December 2021

An AI fair lending policy agenda for the federal financial regulators

Michael Akinwumi, John Merrill, Lisa Rice, Kareem Saleh, and Maureen Yap

This report is available online at: https://www.brookings.edu/

The Brookings Economic Studies program analyzes current and emerging economic issues facing the United States and the world, focusing on ideas to achieve broad-based economic growth, a strong labor market, sound fiscal and monetary policy, and economic opportunity and social mobility. The research aims to increase understanding of how the economy works and what can be done to make it work better.

B | Economic Studies
at BROOKINGS

DISCLOSURE

Algorithms, including artificial intelligence and machine learning models (AI/ML), increasingly dictate many core aspects of everyday life. Whether applying for a job or a loan, renting an apartment, or seeking insurance coverage, AI-powered statistical models decide who will have access to the foundational drivers of opportunity and equality.[1]

These models present both great promise and great risk. They can minimize human subjectivity and bias, facilitate more consistent outcomes, increase efficiencies, and generate more accurate decisions. Properly conceived and managed, algorithmic, and AI-based systems can be opportunity-expanding. At the same time, a variety of factors—including data limitations, lack of diversity in the technology field, and a long history of systemic inequality in America—mean that algorithmic decisions can perpetuate discrimination against historically underserved groups, such as people of color and women.

In light of the growing adoption of AI/ML, federal regulators—including the Consumer Financial Protection Bureau (CFPB), Federal Trade Commission (FTC), the Department of Housing and Urban Development (HUD), Office of the Comptroller of the Currency (OCC), Board of Governors of the Federal Reserve (Federal Reserve), Federal Deposit Insurance Corporation (FDIC), and National Credit Union Administration (NCUA)—have been evaluating how existing laws, regulations, and guidance should be updated to account for the advent of AI in consumer finance. Earlier this year some of these regulators issued a request for information on financial institutions' use of AI and machine learning in the areas of fair lending, cybersecurity, risk management, credit decisions, and other areas.[2]

The adoption of responsible AI/ML policies will continue to receive serious attention from regulators. This paper proposes policy and enforcement steps regulators can take to ensure AI/ML is harnessed to advance financial inclusion and fairness. As many other papers have already focused on methods for embracing the benefits of AI, we focus here on providing recommendations to regulators on how to identify and control for the risks in order to build an equitable market.

---

[1] Note on the language used in this article: There is no universal agreement on definitions for key terms such as "artificial intelligence," "race and ethnicity," and "fairness." We intend in all cases to be inclusive, rather than exclusive, and in no case to diminish the significance of the viewpoint of any person or to injure a person or group through our terminology. For the purposes of this response, we define "artificial intelligence" broadly to include a range of technologies and standardized practices, especially those that rely on machine learning or statistical theory. We use the following language with respect to race and ethnicity: Black, Latino, Asian American, Native Hawaiian or other Pacific Islander, American Indian/Alaska Native, and white. Instead of "fair" or "responsible" AI systems, we generally use the term "non-discriminatory" to refer to AI systems that do not disparately treat or impact people on a prohibited basis, and "equitable" to mean AI systems that promote equitable outcomes, particularly those that address historical discrimination. Finally, the term "bias" has several meanings depending on the context, so, to the extent possible, we have tried to clarify whether we mean racial bias, model bias, or other forms of bias.

[2] Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, Including Machine Learning, 86 Fed. Reg. 16837 (March 31, 2021), https://www.federalreserve.gov/newsevents/pressreleases/files/bcreg20210329a1.pdf.

I.   Background

A. AI/ML and consumer finance

For decades, lenders have used models and algorithms to make credit-related decisions, the most obvious examples being credit underwriting and pricing. Today, models are ubiquitous in consumer markets and are constantly being applied in new ways, such as marketing, customer relations, servicing, and default management. Lenders also commonly rely on models and modeled variables provided by third-party vendors.

Recent increases in computing power and exponential growth in available data have spurred the advancement of even more sophisticated statistical techniques. In particular, entities are increasingly using AI/ML, which involves exposing sophisticated algorithms to historical "training" data to discover complex correlations or relationships between variables in a dataset. The set of discovered relationships—typically referred to as a "model"—is then run against real-world information to predict future outcomes.

In the consumer finance context, AI/ML is similar to traditional forms of statistical analysis in that both are used to identify patterns in historical data to draw inferences and future behavior. What makes AI/ML unique is the ability to analyze much larger amounts of data and discover complex relationships between numerous data points that would normally go undetected by traditional statistical analysis. AI/ML tools are also capable of adapting to new information—or "learning"—without human intervention. These tools are becoming increasingly popular in both the private and public sectors. As two United States senators recently put it, "algorithms are increasingly embedded into every aspect of modern society."[3]

B. The risks posed by AI/ML in consumer finance

While AI/ML models offer benefits, they also have the potential to perpetuate, amplify, and accelerate historical patterns of discrimination. For centuries, laws and policies enacted to create land, housing, and credit opportunities were race-based, denying critical opportunities to Black, Latino, Asian, and Native American individuals. Despite our founding principles of liberty and justice for all, these policies were developed and implemented in a racially discriminatory manner. Federal laws and policies created residential segregation, the dual credit market, institutionalized redlining, and other structural barriers. Families that received opportunities through prior federal investments in housing are some of America's most economically secure citizens. For them, the nation's housing policies served as a foundation of their financial stability and the pathway to future progress. Those who did not benefit from equitable federal investments in housing continue to be excluded.

