Gunnar Mayer

FinTech and Big Data Perpetuate Racism Through Algorithmic Bias in Homeownership Among African Americans

Modern companies are relying on efficient, time-saving algorithms more often as they seek to grow and serve their customer base.[1] Because computers are not inherently biased one can reasonably assume that increased use of algorithms in business decisions completely eradicates racism from those decisions. However, in the mortgage lending industry, this is not the case. Primarily, machine learning algorithms are making decisions based on extensive data sets that are embedded with decades of discrimination.[2] This has generated algorithmic bias that has been shown to assign increased interest rates to African Americans thus perpetuating the discriminatory practice of redlining because they are then forced to live in "high risk" locations.[3] Assigning individuals different interest rates based solely on race is a violation of the Federal Housing Act and has a disparate impact on African Americans because it impedes them from accumulating wealth through homeownership. In this paper I will argue that the mortgage lending industry inadvertently assigns African Americans higher interest rates than equally qualified Caucasians because of racial algorithmic bias.

I will begin with a brief history of wealth accumulation through property holding, why it is important, and why this occurs. Next, I will establish that African Americans receive higher interest rates from mortgage lenders than similarly qualified Caucasians. Third, I will explain

---

[1] Viktor Mayer-Schönberger & Kenneth Cukier, "Big Data: A Revolution That Will Transform How We Live, Work, And Think," (2013) 52–61

[2] James Allen, "The Color of Algorithms: An Analysis and Proposed Research Agenda for Deterring Algorithmic Redlining," *Fordham Urban Law Journal*, no. 2: (2019) 219-221

[3] Allen, "The Color of Algorithms: An Analysis and Proposed Research Agenda for Deterring Algorithmic Redlining," 212

what an algorithm is and why discriminatory data produces algorithmic bias. Algorithmic bias refers to the repeatable "systematic and unfairly discriminate" results that benefit one group over another.[4] Furthermore, I will explain how algorithmic bias in mortgage lending is a violation of the Fair Housing Act. Then justify how this prevents African Americans from homeownership and why owning property is a vital aspect of American stability. Lastly, I will discuss possible technical solutions to algorithmic bias in mortgage lending.

Homeownership is important because a home is widely recognized as one of the most reliable assets to accumulate wealth and move into the middle class. The equity that is gained from homeownership is the main source of wealth for middle class citizens in America.[5] A study published in 2008 took a sample of 42,129 homes from the Panel Study of Income Dynamics (PSID) to compare property appreciation rates over 9 years. The PSID is the oldest household survey in the world. The median appreciation for all in the sample was around 4.4 percent, there were no correlated difference between being high-income, low-income, African American or Caucasian.[6] This supports that once an individual owns a home, they have a stable investment and begin accumulating wealth. While appreciation rates today are strong investments, they are incomparable to the wealth accumulation that occurred in suburbs after World War 2.[7]

---

[4] Batya Friedman and Helen Nissenbaum, "Bias in Computer Systems," ACM Transactions on Information Systems (TOIS), no. 3 1996: 330

[5] Richard Rothstein, *The Color of Law: a Forgotten History of How Our Government Segregated America* (New York: Liveright Publishing Corporation, a division of W. W. Norton & Company, 2018) 185.

[6] Thomas P Boehm and Alan Schlottmann, "Wealth Accumulation and Homeownership: Evidence for Low-Income Households," Cityscape 10, no. 2 (2008): 235

[7] Gregory N. Mankiw and David N. Weil, "The Baby Boom, The Baby Bust, and the Housing Market." *NBER Working Papers* (1988): 236

Wealth accumulation stemming from homeownership is due to rising property values. Property values increase with demand and the demand for housing has never been higher than between 1950 and 1970. This postwar housing bubble was fueled by veterans needing homes and creating the Baby Boom.[8] The federal government used the GI Bill to prevent African Americans from integrating into the "postwar suburbanization" suburbs that were being built all over the country.[9] To illustrate the unusually high wealth accumulation that originated from this period Richard Rothstein gives an example of two segregated postwar suburbs, Lakeview and Levittown.[10] Over the course of three generations, white property owners in Levittown realized gains of over $200,000, while African Americans forced to live in Lakeview only gained $45,000 in equity appreciation over the same time period.[11] Homeownership and wealth accumulation is a crucial part of elevating economically into the middle class. Yet, even in a new era of FinTech utilizing methods that would seem to eradicate racial algorithmic bias continues to cause a disparate impact on African Americans. After being historically discriminated against for decades, the data used to generate decisions, is intrinsically altered and the algorithmic bias continues to hinder them from having the equal opportunity to wealth accumulation through impacting qualifying decisions for homeownership financing.

