approach. Such a system would require the Agencies to indicate what level of unfairness triggers a need for a new model. A post-processing measure, such as an adverse impact ratio, would provide clarity. Subsequently, the Agencies should indicate a minimum standard for fairness and appropriate ratios for improvements in fairness relative to losses in model quality.

c) Provide guidance on how financial institutions should test their algorithmic underwriting models to mitigate against biases against protected class members.

Methods for mitigating bias in AI/ML models take one of three general forms: pre-process, where disparities in training data are addressed; in-process, where algorithms are trained to remove disparities while “learning;” and post-process, where models are altered in response to observed outcomes.

Pre-processing approaches consider the training data set used by the modeler. For the context of protecting against discriminatory outcomes, pre-processing can avoid situations where the data used to build a model is incomplete, not representative of the broader population, or if it is inconsistent with real-world contexts. If such procedures are not implemented, then a “garbage-in garbage-out” problem may occur. An AI/ML model is only as good as its data.

The agencies can implement methods to safeguard against overfitting by creating a publicly maintained data set for testing. While compliance efforts have tended to focus on outcomes,³⁵ an emphasis on pre-processing techniques has its own merits.

In-processing techniques consist of interventions made during the training period of model building. Adversarial debiasing allows a modeler to see which attributes contribute the most to unfairness and then change the model’s weightings to optimize for fairness.³⁶

A virtue of pre-processing and in-processing techniques is that lenders can conduct these efforts before introducing a model to the marketplace.

Post-processing approaches rely on means for measuring bias in completed models. “Drop-one” systems, where variables are removed iteratively to see their “marginal effect” on a fairness criterion, are an example of a post-processing approach. Others may involve re-construction of an algorithm rarely than merely the “drop” of one variable. Post-processing techniques can identify sources of discrimination and measure the impact of a change in a model to fairness, accuracy, or other criteria.

AIRs and standardized mean differences do not suggest specific changes, but they can measure disparate impact in a model. Although they do bear the virtue of simplicity, post-process approaches can lead to sub-optimal resolutions. These approaches can be used to create comparisons between “unfair” and “fair”

³⁵ Patrick Hall, Benjamin Cox, Steven Dickerson, Arjun Ravi Kannan, Raghu Kulkarni, and Nicholas Schmidt. “A United States Fair Lending Perspective on Machine Learning.” *Frontiers of Artificial Intelligence*, June 7, 2021. <https://doi.org/10.3389/frai.2021.695301>.

³⁶ Using Adversarial Debiasing to Reduce Model Bias and One Example of Bias Mitigation in In-Processing Stage. “Using Adversarial Debiasing to Reduce Model Bias: One Example of Bias Mitigation in In-Processing Stage.” *Towards Data Science* (blog), April 21, 2020. <https://towardsdatascience.com/reducing-bias-from-models-built-on-the-adult-dataset-using-adversarial-debiasing-330f2ef3a3b4>.

models, which could then become the basis for some lenders to assert that fairness is costly, or alternatively to make it satisfactory to stop at a “slightly less unfair” model as opposed to re-iterating to a fair one.

Post-processing techniques are relatively simple to implement. From a regulatory perspective, they have value because they allow an agency to make a judgment about a system without stipulating how it should be changed. Nonetheless, the best way to simplify a model is not through post-processing but instead by beginning with an explainable training data set and model.

d) State that lenders are responsible for algorithms and predictive models that have a disparate impact on members of protected classes, even if they are created or maintained by third parties. State that lenders should conduct their model risk assessments using independent sources.

Agencies should encourage lenders to conduct periodic audits of their training data sets and models.

e) Provide guidance on the collection and analysis of demographic data for self-testing privileges.