Algorithmic systems often have disproportionately negative effects on people and communities of color, particularly with respect to credit, because they reflect the dual credit market that resulted

---

[3] Letter from Senators Corey Booker and Ron Wyden to the FTC,
https://www.scribd.com/document/437955271/Booker-Wyden-FTC-Letter.

from our country's long history of discrimination.[4] This risk is heightened by the aspects of AI/ML models that make them unique: the ability to use vast amounts of data, the ability to discover complex relationships between seemingly unrelated variables, and the fact that it can be difficult or impossible to understand how these models reach conclusions. Because models are trained on historical data that reflect and detect existing discriminatory patterns or biases, their outputs will reflect and perpetuate those same problems.[5]

Examples of discriminatory models abound, particularly in the finance and housing space. In the housing context, tenant screening algorithms offered by consumer reporting agencies have had serious discriminatory effects.[6] Credit scoring systems have been found to discriminate against people of color.[7] Recent research has raised concerns about the connection between Fannie Mae and Freddie Mac's use of automated underwriting systems and the Classic FICO credit score model and the disproportionate denials of home loans for Black and Latino borrowers.[8]

These examples are not surprising because the financial industry has for centuries excluded people and communities from mainstream, affordable credit based on race and national origin.[9] There has never been a time when people of color have had full and fair access to mainstream financial services. This is in part due to the separate and unequal financial services landscape, in which mainstream creditors are concentrated in predominantly white communities and non-traditional, higher-cost lenders, such as payday lenders, check cashers, and title money lenders, are hyper-concentrated in predominantly Black and Latino communities.[10]

Communities of color have been presented with unnecessarily limited choices in lending products, and many of the products that *have* been made available to these communities have been designed to fail those borrowers, resulting in devastating defaults.[11] For example, borrowers of color with

---

[4] *See, e.g.*, Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 677-87 (2016) (discussing how data mining for models may reflect societal discrimination); Carol A. Evans, Federal Reserve, *Keeping Fintech Fair: Thinking About Fair Lending and UDAP Risks*, CONSUMER COMPLIANCE OUTLOOK (2017), https://www.frbsf.org/banking/files/Fintech-Lending-Fair-Lending-and-UDAp-Risks.pdf; FTC, BIG DATA: A TOOL FOR INCLUSION OR EXCLUSION? UNDERSTANDING THE ISSUES (2016), https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf; EXECUTIVE OFFICE OF THE PRESIDENT, BIG DATA: A REPORT ON ALGORITHMIC SYSTEMS, OPPORTUNITY, AND CIVIL RIGHTS (2016), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf; Mark MacCarthy, *Standards of Fairness for Disparate Impact Assessment of Big Data Algorithms*, 48 CUMB. L. REV. 67, 75-76 (2018).

[5] Evans, *supra* note 3; Barocas, *supra* note 3, at 674.

[6] *Conn. Fair Hous. Ctr. v. Corelogic Rental Prop. Solutions, LLC*, 478 F. Supp. 3d 259 (D. Conn. 2020) (denying motion for summary judgment to dismiss Fair Housing Act disparate impact and disparate treatment claims based on tenant screening algorithm).

[7] Sarah Ludwig, *Credit Scores in America Perpetuate Racial Injustice. Here's How*, THE GUARDIAN (Oct. 13, 2015), https://www.theguardian.com/commentisfree/2015/oct/13/your-credit-score-is-racist-heres-why.

[8] *See* Emmanuel Martinez and Lauren Kirchner, *The Secret Bias Hidden in Mortgage-Approval Algorithms*, The Markup (Aug. 25, 2021), https://themarkup.org/denied/2021/08/25/the-secret-bias-hidden-in-mortgage-approval-algorithms.

[9] *See, e.g.*, Lisa Rice & Deidre Swesnik, *Discriminatory Effects of Credit Scoring on Communities of Color*, 46 SUFFOLK U. L. REV. 935, 940 (2013).

[10] Cheryl Young & Felipe Chacon, *50 Years After the Fair Housing Act – Inequality Lingers*, TRULIA (April 19, 2018), https://www.trulia.com/research/50-years-fair-housing/.

[11] Rice, *supra* note 7, at 944.

high credit scores have been steered into subprime mortgages, even when they qualified for prime credit.[12] Models trained on this historic data will reflect and perpetuate the discriminatory steering that led to disproportionate defaults by borrowers of color.[13]

Biased feedback loops can also drive unfair outcomes by amplifying discriminatory information within the AI/ML system. For example, a consumer who lives in a segregated community that is also a credit desert might access credit from a payday lender because that is the only creditor in her community. However, even when the consumer pays off the debt on time, her positive payments will not be reported to a credit repository, and she loses out on any boost she might have received from having a history of timely payments. With a lower credit score, she will become the target of finance lenders who peddle credit offers to her.[14] When she accepts an offer from the finance lender, her credit score is further dinged because of the type of credit she accessed. Thus, living in a credit desert prompts accessing credit from one fringe lender that creates biased feedback that attracts more fringe lenders, resulting in a lowered credit score and further barriers to accessing credit in the financial mainstream.

