

# EQUITABLE ALGORITHMS: EXAMINING WAYS TO REDUCE AI BIAS IN FINANCIAL SERVICES

---

## HEARING BEFORE THE TASK FORCE ON ARTIFICIAL INTELLIGENCE OF THE COMMITTEE ON FINANCIAL SERVICES U.S. HOUSE OF REPRESENTATIVES ONE HUNDRED SIXTEENTH CONGRESS SECOND SESSION

---

FEBRUARY 12, 2020

---

Printed for the use of the Committee on Financial Services

**Serial No. 116-87**



---

U.S. GOVERNMENT PUBLISHING OFFICE

42-821 PDF

WASHINGTON : 2021

## HOUSE COMMITTEE ON FINANCIAL SERVICES

MAXINE WATERS, California, *Chairwoman*

CAROLYN B. MALONEY, New York	PATRICK McHENRY, North Carolina,
NYDIA M. VELAZQUEZ, New York	<i>Ranking Member</i>
BRAD SHERMAN, California	ANN WAGNER, Missouri
GREGORY W. MEEKS, New York	FRANK D. LUCAS, Oklahoma
WM. LACY CLAY, Missouri	BILL POSEY, Florida
DAVID SCOTT, Georgia	BLAINE LUETKEMEYER, Missouri
AL GREEN, Texas	BILL HUIZENGA, Michigan
EMANUEL CLEAVER, Missouri	SEAN P. DUFFY, Wisconsin
ED PERLMUTTER, Colorado	STEVE STIVERS, Ohio
JIM A. HIMES, Connecticut	ANDY BARR, Kentucky
BILL FOSTER, Illinois	SCOTT TIPTON, Colorado
JOYCE BEATTY, Ohio	ROGER WILLIAMS, Texas
DENNY HECK, Washington	FRENCH HILL, Arkansas
JUAN VARGAS, California	TOM EMMER, Minnesota
JOSH GOTTHEIMER, New Jersey	LEE M. ZELDIN, New York
VICENTE GONZALEZ, Texas	BARRY LOUDERMILK, Georgia
AL LAWSON, Florida	ALEXANDER X. MOONEY, West Virginia
MICHAEL SAN NICOLAS, Guam	WARREN DAVIDSON, Ohio
RASHIDA TLAIB, Michigan	TED BUDD, North Carolina
KATIE PORTER, California	DAVID KUSTOFF, Tennessee
CINDY AXNE, Iowa	TREY HOLLINGSWORTH, Indiana
SEAN CASTEN, Illinois	ANTHONY GONZALEZ, Ohio
AYANNA PRESSLEY, Massachusetts	JOHN ROSE, Tennessee
BEN McADAMS, Utah	BRYAN STEIL, Wisconsin
ALEXANDRIA OCASIO-CORTEZ, New York	LANCE GOODEN, Texas
JENNIFER WEXTON, Virginia	DENVER RIGGLEMAN, Virginia
STEPHEN F. LYNCH, Massachusetts	WILLIAM TIMMONS, South Carolina
TULSI GABBARD, Hawaii	VAN TAYLOR, Texas
ALMA ADAMS, North Carolina	
MADELEINE DEAN, Pennsylvania	
JESUS "CHUY" GARCIA, Illinois	
SYLVIA GARCIA, Texas	
DEAN PHILLIPS, Minnesota	

CHARLA OUERTATANI, *Staff Director*

## TASK FORCE ON ARTIFICIAL INTELLIGENCE

BILL FOSTER, Illinois, *Chairman*

EMANUEL CLEAVER, Missouri  
KATIE PORTER, California  
SEAN CASTEN, Illinois  
ALMA ADAMS, North Carolina  
SYLVIA GARCIA, Texas  
DEAN PHILLIPS, Minnesota

BARRY LOUDERMILK, Georgia, *Ranking  
Member*  
TED BUDD, North Carolina  
TREY HOLLINGSWORTH, Indiana  
ANTHONY GONZALEZ, Ohio  
DENVER RIGGLEMAN, Virginia





# CONTENTS

---

	Page
Hearing held on:	
February 12, 2020 .....	1
Appendix:	
February 12, 2020 .....	33