---

[8] Gregory N. Mankiw and David N. Weil, "The Baby Boom, The Baby Bust, and the Housing Market." *NBER Working Papers* (1988): 236

[9] Rothstein, *The Color of Law,* 167

[10] Rothstein, *The Color of Law,* 182

[11] Rothstein, *The Color of Law,* 182

Lenders have historically discriminated against African Americans through face to face interactions, and now this practice is being represented by algorithmic bias.[12] Over the years, many studies have demonstrated human lenders have been blatantly racist when assigning minorities interest rates for mortgage loan repayment. One study found that the total amount of extra interest paid a year caused by discrimination against just Latinx and African Americans is "$765 million."[13] African Americans alone are paying "7.9 and 3.6 basis points more in interest for home-purchase and refinance mortgages, respectively."[14] These figures are drawn directly from government published data to display how implicit and explicit racial bias from human lenders effects African Americans and Latinx. FinTech lenders typically decrease this disparity by using algorithms that are not intended to have racial consequences.[15] However, the credit assessments created by these algorithms should in no way be seen as creating equal opportunity for homeownership.

FinTech lending algorithms decrease this racial disparity by 40%, bringing the basis point difference down to 5.3 for purchase mortgages and 2.0 for refinance mortgages. [16] However, any discrimination in such a large market is impermissible. Paying one basis point more has

---

[12] Robert Bartlett, Adair Morse, Richard Stanton, and Nancy Wallace, "Consumer-Lending Discrimination in the Fintech Era," Working Paper Series 25943, (2019): 1

[13] Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace, "Consumer-Lending Discrimination in the Fintech Era," 5

[14] Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace, "Consumer-Lending Discrimination in the Fintech Era,"

[15] Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace, "Consumer-Lending Discrimination in the Fintech Era," 4-5

[16] Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace, "Consumer-Lending Discrimination in the Fintech Era," 4-5

significant effects on an already largely indebted population.[17] There is an unacceptable difference in basis points between equitable Caucasians and equitable African Americans and Latinx. On an individual level this difference may appear negligible, however, an aggregate of this sample size totals "$765 million" more in interest.[18] Algorithmic bias resulting in the increase of even one basis point more, causes African Americans to pay millions of dollars in additional interest.

The previously cited data was computed by studying omitted variables that connect to race. Published data was also taken from the HMDA, ATTOM, and McDash/Equifax, then merged to form one large data set of accepted loans based on similar variables. Some of the variables used include name, date, interest rates, performance information, contract terms, the mortgage lender, and borrower information.[19] Unless otherwise stated, all of the data utilized to display the racial disparity between lending interest rates for African Americans and Caucasians is represented as the direct result of algorithmic bias.

Similar studies on the consequences of algorithmic bias also support that African Americans are heavily discriminated against in the mortgage loan process. In 2004 to 2007, when compared to other borrowers, African Americans were 105% more likely to be in a high cost mortgage.[20] This directly effects their ability to own property and forces African Americans

---

[17] Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace, "Consumer-Lending Discrimination in the Fintech Era," 4-5

[18] Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace, "Consumer-Lending Discrimination in the Fintech Era," 5

[19] Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace, "Consumer-Lending Discrimination in the Fintech Era," 10

[20] Patrick J. Bayer, Fernando Vendramel Ferreira, and Stephen L. Ross, "What Drives Racial and Ethnic Differences in High Cost Mortgages?: The Role of High Risk Lenders," NBER Working Paper Series: No. 22004 (2016): 114

to occupy low income neighborhoods. Another study conducted in 2015 used data from the Survey of Consumer Finances and found that African American borrowers paid an average of 29 basis points more than similarly qualified Caucasians. In addition, if the borrowers were young with low education, subprime borrowers, or women, the difference was even greater.[21] This consistent mortgage interest increase when lending to African Americans continues to restrict their ability to own property in today's society. This is a direct violation of African American's property and lending rights in the United States of America.

Assigning higher interest rates on mortgage loans to African Americans solely based on race is a violation of the Fair Housing Act. The Fair Housing Act popularly called the FHA[22] prohibits discrimination in the financing of real estate based on race, color, sex, religion, familial status, national origin, or disability.[23] Thus it is a violation of the FHA when the only quantifiable reason for increased interest rates is race. When comparing the mortgage interest rates of African American and Caucasians it is apparent that African Americans are being discriminated against. After controlling for all variables and eliminating outliers, numerous studies have proven that Caucasians are consistently receiving lower interest rates than identically qualified African Americans. The FHA states that race is a protected characteristic, and as such it cannot be used as a factor for differentiating interest rates in regard to the financing of property. As per the definition in the Code of Laws of the United States of America this proves that algorithmic bias is a violation of the FHA.