In all these ways and more, models can have a serious discriminatory impact. As the use and sophistication of models increases, so does the risk of discrimination.

C. The applicable legal framework

In the consumer finance context, the potential for algorithms and AI to discriminate implicates two main statutes: the Equal Credit Opportunity Act (ECOA) and the Fair Housing Act. ECOA prohibits creditors from discriminating in any aspect of a credit transaction on the basis of race, color, religion, national origin, sex, marital status, age, receipt of income from any public assistance program, or because a person has exercised legal rights under the ECOA.[15] The Fair Housing Act prohibits discrimination in the sale or rental of housing, as well as mortgage discrimination, on the basis of race, color, religion, sex, handicap, familial status, or national origin.[16]

ECOA and the Fair Housing Act both ban two types of discrimination: "disparate treatment" and "disparate impact." Disparate treatment is the act of intentionally treating someone differently on a prohibited basis (e.g., because of their race, sex, religion, etc.). With models, disparate treatment can occur at the *input* or design stage, for example by incorporating a prohibited basis (such as race or sex) or a close proxy for a prohibited basis as a factor in a model. Unlike disparate treatment, disparate impact does not require intent to discriminate. Disparate impact occurs when

---

[12] *Id*. at 944-45.

[13] *Id*. at 949.

[14] *See* Carol A. Evans and Westra Miller, *From Catalogs to Clicks: The Fair Lending Implications of Target, Internet Marketing*, Federal Reserve Consumer Compliance Outlook (2019), https://consumercomplianceoutlook.org/2019/third-issue/from-catalogs-to-clicks-the-fair-lending-implications-of-targeted-internet-marketing/ (raising concerns about digital redlining that might render some advertisements invisible to certain users, disproportionately impacting users based on protected characteristics, such as race and sex).

[15] *Equal Credit Opportunity Act*, 15 U.S.C. § 1691(a), 112th Congress, 2011. https://www.govinfo.gov/app/details/USCODE-2011-title15/USCODE-2011-title15-chap41-subchapIV-sec1691.

[16] *Public Health and Welfare*, 42 U.S.C. §§ 3604-3605, 111th Congress, 2010. https://www.govinfo.gov/app/details/USCODE-2010-title42/USCODE-2010-title42-chap45-subchapI-sec3604.

a facially neutral policy has a disproportionately adverse effect on a prohibited basis, and the policy either is not necessary to advance a legitimate business interest or that interest could be achieved in a less discriminatory way.[17]

II. Recommendations for mitigating AI/ML Risks

In some respects, the U.S. federal financial regulators are behind in advancing non-discriminatory and equitable technology for financial services.[18] Moreover, the propensity of AI decision-making to automate and exacerbate historical prejudice and disadvantage, together with its imprimatur of truth and its ever-expanding use for life-altering decisions, makes discriminatory AI one of the defining civil rights issues of our time. Acting now to minimize harm from existing technologies and taking the necessary steps to ensure all AI systems generate non-discriminatory and equitable outcomes will create a stronger and more just economy.

The transition from incumbent models to AI-based systems presents an important opportunity to address what is wrong in the status quo—baked-in disparate impact and a limited view of the recourse for consumers who are harmed by current practices—and to rethink appropriate guardrails to promote a safe, fair, and inclusive financial sector. The federal financial regulators have an opportunity to rethink comprehensively how they regulate key decisions that determine who has access to financial services and on what terms. It is critically important for regulators to use all the tools at their disposal to ensure that institutions do not use AI-based systems in ways that reproduce historical discrimination and injustice.

A. Set clear expectations for best practices in fair lending testing, including a rigorous search for less discriminatory alternatives

Existing civil rights laws and policies provide a framework for financial institutions to analyze fair lending risk in AI/ML and for regulators to engage in supervisory or enforcement actions, where appropriate. However, because of the ever-expanding role of AI/ML in consumer finance and because using AI/ML and other advanced algorithms to make credit decisions is high-risk, additional guidance is needed. Regulatory guidance that is tailored to model development and testing would be an important step towards mitigating the fair lending risks posed by AI/ML.

Below we propose several measures that would mitigate those risks.

---

[17] *See, e.g.*, *Reyes v. Waples Mobile Home Park Ltd. P'ship*, 903 F.3d 415, 424 (4th Cir. 2018) (discussing the disparate impact standard under the Fair Housing Act). Because the disparate impact doctrine is concerned with the *effects* of a process rather than its *intent*, it is sometimes referred to as the "effects test." *See, e.g.,* 12 C.F.R. § 1002, Supp. I, ¶ .6(a)-2 (CFPB official commentary to Regulation B, the regulation corresponding to ECOA).