## WITNESSES

WEDNESDAY, FEBRUARY 12, 2020

Ghani, Rayid, Distinguished Career Professor, Machine Learning Department and the Heinz College of Information Systems and Public Policy, Carnegie Mellon University .....	12
Henry-Nickie, Makada, David M. Rubenstein Fellow, Governance Studies, Race, Prosperity, and Inclusion Initiative, Brookings Institution .....	6
Kearns, Michael, Professor and National Center Chair, Department of Com- puter and Information Science, University of Pennsylvania .....	8
Thomas, Philip S., Assistant Professor and Co-Director of the Autonomous Learning Lab, College of Information and Computer Sciences, University of Massachusetts Amherst .....	4
Williams, Bari A., Attorney and Emerging Tech AI & Privacy Advisor .....	10

## APPENDIX

Prepared statements:	
Ghani, Rayid .....	34
Henry-Nickie, Makada .....	43
Kearns, Michael .....	49
Thomas, Philip S. ....	52
Williams, Bari A. ....	55

## ADDITIONAL MATERIAL SUBMITTED FOR THE RECORD

Foster, Hon. Bill:	
Written statement of BSA/The Software Alliance .....	62
Written statement of the Future of Privacy Forum .....	71
Written statement of ORCAA .....	88
Student Borrower Protection Center report entitled, “Educational Red- lining,” dated February 2020 .....	90
Response from Upstart to the Student Borrower Protection Center’s Feb- ruary 2020 report .....	120



## **EQUITABLE ALGORITHMS: EXAMINING WAYS TO REDUCE AI BIAS IN FINANCIAL SERVICES**

**Wednesday, February 12, 2020**

U.S. HOUSE OF REPRESENTATIVES,  
TASK FORCE ON ARTIFICIAL INTELLIGENCE,  
COMMITTEE ON FINANCIAL SERVICES,  
*Washington, D.C.*

The task force met, pursuant to notice, at 2:05 p.m., in room 2128, Rayburn House Office Building, Hon. Bill Foster [chairman of the task force] presiding.

Members present: Representatives Foster, Cleaver, Porter, Casten; Loudermilk, Budd, Hollingsworth, Gonzalez of Ohio, and Riggleman.

Chairman FOSTER. The Task Force on Artificial Intelligence will now come to order. It is my understanding that there is an ongoing markup in the Judiciary Committee, which is competing for Members' attention, and I suspect they will be coming in and out over the course of this hearing.

Without objection, the Chair is authorized to declare a recess of the task force at any time. Also, without objection, members of the full Financial Services Committee who are not members of this task force are authorized to participate in today's hearing, consistent with the committee's practice.

Today's hearing is entitled, "Equitable Algorithms: Examining Ways to Reduce AI Bias in Financial Services."

I will now recognize myself for 5 minutes for an opening statement. First, thank you, everyone, for joining us today for what should be a very interesting hearing of the task force.

Today, we are looking to explore what it means to design ethical algorithms that are transparent and fair. In short, how do we program fairness into our AI models and make sure that they can explain their decisions to us? This is an especially timely topic. It seems as though every week, we are hearing stories and questions about biased algorithms in the lending space, from credit cards that discriminate against women, to loans that discriminate based on where you went to school.

I think many of these issues can be a lot more complicated and nuanced than how they are portrayed in the media, but it is clear that the use of AI is hitting a nerve with a lot of folks.

For us as consumers to understand what is happening, we need to take a deeper look under the hood. First off, there are literally dozens of definitions of fairness to look at. As policymakers, we

need to be able to explicitly state what kinds of fairness we are looking for, and how you balance multiple definitions of fairness against each other. Because, while we have fair lending laws in the form of the Equal Credit Opportunity Act and the Fair Housing Act, translating these into analog laws into machine learning models is easier said than done. It is incumbent upon us to clearly state what our goals are, and to try to quantify the tradeoffs that we are willing to accept between accuracy and fairness.

Equally important to designing ethical algorithms, however, is finding ways to ensure that they are working as they are intended to work. AI models present novel issues for resource-strapped regulators that aren't necessarily present in traditional lending models. For example, AI models continuously train and learn from new data, which means that the models themselves must adapt and change.

Another challenge is in this biased data, and I am reminded at this point of the saying from the great sage, Tom Lehrer, who said that life is like a sewer, what you get out of it depends on what you put into it, and AI algorithms are very similar, where the algorithms are like sewers, and sewage in will generate sewage out. Maybe our job on this committee is to define the correct primary, secondary, and tertiary sewage treatment systems to make sure that what comes out of the algorithms is of higher quality than what goes into them.