---

[21] Carolina K. Reid, Debbie Bocian, Wei Li, and Roberto G. Quercia. "Revisiting the Subprime Crisis: The Dual Mortgage Market and Mortgage Defaults by Race and Ethnicity," (2017): 482

[22] Fair Housing Act, 42 USC § 3601 et seq (1968)

[23] Fair Housing Act, 42 USC § 3605(a) (1968)

In order to understand how racist data produces algorithmic bias, one must first grasp how algorithms function. The types of algorithms used by mortgage lending services are commonly referred to as machine learning algorithms. Software engineers create business rules for the algorithm to initially make decisions based on the data inputs. The decision rules are in the format of "if this, then do this" and are commonly referred to as a decision tree.[24] Decision trees can be thought of similarly to the neural pathways in the human brain. Then training data is inputted into the algorithm so that it can begin to mutate and make its own rules. The algorithm learns to predict outcomes by comparing millions of factors and characteristics to previous outcomes. After the training data set is inputted the algorithm constantly updates as new customers become the new data set. Rapidly as new data is inputted, such a convoluted and confusing workspace is created that even the programmers cannot follow along.[25] In the case of mortgages, the kind of data sets used to train lending algorithms are composed of many characteristics such as income, credit score, and past defaults. After searching through the training sets, the algorithms learn to very accurately predict the default rate of prospective borrowers and assign them a corresponding interest rate based on the computed default risk.[26]

The big data being used has been influenced by a history of de jure and de factor discrimination in American property lending. In the past, African Americans were systematically disadvantaged, resulting in higher default rates than Caucasians. African Americans defaulted on more loans because of a multitude of reasons that were out of their control including; explicitly

---

[24] Talia B. Gillis and Jann L. Spiess, "Big Data and Discrimination," University of Chicago Law Review 86, no. 2 (March 2019): 470

[25] Talia B. Gillis and Jann L. Spiess, "Big Data and Discrimination," 469–478

[26] Talia B. Gillis and Jann L. Spiess, "Big Data and Discrimination," 480-482

biased lenders giving them higher interest rates even if they qualified for higher loan amounts and lower interest rates .[27] The government enforced redlining which led to African Americans residing in low-income and low property value areas. People living in low property value and low-income areas typically have low wage jobs and minuscule savings.[28] If an emergency occurs to these people, they cannot cover everything, and this leads to a higher percentage of the population defaulting on their mortgage loan, which in turn leads to higher interest rates. Through de jure segregation African Americans were destined to default on their mortgages more often and this has severely influenced the data being used in algorithms today.

Algorithmic bias is the byproduct of using racially biased big data to teach algorithms how to make decisions. "Big data is the process of aggregating massive amounts of information from various online platforms and data capturing entities for the purpose of identifying potential patterns."[29] In order for machine learning to predict future results the algorithm must learn about past factors and outcomes, and unfortunately the housing and credit industry has been historically racist and the data holds that history.[30] Machine learning does not take into account that the data might be flawed, it simply predicts outcomes. The data that is used to teach the algorithms carries the disparity of higher interest rates for African Americans. In addition, it also recognizes that African Americans have historically defaulted on their mortgages at a higher percentage, but it does not consider the hundreds of years of systematic oppression designed to

---

[27] Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace, "Consumer-Lending Discrimination in the Fintech Era," 1

[28] Rothstein, *The Color of Law*, 185

[29] Allen, "The Color of Algorithms: An Analysis and Proposed Research Agenda for Deterring Algorithmic Redlining," 226

[30] Rothstein, *The Color of Law*,101-139

create an inferior class. This bias is what leads to algorithmic bias in machine learning algorithms and it continues to impact African American's abilities to own a home.

Unlike implicit human bias, it is possible to improve algorithms until they work exactly as desired. The ideal solution would be to remove race from all calculations when assigning individuals mortgage loan rates. An interesting study conducted by James Allen reveals that simply removing race as a variable from lending algorithms increases the difference between Caucasian and African American lending basis points.[31] This demonstrates that even when race is not taken into consideration African Americans still pay higher interest rates than Caucasians. The only explanation for this phenomenon is that the data the algorithms are using to make decisions is in fact discriminatory. In order to control for this discrimination more complicated solutions must be presented.