[18] For example, the European Union has leapfrogged ahead of the U.S. by releasing a proposed regulation for AI. *See* European Commission, *Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence* (also known as the "Artificial Intelligence Act") (Apr. 21, 2021), https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206.

1. **Set clear and robust regulatory expectations regarding fair lending testing to ensure AI models are non-discriminatory and equitable**

Federal financial regulators can be more effective in ensuring compliance with fair lending laws by setting clear and robust regulatory expectations regarding fair lending testing to ensure AI models are non-discriminatory and equitable. At this time, for many lenders, the model development process simply attempts to ensure fairness by (1) removing protected class characteristics and (2) removing variables that could serve as proxies for protected class membership. This type of review is only a minimum baseline for ensuring fair lending compliance, but even this review is not uniform across market players. Consumer finance now encompasses a variety of non-bank market players—such as data providers, third-party modelers, and financial technology firms (fintechs)—that lack the history of supervision and compliance management. They may be less familiar with the full scope of their fair lending obligations and may lack the controls to manage the risk. At a minimum, the federal financial regulators should ensure that all entities are excluding protected class characteristics and proxies as model inputs.[19]

Removing these variables, however, is not sufficient to eliminate discrimination and comply with fair lending laws. As explained, algorithmic decisioning systems can also drive disparate impact, which can (and does) occur even absent using protected class or proxy variables. Guidance should set the expectation that high-risk models—i.e., models that can have a significant impact on the consumer, such as models associated with credit decisions—will be evaluated and tested for disparate impact on a prohibited basis at each stage of the model development cycle.

Despite the need for greater certainty, regulators have not clarified and updated fair lending examination procedures and testing methodologies for several years. As a result, many financial institutions using AI/ML models are uncertain about what methodologies they should use to assess their models and what metrics their models are expected to follow. Regulators can ensure more consistent compliance by explaining the metrics and methodologies they will use for evaluating an AI/ML model's compliance with fair lending laws.

2. **Clarify that the federal financial regulators will conduct a rigorous search for less discriminatory alternatives as part of fair lending examinations, and set expectations that lenders should do the same**

The touchstone of disparate impact law has always been that an entity must adopt an available, less discriminatory alternative (LDA) to a practice that has discriminatory effect, so long as the alternative can satisfy the entity's legitimate needs. Consistent with this central requirement, responsible financial institutions routinely search for and adopt LDAs when fair lending testing reveals a disparate impact on a prohibited basis. But not all do. In the absence of a robust fair lending compliance framework, the institutions that fail to search for and adopt LDAs will unnecessarily perpetuate discrimination and structural inequality. Private enforcement against these institutions is difficult because outside parties lack the resources and/or transparency to police all models across all lenders.

---

[19] In doing so, though, regulators should balance prohibiting these illegal uses with encouraging model training improvements, as described in Section II.B below.

Given private enforcement challenges, consistent and widespread adoption of LDAs can only happen if the federal financial regulators conduct a rigorous search for LDAs and expect the lenders to do the same as part of a robust compliance management system. Accordingly, regulators should take the following steps to ensure that all financial institutions are complying with this central tenet of disparate impact law:

a. Inform financial institutions that regulators will conduct a rigorous search for LDAs during fair lending examinations so that lenders also feel compelled to search for LDAs to mitigate their legal risk. Also inform financial institutions how regulators will search for LDAs, so that lenders can mirror this process in their own self-assessments.

b. Inform financial institutions that they are expected to conduct a rigorous LDA search as part of a robust compliance management system, and to advance the policy goals of furthering financial inclusion and racial equity.

c. Remind lenders that self-identification and prompt corrective action will receive favorable consideration under the Uniform Interagency Consumer Compliance Rating System[20] and the CFPB's Bulletin on Responsible Business Conduct.[21] This would send a signal that self-identifying and correcting likely fair lending violations will be viewed favorably during supervisory and enforcement matters.

The utility of disparate impact and the LDA requirement as a tool for ensuring equal access to credit lies not only in enforcement against existing or past violations but in shaping the ongoing processes by which lenders create and maintain the policies and models they use for credit underwriting and pricing. Taking the foregoing steps would help ensure that innovation increases access to credit without unlawful discrimination.

3. Broaden Model Risk Management Guidance to incorporate fair lending risk

For years, financial regulators like the OCC and Federal Reserve have articulated Model Risk Management ("MRM") Guidance, which is principally concerned with mitigating financial safety and soundness risks that arise from issues of model design, construction, and quality.[22] The MRM Guidance does not account for or articulate principles for guarding against the risks that models cause or the perpetuation of discrimination. Broadening the MRM Guidance scope would ensure institutions are guarding against discrimination risks throughout the model development and use process. In particular, regulators should clearly define "model risk" to include the risk of

---

[20] Federal Financial Institutions Examination Council (FFIEC), *Uniform Interagency Consumer Compliance Rating System*, 81 Fed. Reg. No. 79473-02 (Nov. 14, 2016), https://www.ffiec.gov/press/PDF/FFIEC_CCR_SystemFR_Notice.pdf.