And because AI models often train on historical data that reflects historical biases, which we hope will disappear over time, that means the models must correct for them as we wish today, and hopefully, those corrections will become less important in the future.

But as more alternative data points are added to the underwriting models, the risk that a model will use such data as a proxy for prohibited characteristics, like race or age, only increase. One potential solution that we keep hearing about is the idea that these algorithms or their outputs should be audited by expert third parties.

As an analogy, we have all subscribed to the idea that companies' financial statements should be audited by qualified accountants to ensure that they are in compliance with Generally Accepted Accounting Principles (GAAP).

Another idea is to require companies to regularly self-test and perform benchmarking analyses that are submitted to regulators for review. This recognizes that model development is an iterative process, and we need agile ways to review and respond to changing models.

These are just a few of the many good ideas that have been discussed. I am excited to have this conversation, to see how we can make AI be the best version of itself, and how to design algorithmic models that best capture the ideals of fairness and transparency that are reflected in our fair lending laws.

We want to make sure that the biases of the analog world are not repeated in the AI and machine learning world. And with that, I now recognize the ranking member of the task force, my friend from Georgia, Mr. Loudermilk, for 5 minutes.

Mr. LOUDERMILK. Thank you, Mr. Chairman. And I thank all of you for being here today as we discuss this very important subject. We are going to discuss ways to identify and reduce bias in artificial intelligence in financial services. We have talked about this issue in concept numerous times, but we have not yet gotten deep into what algorithm explainability really means. So, I appreciate the chairman holding this hearing.

Analytical models of AI and machine learning are best understood, at least to me, when they are broken into three basic models: descriptive analytics, which analyzes past data; predictive analytics, which predicts future outcomes based on past data; and prescribing analytics, where the algorithm recommends a course of action based on past data.

There is also a fourth emerging model, which I refer to as the “execution model,” which automatically takes action based on other AI systems’ outputs. I believe the execution model deserves the most attention from policymakers because it can remove the human element in decision-making.

There are a number of noteworthy recent developments in artificial intelligence that I hope we can discuss today. First, the White House Office of Science and Technology Policy recently released principles for how Federal agencies can regulate the development of AI in the private sector. The intent of the principles is to govern AI with the direction on the technical and ethical aspects without stifling innovation.

The principles recommended providing opportunities for public feedback during the rulemaking process, considering fairness and nondiscrimination regarding the decisions made by AI applications, and basing the regulatory approach on scientific data.

The U.S. Chief Technology Officer said the principles are designed to ensure public engagement, limit regulatory overreach, and promote trustworthy technology. Some private-sector organizations recommend principles for companies using AI, which include designating a lead AI ethics official, making sure the customer knows when they are interacting with AI, explaining how the AI arrived at its result, and testing AI for bias. I believe the latter two, explaining results and testing for bias, are important to ensure appropriate use of AI by private sector businesses.

A basic but central part of explainability is making sure businesses and their regulators are able to know the building blocks of what went into an algorithm when it was being constructed. In other words, coders should maintain full records of what is going into the model when it is being trained, ranked by order of importance. This is also known as “logging,” and can help isolate sources of bias.

A similar concept is present in credit scoring. Credit scores are generated by algorithms to create a number that predicts a person’s ability to repay a loan.

Importantly, it is easy to figure out what is bringing someone’s score up or down, because the factors that go into the score are transparent. Having a long credit history brings the score up, while the use of available credit brings it down. Additionally, on-time payments are weighted higher than the number of inquiries.

With that said, recordkeeping is a starting point, and certainly is not a silver bullet solution to the explainability problem, especially with more complex algorithms.

With explainability, we also need to define what fairness is. There needs to be a benchmark to compare algorithm results and evaluate the fairness of an algorithm's decisions. These kinds of paper trails can help get to the bottom of suspected bias in loan underwriting decisions. It is also important to be able to test algorithms to see if there is any bias present.

If there is suspected bias, companies can take a subset of the data based on sensitive features like gender and race to see if there is disparate impact on a particular group. Aside from testing for bias, testing can also help companies verify if the algorithm is arriving at its expected results.

