Institute for Youth in Policy

5214F Diamond Heights Blvd #3055
San Francisco, CA 94109
+1 (715) 469-6884
fellowship@yipinstitute.org
www.yipinstitute.org/fellowship/

Algorithmic Bias in Credit Scoring: Regulating AI for Fair Economic Outcomes

Diya Sharma

I. Executive summary

Algorithmic systems increasingly shape financial outcomes, yet their opaque, black-box nature can entrench systemic inequalities and undermine consumers' rights to equitable lending practices. This brief examines how the lack of transparency in algorithmic credit scoring limits economic opportunities and discusses how implementing a more comprehensive policy framework could address these issues.

II. Overview

Banks, financial technology firms, and various third-party lenders utilize algorithmic credit scoring systems to assess an individual's creditworthiness. These systems are built using machine learning models that are trained on extensive datasets. They often integrate unconventional data points, such as digital footprints and online purchasing behaviors, alongside traditionally-used factors like zip codes and education levels. The model assigns specific weights to these factors, determined through systematic analysis, with the objective of minimizing the prediction error in comparison to established patterns within a testing dataset. While such models are designed to enhance objectivity, their black-box algorithms fall short in identifying potential discriminatory practices. Consequently, algorithmic credit scoring may perpetuate existing inequalities by encoding biased data into predictive frameworks.

Weights are determined systematically by the model itself, with the goal being to determine weights that minimize the error between the model's prediction and the actual pattern in the testing dataset, while maximising the translatability on test datasets, which are used in practice. While consumers are promised greater objectivity in outcomes, the use of black-box algorithms cannot determine if they were subject to discriminatory scoring. As a result, algorithmic credit scoring can perpetuate historical inequalities by embedding biased data into predictive models.

A. Relevance

Credit access is key to economic security and upward mobility. However, a 2023 report by the Consumer Financial Protection Bureau finds that over 45 million Americans remain either credit invisible or unscorable under traditional systems. Algorithmic credit scoring was initially heralded as a way to increase access to financial institutions for these people. Unfortunately, a 2021 study from the Brookings Institution found that Black and Latino loan applicants were 40-80% more likely to be denied by algorithmic lending platforms compared to their white counterparts, even after accounting for the variables for income and debt-to-income ratios. As the number of people, notably ethnic minorities within the United States, applying for loans increases, financial institutions increasingly depend on algorithmic credit scoring to evaluate the financial health of potential borrowers.

III. History

A. Current Stances

Credit scoring has long been a site of racial and economic exclusion. The earliest credit scoring systems in the U.S. were developed in the mid-20th century and often relied on highly subjective judgments that disproportionately favored rich, white applicants. The introduction of the FICO score in the 1980s offered a standardized method, but still reflected patterns of structural inequality. Today, newer machine learning systems are replacing traditional scores. Unlike the FICO score system, many modern credit scoring algorithms operate as black boxes, wherein the actual model weights and code are not revealed externally. The defense for this, provided by lenders, is that algorithms need to be classified as intellectual property to incentivise innovation.

There is an increasing agreement among advocates for digital rights and regulatory bodies that transparency in algorithms is vital for achieving equitable lending practices. In 2022, the White House Office of Science and Technology Policy (OSTP) introduced the Blueprint for an AI Bill of Rights, which explicitly emphasized the need for clear explanations and transparency in important decisions such as lending. Nonetheless, at present, there are no federal regulations that mandate significant transparency within private credit scoring systems.

Certain racial groups have lower average credit scores due to historical patterns of economic inequality, such as redlining and employment discrimination. The datasets that algorithms are trained on often do not account for these inequities, which perpetuates the cycle of economic disadvantage.

IV. Policy Problem

A. Stakeholders

The challenge of algorithmic transparency in credit scoring involves a wide array of stakeholders, each with unique interests and responsibilities.

At the forefront are consumers, especially those from historically marginalized groups, who are disproportionately affected by biased outcomes that restrict their access to credit and, consequently, their economic advancement.

Financial institutions, which include banks and fintech companies, play a pivotal role in implementing these algorithms, as they must navigate the balance between operational efficiency and regulatory compliance, as well as maintaining their reputations.

Regulatory agencies, such as the Consumer Financial Protection Bureau (CFPB) and various state-level technology oversight bodies, are charged with the responsibility of ensuring that lending practices remain fair and free from discrimination. Additionally, civil rights and digital rights organizations are essential advocates for fairness and transparency in algorithmic decision-making, working to protect vulnerable populations.

Lastly, technology developers are tasked with creating and licensing the algorithms that underpin credit scoring systems. Unfortunately, the personal biases of these developers can often lead to biased elements being inadvertently incorporated into the algorithms themselves.

B. Risks of Indifference

Allowing opaque and biased algorithmic credit scoring to persist carries significant risks. It entrenches systemic discrimination in lending, which locks marginalized groups into cycles of economic disadvantage and reduces their ability to accrue wealth. Traditional credit scoring models are understandable to the typical consumer who takes the time to understand how to impact their credit score. However, unlike credit scoring models, lending platforms can input a data variable with no requirement to disclose the models that impact decisioning

Public trust in both financial institutions and emerging financial technologies would erode as consumers perceive the system to be inaccessible. The lack of transparency also severely limits consumer recourse. Individuals would not be able to challenge or even understand any adverse credit decisions without insight into the model's reasoning.

Moreover, financial institutions themselves face considerable reputational and legal risks if hidden biases are exposed after widespread harm has occurred, particularly in a political and regulatory climate increasingly focused on AI accountability.