Efforts by lenders to conduct self-testing for compliance with fair lending rules confront challenges due to a lack of clarity from the Agencies on collecting demographic data. At the moment, mortgage lenders must collect and report data on aspects of demographic data under Regulation C implementing the Home Mortgage Disclosure Act, and a similar requirement is expected to be implemented for the collection of small business lending data pursuant to Section 1071 of the Dodd-Frank Wall Street Reform and Consumer Protection Act. However, Regulation B is not clear on when and for what purposes lenders can collect data when offering other types of credit products.³⁷

As a result, compliance professionals use statistical methods to approximate the demographic makeup of their customers. In 2014, the CFPB published a report on the effectiveness of the Bayesian Improved Surname Geocoding (BISG) proxy method, noting that it presented improvements over previous approaches that only used surnames or geocoding.³⁸

Nonetheless, BISG still has certain shortcomings. Most notably, while compliance professionals can use BISG to estimate the overall demographic makeup of a group of applicants, they cannot have certainty over the demographic identity of individual applicants. Moreover, given that rates of inter-racial marriage are increasing, the predictive power of the surnames in BISG may decrease. The same concerns exist for geocoding in areas with high rates of gentrification. As well, the predictive power of any BISG analysis faces constraints with smaller sample sizes. Despite these limitations, BISG is accepted by both regulators and responsible lenders as a diagnostic tool to proxy for race and ethnicity because it is reasonably accurate.

Lenders should continue to enhance their proxy methodologies through the use of self-testing. To facilitate improvements and simplify the process, lenders could benefit from considering demographic

³⁷ Consumer Financial Protection Bureau. “12 CFR § 1002.15(b)(1)(ii) and commentary - Incentives for Self-Testing and Self-Correction,” n.d. <https://www.consumerfinance.gov/rules-policy/regulations/1002/15/>.

³⁸ Consumer Financial Protection Bureau. “Using Publicly Available Information to Proxy for Unidentified Race and Ethnicity: A Methodology and Assessment.” Washington, D.C.: CFPB, Summer 2014. https://files.consumerfinance.gov/f/201409_cfpb_report_proxy-methodology.pdf.

data to assist with their internal fair lending compliance procedures. However, they often hesitate to do so for fear of exposing their organizations to additional compliance risk. We believe that lenders could benefit from new guidance that clarifies the extent to which they can use demographic data for self-testing, as long as that data is not included in marketing, underwriting, servicing, fraud protection, or other credit-related decisions.

In addition, please see part B of our response to Questions 4 and 5.

Conclusion

NCRC applauds the decision by the Agencies to address the use of artificial intelligence, machine learning, and alternative data.

NCRC believes that future supervision, rulemaking, and enforcement of AI and ML should focus on the principles of equity, transparency, and accountability.

Our comments have focused on concerns related to the explainability of models, accountability, the use of alternative data, and the need to build safeguards to prevent discriminatory practices.

NCRC urges the Agencies to ensure that the use of AI and ML develops in ways that are non-discriminatory. As the use of these technologies becomes increasingly widespread, urgency will build for regulators to establish safeguards. The Agencies must address these questions as soon as possible, as we face an inflection point where algorithmic underwriting may develop in ways that either add to or undermine access to fair, safe, and inclusive financial services.

We encourage the Agencies to provide guidance on how financial institutions should deploy artificial intelligence and machine learning. If the Agencies can give clear guidance on the issues outlined in our comment, it will clarify that well-intentioned actors need to build and market safe products. Indeed, without substantial changes, lenders will be unable to adequately explain adverse decisions or establish compliance systems that empower the kinds of frequent self-testing efforts that machine learning requires.

To reduce the chance of digital redlining, the Agencies should insist that all lenders build explainable models. Explainable models create safeguards against digital redlining.

Financial service regulators should hold all lenders accountable for testing their models for disparate impact, including those using alternative data developed by the lender or purchased from a third party. We encourage the Agencies to provide guidance on how and when lenders should implement alternatives.

Making a commitment to equity must include explicitly identifying discriminatory practices as a type of risk, by providing guidance on the disparate impact standard by updating the language to state that the creditor practice must meet a "substantial, legitimate, nondiscriminatory interest, and by emphasizing diversity in hiring.

Please contact me, Brad Blower (bblower@ncrc.org), or Adam Rust (arust@ncrc.org) if we can provide clarity on any of these issues or others.