It is also important for companies and regulators to verify the input data for accuracy, completeness, and appropriateness. Flawed data likely results in flawed algorithm outcomes.

I look forward to the discussion on this issue, and I yield back.

Chairman FOSTER. Thank you.

Today, we welcome the testimony of Dr. Philip Thomas, assistant professor and co-director of the Autonomous Learning Lab, College of Information and Computer Sciences at the University of Massachusetts Amherst; Dr. Makada Henry-Nickie, the David M. Rubenstein Fellow for the Governance Studies, Race, Prosperity, and Inclusion Initiative at the Brookings Institution; Dr. Michael Kearns, professor and national center chair, the Department of Computer and Information Science at the University of Pennsylvania; Ms. Bari A. Williams, attorney and emerging tech AI and privacy adviser; and Mr. Rayid Ghani, the distinguished career professor in the machine learning department at Heinz College of Information Systems and Public Policy at Carnegie Mellon University.

Witnesses are reminded that your oral testimony will be limited to 5 minutes. And without objection, your written statements will be made a part of the record.

Dr. Thomas, you are now recognized for 5 minutes to give an oral presentation of your testimony.

**STATEMENT OF PHILIP S. THOMAS, ASSISTANT PROFESSOR  
AND CO-DIRECTOR OF THE AUTONOMOUS LEARNING LAB,  
COLLEGE OF INFORMATION AND COMPUTER SCIENCES,  
UNIVERSITY OF MASSACHUSETTS AMHERST**

Mr. THOMAS. Thank you. Chairman Foster, Ranking Member Loudermilk, and members of the task force, thank you for the opportunity to testify today.

I am Philip Thomas, an assistant professor at the University of Massachusetts Amherst.

My goal as a machine learning researcher is to ensure that systems that use machine learning algorithms are safe and fair, properties that may be critical to the responsible use of AI in finance.

Towards this goal, in a recent science paper, my co-authors and I proposed a new type of machine learning algorithm which we call a Seldonian algorithm. Seldonian algorithms make it easier for the people using AI to ensure that the systems they create are safe and

fair. We have shown how Seldonian algorithms can avoid unfair behavior when applied to a variety of applications, including optimizing online tutorials to improve student performance, influencing criminal sentencing, and deciding which loan applications should be approved.

While our work with loan application data may appear most relevant to this task force, that work was in a subfield of machine learning called contextual bandits. The added complexity of the contextual bandit setting would not benefit this discussion, and so I will instead focus on an example in a more common and straightforward setting called regression.

In this example, we used entrance exam scores to predict what the GPAs of new university applicants would be if they were accepted. The GPA prediction problem resembles many problems in finance, for example, rating applications for a job or a loan. The fairness issues that I will discuss are the same across these applications.

In the GPA prediction study, we found that three standard machine learning algorithms overpredicted the GPAs of male applicants on average and underpredicted the GPAs of female applicants on average, with a total bias of around 0.3 GPA points in favor of male applicants. The Seldonian algorithm successfully limited this bias to below 0.05 GPA points, with only a small reduction in predictive accuracy.

The rapidly growing community of machine learning researchers studying issues related to fairness has produced many similar AI systems that can effectively preclude a variety of types of unfair behavior across a variety of applications. With the development of these fair algorithms, machine learning is reaching the point where it can be applied responsibly to financial applications, including influencing hiring and loan approval decisions.

I will now discuss technical issues related to ensuring the fairness of algorithms which might inform future regulations aimed at ensuring the responsible use of AI in finance.

First, there are many definitions of fairness. Consider our GPA prediction example. One definition of fairness requires the average predictions to be the same for each gender. Under this definition, a system that tends to predict a lower GPA if you are of a particular gender would be deemed unfair.

Another definition requires the average error of predictions to be the same for each gender. Under this definition, a system that tends to overpredict the GPAs of one gender and underpredict for another would be deemed unfair.

Although both of these might appear to be desirable requirements for a fair system, for this problem, it is not possible to satisfy both simultaneously. Any system, human or machine, that produces the same average prediction for each gender necessarily overpredicts more for one gender and vice versa. The machine learning community has generated more than 20 possible definitions of fairness, many of which are known to be conflicting in this way.

In any effort to regulate the use of machine learning to ensure fairness, a critical first step is to define precisely what fairness means. This may require recognizing that certain behaviors that appear to be unfair may necessarily be permissible in order to en-

able the enforcement of a conflicting and more appropriate notion of fairness.