[21] CFPB, *Responsible Business Conduct: Self-Assessing, Self-Reporting, Remediating, and Cooperating*, CFPB Bulletin 2020-01 (Mar. 6, 2020), *available at* https://files.consumerfinance.gov/f/documents/cfpb_bulletin-2020-01_responsible-business-conduct.pdf.

[22] *See* OCC and Federal Reserve, *Supervisory Guidance on Model Risk Management*, SR 11-7 at 3 (Apr. 4, 2011) ("Model Risk Management Guidance"), https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf (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").

discriminatory or inequitable outcomes for consumers rather than just the risk of financial loss to a financial institution.

Effective model risk management practices would aid compliance with fair lending laws in several ways. First, model risk management practices can facilitate variable reviews by ensuring institutions understand the quality of data used and can identify potential issues, such as datasets that are over- or under-representative for certain populations. Second, model risk management practices are essential to ensuring that models, and variables used within models, meet a legitimate business purpose by establishing that models meet performance standards to achieve the goals for which they were developed. Third, model risk management practices establish a routine cadence for reviewing model performance. Fair lending reviews should, at a minimum, occur at the same periodic intervals to ensure that models remain effective and are not causing new disparities because of, for example, demographic changes in applicant and borrower populations.

To provide one example of how revising the MRM Guidance would further fair lending objectives, the MRM Guidance instructs that data and information used in a model should be representative of a bank's portfolio and market conditions.[23] As conceived of in the MRM Guidance, the risk associated with unrepresentative data is narrowly limited to issues of financial loss. It does not include the very real risk that unrepresentative data could produce discriminatory outcomes. Regulators should clarify that data should be evaluated to ensure that it is representative of protected classes. Enhancing data representativeness would mitigate the risk of demographic skews in training data being reproduced in model outcomes and causing financial exclusion of certain groups.

One way to enhance data representativeness for protected classes would be to encourage lenders to build models using data from Minority Depository Institutions (MDIs) and Community Development Financial Institutions (CDFIs), which have a history of successfully serving minority and other underserved communities; adding their data to a training dataset would make the dataset more representative. Unfortunately, many MDIs and CDFIs have struggled to report data to consumer reporting agencies in part due to minimum reporting requirements that are difficult for them to satisfy. Regulators should work with both consumer reporting agencies and institutions like MDIs and CDFIs to identify and overcome obstacles to the incorporation of this type of data in mainstream models.

4. Provide guidance on evaluating third-party scores and models

Financial institutions routinely rely on third-party credit scores and models to make major financial decisions. These scores and models often incorporate AI/ML methods. Third-party credit scores and other third-party models can drive discrimination, and there is no basis for immunizing them from fair lending laws. Accordingly, regulators should make clear that fair lending expectations and mitigation measures apply as much to third-party credit scores and models as they do to institutions' own models.

More specifically, regulators should clarify that, in connection with supervisory examinations, they may conduct rigorous searches for disparate impact and less discriminatory alternatives

---

[23] *See* Model Risk Management Guidance at 6.

related to third-party scores and models and expect the lenders to do the same as part of a robust compliance management system. The Federal Reserve Board, FDIC, and OCC recently released the "Proposed Interagency Guidance on Third-Party Relationships: Risk Management," which states: "When circumstances warrant, the agencies may use their authorities to examine the functions or operations performed by a third party on the banking organization's behalf. Such examinations may evaluate...the third party's ability to...comply with applicable laws and regulations, including those related to consumer protection (including with respect to fair lending and unfair or deceptive acts or practices) ...."[24] While this guidance is helpful, the regulators can be more effective in ensuring compliance by setting clear, specific, and robust regulatory expectations regarding fair lending testing for third-party scores and models. For example, regulators should clarify that protected class and proxy information should be removed, that credit scores and third-party models should be tested for disparate impact, and that entities are expected to conduct rigorous searches for less discriminatory alternative models as part of a robust compliance management program.[25]

5. Provide guidance clarifying the appropriate use of AI/ML during purported pre-application screens

Concerns have been raised about the failure to conduct fair lending testing on AI/ML models that are used in purported pre-application screens such as models designed to predict whether a potential customer is attempting to commit fraud. As with underwriting and pricing models, these models raise the risk of discrimination and unnecessary exclusion of applicants on a prohibited basis. Unfortunately, some lenders are using these pre-application screens to artificially limit the applicant pool that is subject to fair lending scrutiny. They do so by excluding from the testing pool those prospective borrowers who were purportedly rejected for so-called "fraud"-based or other reasons rather than credit-related reasons. In some cases, "fraud"[26] is even defined as a likelihood that the applicant will not repay the loan—for example, that an applicant may max out a credit line and be unwilling to pay back the debt. This practice can artificially distort the lender's applicant pool that is subject to fair lending testing and understate denial rates for protected class applicants.

Regulators should clarify that lenders cannot evade civil rights and consumer protection laws by classifying AI/ML models as fraud detection rather than credit models and that any model used to screen out applicants must be subject to the same fair lending monitoring as other models used in the credit process.

---

[24] Federal Reserve Board, FDIC, OCC, *Proposed Interagency Guidance on Third-Party Relationships: Risk Management*, 86 Fed. Reg. 38182, 38195 (July 19, 2021), https://www.govinfo.gov/content/pkg/FR-2021-07-19/pdf/2021-15308.pdf.