Sincerely,

National Community Reinvestment Coalition

Affordable Homeownership Foundation, Inc.
California Reinvestment Coalition
CASA of Oregon
Clarifi
Community Service Network
Educational
Fair Finance Watch
Fair Housing Center of Central Indiana
HOME of Greater Cincinnati
Housing Justice Center
Housing on Merit
LINC UP
MakingChange, Inc
Maryland Consumer Rights Coalition
Metro North Community Development Corp.
National Association of American Veterans, Inc.
Northwest Indiana Reinvestment Alliance
Olive Hill Community Economic Development Corporation, Inc
People's Self-Help Housing
Public Good Law Center
Southern Dallas Progress Community Development Corporation
Stark County Minority Business Association
United South Broadway Corporation

APPENDIX:

Statement on Request for Guidance on Implementation of Disparate Impact Rules under ECOA

A proposal by National Community Reinvestment Coalition's Innovation Council for Financial Inclusion.



Preventing discrimination in the use of algorithms and predictive models is crucial for a fair financial system in the digital age. As a group of both consumer advocates and financial services companies, we have found a shared interest in encouraging a fair lending regulatory framework that can truly address the risk of digital discrimination, while also promoting technology and data innovation that has the potential to increase financial inclusion and lower prices for consumers. We believe the avoidance of disparate impact is the core of the solution.

We appreciate disparate impact's statistical, outcomes-based approach to identifying discrimination. By assessing outcomes, rather than inputs, disparate impact addresses discrimination that can arise when decisions are the result of algorithms or data, rather than human intent. We also believe this outcomes-based approach establishes disparate impact as a pro-innovation framework for preventing discrimination. This is because it can accommodate advances in credit modeling, artificial intelligence, machine learning, and alternative data, which have the potential to increase financial inclusion, while at the same time holding these technologies accountable for addressing potential discriminatory impact. This combination of innovation and outcomes-based accountability will produce the most fair, inclusive, consumer friendly financial services ecosystem, and allow innovation to help address the "financial services deserts" by bringing the benefits of the financial system to those who are currently underserved.

We ask that the CFPB update its guidance on disparate impact to reinforce the Equal Credit Opportunity Act (ECOA) and Regulation B, for the digital age, in a manner that is designed to encourage a fair, innovative and more inclusive financial system. We believe the need for updated guidance is even greater as a result of HUD's 2020 Disparate Impact Rule. HUD's rule, currently enjoined in federal court, undermines the disparate impact framework when used under the Fair Housing Act, and if followed could become a misguided template for regulation of disparate impact under ECOA,

To encourage innovation and financial inclusion, the CFPB should provide further guidance on disparate impact under ECOA and Regulation B by:

a) Specifically stating that the disparate impact framework applies to both traditional and technological underwriting techniques, including those that use artificial intelligence, machine learning, algorithms, and the use of alternative data.

- b) Aligning the “legitimate business need” standard to the 2013 HUD rule. We call on the CFPB to update Regulation B and its Commentary to establish that a creditor practice must meet a “substantial, legitimate, and non-discriminatory interest” that cannot reasonably be achieved through another practice that has a less discriminatory effect. This clarification is important to reduce uncertainty about what constitutes a “legitimate business need” and to minimize the risk that a lender could assert that greater profit alone is a sufficient business justification without considering the harm caused from disparate impact. We suggest the CFPB also provide examples of how lenders can demonstrate legitimate business need.
- c) Establishing statistical standards of “practical significance” to clarify when disparities would rise to the level of potentially constituting grounds for a disparate impact discrimination claim. This would increase accountability while providing clarity and ease of administration for responsible compliance programs. Such clarity could have the added benefit of giving lenders the freedom to further innovate their lending.
- d) Clarifying how lenders may deploy alternative analyses to search for less-discriminatory alternatives to a practice or data variable found to result in a disparity. Guidelines should address *when* and *how* a review for a less-discriminatory alternative is appropriate, and the extent that a guidance could include use case-specific models for defensible practices.
- e) Providing additional guidance on the use of the self-testing privilege to gather data on customer race, ethnicity, or gender, and the use of self-testing, including methods other than BISG.

This guidance would provide more regulatory certainty in loan product types beyond mortgage lending. With additional clarity from the Bureau, more lenders could implement robust self-testing and remediation protocols.

We believe that the above guidance will foster more effective monitoring of disparate impact in compliance with ECOA, while also providing greater access to affordable credit.