Although the task of selecting the appropriate definition of fairness should likely fall to regulators and social scientists, machine learning researchers can inform this decision by providing guidance with regard to which definitions are possible to enforce simultaneously, what unexpected behavior might result from a particular definition of fairness, and how much or how little different definitions of fairness might impact profitability.

Regulations could also protect companies. Fintech companies that make every attempt to be fair using AI systems that satisfy a reasonable definition of fairness may still be accused of racist or sexist behavior for failing to enforce a conflicting definition of fairness. Regulation could protect these companies by providing an agreed-upon, appropriate, and satisfiable definition of what it means for their systems to be fair.

Once a definition of fairness has been selected, machine learning researchers can work on developing algorithms that will enforce the chosen definition. For example, our latest Seldonian algorithms are already compatible with an extremely broad class of fairness definitions and might be immediately applicable.

Still, there is no silver bullet algorithm for remedying bias and discrimination in AI. The creation of fair AI systems may require use-specific considerations across the entire AI pipeline, from the initial collection of data, through monitoring the final deployed system.

Several other questions must be answered for regulations to be effective and fair. For example, will fairness requirements that appear reasonable for the short term have the long-term effect of reinforcing existing social inequalities? How should fairness requirements account for the fact that changing demographics can result in a system that was fair last month not being fair today? And when unfair behavior occurs, how can regulators determine whether this is due to the improper use of machine learning? Thank you again for the opportunity to testify today. I look forward to your questions.

[The prepared statement of Dr. Thomas can be found on page 52 of the appendix.]

Chairman FOSTER. Thank you.

Dr. Henry-Nickie, you are now recognized for 5 minutes to give an oral presentation of your testimony.

**STATEMENT OF MAKADA HENRY-NICKIE, DAVID M. RUBENSTEIN FELLOW, GOVERNANCE STUDIES, RACE, PROSPERITY, AND INCLUSION INITIATIVE, BROOKINGS INSTITUTION**

Ms. HENRY-NICKIE. Chairman Foster, Ranking Member Loudermilk, and distinguished members of the task force, thank you for the opportunity to testify today. I am Makada Henry-Nickie, a fellow at the Brookings Institution, where my research covers issues of consumer financial protection.

I am pleased to share my perspective on both the opportunities and challenges of integrating AI into financial services. As this committee knows, market interest in AI is soaring. AI technologies



have permanently reshaped the financial marketplace and altered consumer preferences and expectations of banks. I want to point out a few key trends that underscore this premise.

First, layering AI onto the financial value chain is unlocking enormous opportunities for banks. Consider that J.P. Morgan, for example, just installed a contract software that takes mere seconds to review the same number of documents that previously required about 360,000 manpower hours to complete.

Second, AI is creating new surface areas for banks to cross-sell products to customers, and this means more revenue.

Finally, consumers are increasingly open to embracing AI in banking. According to Adobe Analytics, 44 percent of Gen Z and 31 percent of millennials have interacted with a chat bot. And they prefer, overwhelmingly, to interact with a chat bot as opposed to a human representative. Taken together, these trends suggest that AI is undoubtedly shaping the future of banking.

The story of AI in financial services is not all bad, and innovative fintechs have made salient contributions that make financial services more inclusive and more accessible for consumers. Micro savings apps, for example, have empowered millions of consumers to save more and to do so automatically.

Digit has used machine learning to help its clients save over \$2.5 billion; that is an average of \$2,000 annually. In credit markets, a combination of machine learning and alternative data is slowly showing some early promise. When I say, “alternative data,” I am not referring to the format of your email address. I am talking about practical, alternative factors such as rental payment and utility payment histories, among others. A 2019 FINRA study showed that these variables can reliably predict a consumer’s ability to repay.

Furthermore, the results of CFPB’s No Action Letter review also supports this idea of early promise. According to CFPB, Upstart, through its use of machine learning and alternative data, was able to increase loan approval by nearly 30 percent and lower APRs by as much as 17 percent. Crucially, the CFPB reported that the fintech’s data showed no evidence of fair lending disparities.

Meanwhile, a UC Berkeley study found that algorithmic lending substantially decreased pricing disparities and eliminated underwriting discrimination for Black and Hispanic borrowers.