[25] For example, the Federal Housing Finance Agency ("FHFA") should issue guidance to clarify that the Credit Score Validation Rule requires a rigorous search for less discriminatory alternatives, consistent with disparate impact legal precedent. *See* FHFA, *Validation and Approval of Credit Score Models*, 84 Fed. Reg. 41886 (Aug. 16, 2019).

[26] Ken Pruett, "There is more to fraud than just identity theft", https://www.experian.com/blogs/insights/2009/08/there-is-more-to-fraud-than-just-identity-theft/

B. Provide clear guidance on the use of protected class data to improve credit outcomes

Any disparate impact analysis of credit outcomes requires awareness or estimation of protected class status. It is lawful—and often necessary—for institutions to make protected-class neutral changes to practices (including models) to decrease any outcome disparities observed during fair lending testing. For example, institutions may change decision thresholds or remove or substitute model variables to reduce observed outcome disparities.

Institutions should also actively mitigate bias and discrimination risks during model development. AI/ML researchers are exploring fairness enhancement techniques to be used during model pre-processing and in-processing, and evidence exists that these techniques could significantly improve model fairness. Some of these techniques use protected class data during model training but do not use that information while scoring real-world applications once the model is in production. This raises the question of the ways in which the awareness or use of protected class data during training is permissible under the fair lending laws. If protected class data is being used for a salutary purpose during model training—such as to improve credit outcomes for historically disadvantaged groups—there would seem to be a strong policy rationale for permitting it, but there is no regulatory guidance on this subject. Regulators should provide clear guidance to clarify the permissible use of protected class data at each stage of the model development process in order to encourage developers to seek optimal outcomes whenever possible.

C. Consider improving race and gender imputation methodologies

Fair lending analyses of AI/ML models—as with any fair lending analysis—require some awareness of applicants' protected class status. In the mortgage context, lenders are permitted to solicit this information, but ECOA and Regulation B prohibit creditors from collecting it from non-mortgage credit applicants. As a result, regulators and industry participants rely on methodologies to estimate the protected class status of non-mortgage credit applicants to test whether their policies and procedures have a disparate impact or result in disparate treatment. The CFPB, for example, uses Bayesian Improved Surname Geocoding (BISG), which is also used by some lenders and other entities.[27] BISG can be useful as part of a robust fair lending compliance management system. Using publicly available data on names and geographies, BISG can allow agencies and lenders to improve models and other policies that cause disparities in non-mortgage credit on a prohibited basis.[28]

Regulators should continue to research ways to further improve protected class status imputation methodologies using additional data sources and more advanced mathematical techniques. Estimating protected class status of non-mortgage credit applicants is only necessary because Regulation B prohibits creditors from collecting such information directly from those applicants.[29]

---

[27] CFPB, *Using Publicly Available Information to Proxy for Unidentified Race and Ethnicity: A Methodology and Assessment* (Summer 2014), https://files.consumerfinance.gov/f/201409_cfpb_report_proxy-methodology.pdf.

[28] For an alternative method used by the Federal Reserve Board, *see* Federal Reserve, *Consumer Compliance Outlook Live: Indirect Auto Lending - Fair Lending Considerations* (Aug. 6, 2013), https://docs.google.com/document/d/1zNEz7OAcgv7XG9z-0NSQmsX3ssIzlw1QfquROW0sLEQ/edit?usp=sharing.

[29] *Equal Credit Opportunity Act*, 12 C.F.R. §1002.5(b), 112th Congress, 2012. https://www.govinfo.gov/app/details/CFR-2012-title12-vol8/CFR-2012-title12-vol8-sec1002-5.

The CFPB should consider amending Regulation B to require lenders to collect protected class data as a part of all credit applications, just as they do for mortgage applications.

D. Ensure lenders provide useful adverse action notices

AI/ML explainability for individual decisions is important for generating adverse action reasons in accordance with ECOA and Regulation B.[30] Regulation B requires that creditors provide adverse action notices to credit applicants that disclose the principal reasons for denial or adverse action.[31] The disclosed reasons must relate to and accurately describe the factors the creditor considered. This requirement is motivated by consumer protection concerns regarding transparency in credit decision making and preventing unlawful discrimination. AI/ML models sometimes have a "black box" quality that makes it difficult to know why a model reached a particular conclusion. Adverse action notices that result from inexplicable AI/ML models are generally not helpful or actionable for the consumer.

Unfortunately, a CFPB blog post regarding the use of AI/ML models when providing adverse action notices seemed to emphasize the "flexibility" of the regulation rather than ensuring that AI providers and users adhere to the letter and spirit of ECOA, which was meant to ensure that consumers could understand the credit denials that impact their lives.[32] The complications raised by AI/ML models do not relieve creditors of their obligations to provide reasons that "relate to and accurately describe the factors actually considered or scored by a creditor."[33] Accordingly, the CFPB should make clear that creditors using AI/ML models must be able to generate adverse action notices that reliably produce consistent, specific reasons that consumers can understand and respond to, as appropriate. As the OCC has emphasized, addressing fair lending risks requires an effective explanation or explainability method; regardless of the model type used: "bank management should be able to explain and defend underwriting and modeling decisions."[34]

There is little current emphasis in Regulation B on ensuring these notices are consumer-friendly or useful. Creditors treat them as formalities and rarely design them to actually assist consumers. As a result, adverse action notices often fail to achieve their purpose of informing consumers why they were denied credit and how they can improve the likelihood of being approved for a similar loan in the future. This concern is exacerbated as models and data become more complicated and interactions between variables less intuitive.