C. Nonpartisan Reasoning

Providing underserved communities with credit access fosters entrepreneurship and boosts local economic outcomes. Transparency, in turn, is a prerequisite for any consumer-led accountability action. It is key to balancing the inherent power dynamic that exists between lenders and lendees, particularly those belonging to marginalized groups.

Finally, from an innovation standpoint, safeguards do not inherently stifle technological progress; rather, they ensure that innovation aligns with public interest, protecting consumers while preserving the competitive advantages that algorithmic models offer.

V. Tried Policy

Existing policies offer partial but insufficient solutions to the problem.

The Equal Credit Opportunity Act (ECOA) prohibits discrimination in lending, but it does not address the opacity that black-box models, the most used type of machine learning model in this scenario, present.

The Fair Credit Reporting Act (FCRA) provides consumers with the right to know what information is used in credit determinations, yet it lacks any governmental enforcement to interpret or push the burden of explaining onto the corporations that actually deploy these models. Some fintech firms have voluntarily experimented with explainable AI and transparency measures, but without public sector regulation, adoption has been inconsistent and competitive pressures often incentivize secrecy over openness.

Ultimately, the combination of personal, financial motives held by corporations and the extensive resources belonging to these corporations and their private investors is outpacing current AI regulation.

The Algorithmic Justice and Online Platform Transparency Act, introduced by Rep. Doris Matsui, addresses privacy concerns insofar as limiting model datasets to the traditional factors that accounted for lending decisions pre-AI.

VI. Policy Options

A policy approach would involve establishing clear standards and accountability mechanisms for algorithmic credit scoring systems to ensure fairness and transparency. Under such a framework, any financial institution using an algorithmic model would be required to submit a Model Impact Assessment (MIA) before deploying the system in consumer-facing decisions. The MIA would be mandated to include documentation about the model, such as the variables used in decision-making and the weightage of each variable. The contents will be stored in a database accessible by governmental enforcement agencies, so as to respect the intellectual property rights and personal motives of companies and their developers.

Consumers who are subject to adverse decisions would have the right to receive a layman's explanation of the model's decision, including the key factors that influenced the outcome, as well as a method to request a model audit conducted by a technology oversight agency or appeal. Institutions would also be required to retain logs on model behavior and make them available for independent audits. A public database could maintain summaries of approved models and historical performance across different demographic groups, notably any group classified as a protected class.

Models that rely on variables considered "high-risk attributes" would be subject to stricter scrutiny, requiring institutions to demonstrate that such models do not produce a disparate impact on protected groups. Institutions showing measurable gains in fairness metrics could be subject to fewer mandatory audits, while an annual "Fairness Leaderboard" could publicly recognize companies with the most equitable outcomes. This would incentivise improvements via reputation boosting.

Enforcement would include unannounced audits and financial penalties per instance of noncompliance with current federal and state AI regulatory law. Confirmed violations would be added to the public database.

VII. Conclusions

This policy brief aims to introduce and emphasize the issue of opaque and unfair algorithmic lending outcomes, through analysis of the past and status quo of these machine learning models, and presenting numerous policy options in order to curb the harmful effects of the models' outputs.

In conclusion, the lack of regulation on algorithmic lending is dangerous and harmful to all consumers, regardless of their status as a protected class, despite those individuals being at a disproportionately higher harm.

All Americans deserve the right to the pursuit of personal happiness, which is often facilitated by loans, used for personal purchases and opening of businesses most commonly.

The historical context of poverty in America is sensitive and incredibly imperative to be accounted for in training data used in these models. ML models have great potential to be more equitable than traditional lending practices and can end cycles of economic disadvantage if regulated and used appropriately.

Acknowledgment

The Institute for Youth in Policy wishes to acknowledge Mason Carlisle, Lilly Kurtz, Asher Cohen, Paul Kramer. and other contributors for developing and maintaining the Fellowship Program within the Institute.

References

[1] Agarwal, S., Muckley, C. B., & Neelakantan, P. (2023). Countering racial discrimination in algorithmic lending: A case for model-agnostic interpretation methods. Economics Letters, 226, 113--120. https://doi.org/10.1016/j.econlet.2023.110342

[2] Garcia, A. C. B., Garcia, M. G. P., & Rigobon, R. (2023). Algorithmic discrimination in the credit domain: What do we know about it? AI & Society, 39(4), 2059–2098. https://doi.org/10.1007/s00146-023-01676-3

[3] Sargeant, H. (2022, November 21). Economic and normative implications of algorithmic credit scoring. Oxford Business Law Blog. https://blogs.law.ox.ac.uk/oblb/blog-post/2023/01/economic-and-normative-implications-algorithmic-credit-scoring

[4] Cavazos, S. (2025). The impact of artificial intelligence on lending: A new form of redlining? Texas A&M Journal of Property Law, 11(2), 331–356. https://doi.org/10.37419/JPL.V11.I2.4

[5] Economic Policy Institute. (2011, October 18). Countrywide Financial: Discriminatory and illegal lending. https://www.epi.org/publication/bp335-boa-countrywide-discriminatory-lending/

[6] Center for Security and Emerging Technology. (2022, September 14). Algorithmic transparency requirements for lending platforms using automated decision systems. Federation of American Scientists. https://fas.org/publication/algorithmic-transparency-requirements-for-lending-platforms-using-automated-decision-systems/

[7] De-Arteaga, M., & Roman, S. (2025). Auditing large language models for bias in credit scoring. arXiv. https://arxiv.org/pdf/2502.17816

[8] Bruckner, M. A. (2018). The promise and perils of algorithmic lenders' use of big data. Chicago-Kent Law Review, 93(1), 3–46. https://scholarship.kentlaw.iit.edu/cklawreview/vol93/iss1/