Both research and market evidence show that, despite the risks, algorithmic models have potential to provide benefits to consumers. However, it is important not to overstate this promise. We have all had a front row seat to the movie. Algorithms propagate bias. This is not an attempt to exaggerate. Numerous cases from various scenes support this claim, from Amazon’s hiring algorithm shown to be biased against women, to Google’s insulting association between African Americans, like myself, and gorillas.

And the same Berkeley study I mentioned earlier found that algorithmic lenders systematically charged Black and Hispanic borrowers higher interest rates. According to the study, minorities paid 5.3 basis points more than their white peers.

In the final analysis, machine learning was not sophisticated enough to break the systematic correlation between race and credit risk. In the end, these borrowers pay an estimated ongoing \$765

million in excess interest payments, instead of saving or paying down student loan debt.

Machine bias is not inevitable, nor is it final. This bias, though, is not benign. AI has enormous consequences for racial, gender, and sexual minorities. This should not be trivialized. Technical solutions alone, though, will not reduce algorithmic bias or ameliorate its effects.

Congress should focus on strengthening the resiliency of the Federal consumer financial protection framework so that consumers are protected. Thank you for your time, and I look forward to your questions.

[The prepared statement of Dr. Henry-Nickie can be found on page 43 of the appendix.]

Chairman FOSTER. Thank you.

Dr. Kearns, you are now recognized for 5 minutes to give an oral presentation of your testimony.

**STATEMENT OF MICHAEL KEARNS, PROFESSOR AND NATIONAL CENTER CHAIR, DEPARTMENT OF COMPUTER AND INFORMATION SCIENCE, UNIVERSITY OF PENNSYLVANIA**

Mr. KEARNS. Thank you for the opportunity to testify today. My name is Michael Kearns, and I am a professor in the Computer and Information Science Department at the University of Pennsylvania. For more than 3 decades, my research has focused on machine learning and related topics. I have consulted extensively in the finance and technology sectors, including on legal and regulatory matters. I discussed the topics and these remarks at greater length in my recent book, "The Ethical Algorithm: The Science of Socially Aware Algorithm Design."

The use of machine learning for algorithmic decision-making has become ubiquitous in the finance industry and beyond. It is applied in consequential decisions for individual consumers such as lending and credit scoring, in the optimization of electronic trading algorithms at large brokerages, and in making forecasts of directional movement of volatility in markets and individual assets.

With major exchanges now being almost entirely electronic and with the speed and convenience of the consumer internet, the benefits of being able to leverage large-scale, fine-grain, historical data sets by machine learning have become apparent.

The dangers and harms of machine learning have also recently alarmed both scientists and the general public. These include violations of fairness, such as racial or gender discrimination in lending or credit decisions, and privacy, such as leaks of sensitive personal information.

It is important to realize that these harms are generally not the result of human malfeasance, such as racist or incompetent software developers. Rather, they are the unintended consequences of the very scientific principles behind machine learning.

Machine learning proceeds by fitting a statistical model to a training data set. In a consumer lending application, such a data set might contain demographic and financial information derived from past loan applicants, along with the outcomes of granted loans.

Machine learning is applied to find a model that can predict loan default probabilities and to make lending decisions accordingly. Because the usual goal or objective is exclusively the accuracy of the model, discriminatory behavior can be inadvertently introduced. For example, if the most accurate model overall has a significantly higher false rejection rate on Black applicants than on white applicants, the standard methodology of machine learning will, indeed, incorporate this bias.

Minority groups often bear the brunt of such discrimination since, by definition, they are less represented in the training data. Note that such biases routinely occur even if the training data itself is collected in an unbiased fashion, which is rarely the case.

Truly unbiased data collection requires a period of what is known as exploration in machine learning, which is rarely applied in practice because it involves, for instance, granting loans randomly, without regard for the properties of applicants.

When the training data is already biased and the basic principles of machine learning can amplify such biases or introduce new ones, we should expect discriminatory behavior of various kinds to be the norm and not the exception.

Fortunately, there is help on the horizon. There is now a large community of machine learning researchers who explicitly seek to modify the classical principles of machine learning in a way that avoids or reduces sources of discriminatory behavior. For instance, rather than simply finding the model that maximizes predictive accuracy, we could add the constraints that different—that the model must not have significantly different false rejection rates across different racial groups.