---

[30] Adverse action notices are also required under the Fair Credit Reporting Act ("FCRA") for certain events including adverse actions as defined in ECOA, and when adverse actions in employment, insurance, and certain other contexts are taken on the basis of information in a consumer report. *See* 15 U.S.C. § 1681a(k). Adverse action is ….

[31] 12 C.F.R. Part 1002, Supp. I, ¶ 9(b). The term adverse action includes denial of credit requested, denial of a request to increase the amount of credit available, and termination of an account (or unfavorable change in the terms of an account) that does not affect all or substantially of a class of the creditor's accounts. 12 C.F.R. 1002.2(c)(1).

[32] Patrice Alexander Ficklin, Tom Pahl, and Paul Watkins, *Innovation Spotlight: Providing Adverse Action Notices When Using AI/ML Models*, , CFPB Blog (July 7, 2020), https://www.consumerfinance.gov/about-us/blog/innovation-spotlight-providing-adverse-action-notices-when-using-ai-ml-models/.

[33] 12 C.F.R. Part 1002, Supp. I, ¶ 9(b)(2)-2.

[34] OCC, *Semiannual Risk Perspective*, 23 (Spring 2019), https://www.occ.treas.gov/publications-and-resources/publications/semiannual-risk-perspective/files/pub-semiannual-risk-perspective-spring-2019.pdf.

The model adverse action notice contained in Regulation B illustrates how adverse action notices often fail to meaningfully assist consumers. For instance, the model notice includes vague reasons, such as "Limited Credit Experience." Although this could be an accurate statement of a denial reason, it does not guide consumer behavior. An adverse action notice that instead states, for example, *you have limited credit experience; consider using a credit-building product, such as a secured loan, or getting a co-signer*, would provide better guidance to the consumer about how to overcome the denial reason. Similarly, the model notice in Regulation B includes "number of recent inquiries on credit bureau report" as a sample denial reason. This denial reason may not be useful because it does not provide information about directionality. To ensure that adverse action notices are fulfilling their statutory purpose, the CFPB should require lenders to provide directionality associated with principal reasons and explore requiring lenders to provide notices containing counterfactuals—the changes the consumer could make that would most significantly improve their chances of receiving credit in the future.

E. Engage in robust supervision and enforcement activities

Regulators should ensure that financial institutions have appropriate compliance management systems that effectively identify and control risks related to AI/ML systems, including the risk of discriminatory or inequitable outcomes for consumers. This approach is consistent with the Uniform Interagency Consumer Compliance Rating System[35] and the Model Risk Management Guidance. The compliance management system should comprehensively cover the roles of board and senior management, policies and procedures, training, monitoring, and consumer complaint resolution. The extent and sophistication of the financial institution's compliance management system should align with the extent, sophistication, and risk associated with the financial institution's usage of the AI system, including the risk that the AI system could amplify historical patterns of discrimination in financial services.

Where a financial institution's use of AI indicates weaknesses in their compliance management system or violations of law, the regulators should use all the tools at their disposal to quickly address and prevent consumer harm, including issuing Matters Requiring Attention; entering into a non-public enforcement action, such as a Memorandum of Understanding; referring a pattern or practice of discrimination to the U.S. Department of Justice; or entering into a public enforcement action. The Agencies have already provided clear guidance (e.g., the Uniform Consumer Compliance Rating System) that financial institutions must appropriately identify, monitor, and address compliance risks, and the regulators should not hesitate to act within the scope of their authority. When possible, the regulators should explain to the public the risks that they have observed and the actions taken in order to bolster the public's trust in robust oversight and provide clear examples to guide the industry.

---

[35] FFIEC, *Uniform Interagency Consumer Compliance Rating System* at 21-22 (Nov. 7, 2016), https://www.ffiec.gov/press/PDF/FFIEC_CCR_SystemFR_Notice.pdf stating that for purposes of a financial institution's consumer compliance rating, examiners will assess the financial institution's Compliance Management System based on the board and management oversight as well as the compliance program, which includes policies and procedures, training, monitoring, and complaint resolution. *See also* CFPB Bulletin 2020-01, *Responsible Business Conduct: Self-Assessing, Self-Reporting, Remediating, and Cooperating* (Mar. 6, 2020), https://www.consumerfinance.gov/compliance/supervisory-guidance/bulletin-responsible-business-conduct/.