One solution to algorithmic bias is to control the interest rates that are decided by the algorithm. This is a popular solution found in multiple studies because it is one of the more viable solutions. It does not require any altering of the previous data sets or machine learning algorithms. However, it is slightly counterintuitive because it would require race to be a factor in an algorithm that is working in tandem to the lending algorithm. In this instance, when race is considered as an input ,it would be able to compare African American interest rates to Caucasians and prevent any disparities it found. By consistently adjusting African American interest rates to similarly qualified Caucasians, any disparities that are the product of discriminatory outputs are corrected.[32]

---

[31] Talia B. Gillis and Jann L. Spiess, "Big Data and Discrimination," 465

[32] Allen, "The Color of Algorithms", 259

Another solution includes implementing transparency in the decisions being made. Currently companies use mathematical representations of data that are too difficult for some seasoned computer scientists to understand. For the government to regulate FinTech companies utilizing lending algorithms they must first understand what is actually occurring when an individual is assigned an interest rate or loan amount. When lending algorithms are created, they need to have a step by step explanation for the decision tree that was created. To increase transparency on a case by case basis, the algorithm should create an explanation of the decisions it made to reach the final assessment. This would also allow the programmers to see where possible problems could arise with the algorithm, leading them to fix the algorithmic bias without government regulation. [33]

If the software engineers do not solve the algorithmic bias, increased transparency would allow auditors, and by extension the government, to assess what changes to anti-discrimination laws need to occur. Currently, lawyers cannot prove disparate impact because they cannot understand what the machine learning algorithms are doing. For instance, around 2010, a number of disparate impact cases surrounding algorithmic bias were brought against different banks such as Countrywide in 2008 and 2011, Wells Fargo in 2012, and Sage Bank in 2015. The lawyers were forced to argue that it was the broker's ability to make the final decision concerning the mortgage interest rate that caused their African American and Hispanic clients to be assigned higher interest rates. Their argument was considered weak because lenders make "opaque human

[33] Kroll A. Oshua, et all, "Accountable Algorithms," University of Pennsylvania Law Review 165 (3): 633–670

decisions."[34] If the mortgage lending algorithms were more transparent, they would have had quantifiable data to support the case for discrimination.

Many studies have demonstrated that mortgage lending algorithms are causing a disparate impact on African American borrowers through algorithmic bias. This is evidenced by a striking amount of data that African Americans are paying a substantial amount more in interest than equally qualified Caucasians even when algorithms are used. The only observable reason for this difference in interest assignment is race, which is a violation of the Fair Housing Act. Looking to the future, we must be careful of algorithmic bias. As businesses seek to become even more efficient algorithms will have a substantial impact on society. Algorithmic bias can materialize in any machine learning algorithm where the data sets being used have been influenced by a discriminatory history, such as hiring practices or setting salaries for women.

[34] Talia B. Gillis and Jann L. Spiess, "Big Data and Discrimination," 462

Allen, James A. 2019. “The Color of Algorithms: An Analysis and Proposed Research Agenda for Deterring Algorithmic Redlining.” Fordham Urban Law Journal, no. 2: 219-270

Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace. “Consumer-Lending Discrimination in the FinTech Era,” 2019.

Bayer, Patrick J., Fernando Vendramel Ferreira, and Stephen L. Ross. 2016. What Drives Racial and Ethnic Differences in High Cost Mortgages?: The Role of High Risk Lenders. NBER Working Paper Series: No. 22004.

Friedman, Batya and Helen Nissenbaum. 1996. “Bias in Computer Systems.” ACM Transactions on Information Systems (TOIS), no. 3: 307-330.

Gregory N. Mankiw, and David N. Weil. 1988. “The Baby Boom, The Baby Bust, and the Housing Market.” NBER Working Papers. Pg 236

Oshua Kroll, A., Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson, And Harlan Yu. “Accountable Algorithms,” University of Pennsylvania Law Review 165 (3): 633–670.

Reid, Carolina K., Debbie Bocian, Wei Li, and Roberto G. Quercia. 2017. “Revisiting the Subprime Crisis: The Dual Mortgage Market and Mortgage Defaults by Race and Ethnicity.” 469- 487.

Rothstein, Richard. The Color of Law: a Forgotten History of How Our Government Segregated America. New York: Liveright Publishing Corporation, a division of W. W. Norton & Company, 2018.

Talia B. Gillis and Jann L. Spiess, “Big Data and Discrimination.” University of Chicago Law Review 86, no. 2 (March 2019): 459-487.

Mayer-Schönberger, Viktor and Kenneth Cukier. 2013 “Big Data: A Revolution That Will Transform How We Live, Work, And Think,”