This constraint can be seen as forcing a balance between accuracy and a particular definition of algorithmic fairness. The modified methodology generally requires us to specify what groups or attributes we wish to protect and what harms do we wish to protect them from. These choices will always be specific to the context and should be made by key stakeholders.

There are some important caveats to this agenda. First of all, there are bad definitions of fairness that should be avoided. One example is forbidding the use of race in lending decisions in the hope that it will prevent racial discrimination. It doesn't, largely because there are so many other variables strongly correlated with race that machine learning can discover as proxies.

Even worse, one can show simple examples where such restrictions will, in fact, harm the very group we sought to protect. Unfortunately, to the extent the consumer finance law incorporates fairness considerations, they are usually of this flawed form that restricts model inputs. It is usually far better to explicitly constrain the model's output behavior, as in the example of equalizing false rejection rates in lending.

I note in closing, though all my remarks have focused on the potential for designing algorithms that are better-behaved, they also point the way to regulatory reform, since most notions of algorithmic fairness can be algorithmically audited. If we are concerned over false rejection rates, or disparities by race, we can systematically test models for such behaviors and measure the violations.

I believe that the consideration of such algorithmic regulatory mechanisms is both timely and necessary, and I have elaborated on this in other recent writings. Thank you.

[The prepared statement of Dr. Kearns can be found on page 49 of the appendix.]

Chairman FOSTER. Thank you.

Ms. Williams, you are now recognized for 5 minutes to give an oral presentation of your testimony.

**STATEMENT OF BARI A. WILLIAMS, ATTORNEY AND  
EMERGING TECH AI & PRIVACY ADVISOR**

Ms. WILLIAMS. Chairman Foster, thank you. Members of the task force, thank you for allowing me to be here. My name is Bari A. Williams, and I am an attorney and start-up adviser based in Oakland, California. I have a B.A. from UC Berkeley, an MBA from St. Mary's College of California, an M.A. in African-American studies from UCLA, and a J.D. from UC Hastings College of the Law.

Primarily, I work in technology transactions, and that also includes writing all of the terms of service, which are what I like to call the things that you scroll, scroll, scroll through, and then accept. I write all of the things that people typically tend not to read. I also focus on privacy and a specialization in AI, and my previous employer, All Turtles, is akin to an AI incubator, and they are concentrated not just on legal and policy, but also help with product production and inclusiveness.

So, in my work in the tech sector, I have been exposed to many different use cases for AI. And the things that you tend to see for now, and a lot of the panelists have also referred to them—criminal justice, lending, understanding predictive behavior—are also responsible for all of the ads that you tend to see, to influence consumer behavior.

So I would say that there are five main issues with AI in financial services, in particular. One, what data sets are being used? And to me, I distill that down to, who fact-checks the fact-checkers? What does it mean to use this particular data set, and why are you choosing to use it?

Two, what hypotheses are set out to be proven by using this data? Meaning, is there a narrative that is already being written and you are looking for examples in which to prove it and to bake that into your code?

Three, how inclusive is the team that is building and helping you test this product? I think that is one thing that has yet to be mentioned on the panel, is, also, how inclusive is the team that is actually creating this product? So who are you building the products with?

Four, what conclusions are drawn from the pattern recognition in the data that the AI provides? That is, who are you building the products for? And then, who is harmed and who stands to benefit?

And, five, how do we ensure bias neutrality, and are there even good reasons to ensure that there is bias neutrality because not all biases are bad?

Data sets in financial services are used to determine your home ownership, your mortgage, and your savings and student loan rates, all of the things that the prior panelists also noted.

I also cited the same study that Dr. Henry-Nickie did as well from UC Berkeley by noting that, yes, she is correct; in that 2017 study, it showed that 19 percent of Black borrowers and almost 14 percent of Latinx borrowers were turned down for a conventional loan, and additionally, the bias was not removed whether it was a face-to-face interaction or it was done using the algorithm. So, in fact, it just seems that the AI technology actually made the efficiency better, to deny people loans and to increase their interest rates.

So there are two mechanisms in which you can drive for fair outcomes. Again, you can pick your favorite definition of “fair.” I think you will see that there are many to choose from. One is to leverage statistical techniques to resample or reweigh a data sample to reduce the bias. I would give you a visual of, essentially, it is someone standing on a box. Imagine someone may be shorter, and you give them a box to stand on so that they are the same height as the person next to them. That is essentially reweighing the data.