F. Release additional data and encourage public research

Researchers and advocacy groups have made immense strides in recent years studying discrimination and models, but these efforts are stymied by a lack of publicly available data. At present, the CFPB and the Federal Housing Finance Agency (FHFA) release some loan-level data through the National Survey of Mortgage Originations (NSMO) and Home Mortgage Disclosure Act (HMDA) databases. However, the data released into these databases is either too limited or too narrow for AI/ML techniques truly to discern how current underwriting and pricing practices could be fairer and more inclusive. For example, there are only about 30,000 records in NSMO, and HMDA does not include performance data or credit scores.

Adding more records to the NSMO database and releasing additional fields in the HMDA database (including credit score) would help researchers and advocacy groups better understand the effectiveness of various AI fairness techniques for underwriting and pricing. Regulators also should consider how to expand these databases to include more detailed data about inquiries, applications, and loan performance after origination. To address any privacy concerns, regulators could implement various measures such as only making detailed inquiry and loan-level information (including non-public HMDA data) available to trusted researchers and advocacy groups under special restrictions designed to protect consumers' privacy rights.

In addition, NSMO and HMDA both are limited to data on mortgage lending. There are no publicly available application-level datasets for other common credit products such as credit cards or auto loans. The absence of datasets for these products precludes researchers and advocacy groups from developing techniques to increase their inclusiveness, including through the use of AI. Lawmakers and regulators should therefore explore the creation of databases that contain key information on non-mortgage credit products. As with mortgages, regulators should evaluate whether inquiry, application, and loan performance data could be made publicly available for these credit products.

Finally, the regulators should encourage and support public research. This support could include funding or issuing research papers, convening conferences involving researchers, advocates, and industry stakeholders, and undertaking other efforts that would advance the state of knowledge on the intersection of AI/ML and discrimination. The regulators should prioritize research that analyzes the efficacy of specific uses of AI in financial services and the impact of AI in financial services for consumers of color and other protected groups.

G. Hire staff with AI and fair lending expertise, ensure diverse teams, and require fair lending training

AI systems are extremely complex, ever-evolving, and increasingly at the center of high-stakes decisions that can impact people and communities of color and other protected groups. The regulators should hire staff with specialized skills and backgrounds in algorithmic systems and fair lending to support rulemaking, supervision, and enforcement efforts that involve lenders who use AI/ML. The use of AI/ML will only continue to increase. Hiring staff with the right skills and experience is necessary now and for the future.

In addition, the regulators should also ensure that regulatory as well as industry staff working on AI issues reflect the diversity of the nation, including diversity based on race, national origin, and

gender. Increasing the diversity of the regulatory and industry staff engaged in AI issues will lead to better outcomes for consumers. Research has shown that diverse teams are more innovative and productive[36] and that companies with more diversity are more profitable.[37] Moreover, people with diverse backgrounds and experiences bring unique and important perspectives to understanding how data impacts different segments of the market.[38] In several instances, it has been people of color who were able to identify potentially discriminatory AI systems.[39]

Finally, the regulators should ensure that all stakeholders involved in AI/ML—including regulators, financial institutions, and tech companies—receive regular training on fair lending and racial equity principles. Trained professionals are better able to identify and recognize issues that may raise red flags. They are also better able to design AI systems that generate non-discriminatory and equitable outcomes. The more stakeholders in the field who are educated about fair lending and equity issues, the more likely that AI tools will expand opportunities for all consumers. Given the ever-evolving nature of AI, the training should be updated and provided on a periodic basis.

III. Conclusion

Although the use of AI in consumer financial services holds great promise, there are also significant risks, including the risk that AI has the potential to perpetuate, amplify, and accelerate historical patterns of discrimination. However, this risk is surmountable. We hope that the policy recommendations described above can provide a roadmap that the federal financial regulators can use to ensure that innovations in AI/ML serve to promote equitable outcomes and uplift the whole of the national financial services market.

---

[36] *See, e.g.,* John Rampton, Why You Need Diversity on Your Team, and 8 Ways to Build It, Entrepreneur (Sept. 26, 2019), https://www.entrepreneur.com/article/338663.

[37] *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).

[38] *See, e.g.,* Inioluwa Deborah Raji, et al., Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing at 39 (2020), https://dl.acm.org/doi/pdf/10.1145/3351095.3372873 (stressing the importance of "standpoint diversity" as algorithm development implicitly encodes developer assumptions of which they may not be aware). *See also* Model Risk Management Guidance at 4 (stating that "[a] guiding principle for managing model risk is 'effective challenge' of models, that is, critical analysis by objective, informed parties who can identify model limitations and assumptions and produce appropriate changes").

[39] *See, e.g.,* Steve Lohr, Facial Recognition is Accurate, If You're a White Guy, N.Y. Times (Feb. 9, 2018), https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html (explaining how Joy Buolamwini, a Black computer scientist, discovered that facial recognition worked well for her white friends but not for her).

B | Economic Studies
at BROOKINGS

The Brookings Economic Studies program analyzes current and emerging economic issues facing the United States and the world, focusing on ideas to achieve broad-based economic growth, a strong labor market, sound fiscal and monetary policy, and economic opportunity and social mobility. The research aims to increase understanding of how the economy works and what can be done to make it work better.

Questions about the research? Email communications@brookings.edu.
Be sure to include the title of this paper in your inquiry.