And the second technique is a fairness regulator, which is essentially a mathematical constraint to ensure fairness in the model to existing algorithms.

So what are other emerging methods or ways that you can use AI for good? Some emerging methods—there is one in particular, that is seen with Zest AI, which is a tech company, and it has created a product called ZAML Fair, which reduces bias in credit assessment by ranking an algorithm’s credit variables by how much they lead to biased outcomes. And then, they muffle the influence of those variables to produce a better model with less biased outcomes.

So if more banks, or even consumer-facing retailers, credit reporting bureaus, used something like this, you may get a better outcome that shows better parity.

What ways can existing laws and regulations help us? It is the same as I tell my kids, and what I tell my clients: A rule is only as good as its enforcement. So, if you act as if the rule doesn’t exist, it might as well not exist.

For example, if a lending model finds that older individuals have a higher default rate on their loans, and then they decide to reduce lending to those individuals based on their age, that can constitute a claim for housing discrimination. That is where you could apply the Fair Housing Act.

Additionally, the U.S. Equal Credit Opportunity Act of 1974, if you show greater disparate impact on the basis of any protected class, you could also use that as a lever as well.

And I don’t abide by the idea that, oh, well, the model did it. There are people who are actually creating the models, and so that means that there is regulation that could be used to actually ensure that the people creating the models are inclusive and diverse as well. Thank you.

[The prepared statement of Ms. Williams can be found on page 55 of the appendix.]

Chairman FOSTER. Thank you.

And, Mr. Ghani, you are now recognized for 5 minutes to give an oral presentation of your testimony.

**STATEMENT OF RAYID GHANI, DISTINGUISHED CAREER PROFESSOR, MACHINE LEARNING DEPARTMENT AND THE HEINZ COLLEGE OF INFORMATION SYSTEMS AND PUBLIC POLICY, CARNEGIE MELLON UNIVERSITY**

Mr. GHANI. Thank you. Chairman Foster, members of the task force, thanks for giving me the opportunity, and for holding this hearing. My name is Rayid Ghani, and I am a professor in the machine learning department in the Heinz College of Information Systems and Public Policy at Carnegie Mellon University.

I have worked in the private sector, in academia, and extensively with governments and nonprofits in the U.S. and globally on developing and using machine learning and AI systems for public policy problems across health, criminal justice, education, public safety, human services, and workforce development in a fair and equitable manner.

AI has a lot of potential in helping tackle critical problems we face today, from improving the health and education of our children, to reducing recidivism, to improving police-community relations, to improving health and safety outcomes and conditions in workplaces and housing.

AI systems can help improve outcomes for everyone and result in a better and more equitable society. At the same time, any AI system affecting people's lives should be explicitly built to increase equity and not just optimize for efficiency.

An AI system designed to explicitly optimize for efficiency has the potential to leave more difficult or costly people to help behind, resulting in increased inequities. It is critical for government agencies and policymakers to ensure that AI systems are developed in a responsible, ethical, and collaborative manner with stakeholders that include, yes, developers who build these systems, and decision-makers who use these systems, but critically including the communities that are being impacted by them.

Since today's hearing is entitled, "Equitable Algorithms," I do want to mention that, contrary to a lot of thinking in this space today, simply developing AI algorithms that are equitable is not sufficient to achieve equitable outcomes. Rather, the goal should be to make entire systems and their outcomes equitable.

Since algorithms are typically not—and shouldn't be—making autonomous decisions in critical situations, we want equity across the entire decision-making process, which includes the AI algorithm but also the decisions made by humans using inputs from those algorithms and the impact of those decisions.

In some recent preliminary work we did with the Los Angeles City attorney's office, we found that we can mitigate the disparities that a potentially biased algorithm may create to potentially result in equitable criminal justice outcomes across racial groups.

Because an AI system requires us to define exactly what we want it to optimize, and which mistakes we think are costlier, financially or socially, than others, and by exactly how much, it forces us to make some of these ethical and societal values explicit. For example, in a system recommending lending decisions, we may have to specify the differential costs of different areas. Flagging somebody as unlikely to pay back a loan and being wrong about it versus predicting someone will pay and not pay back a loan and

being wrong about it, and specify those costs explicitly for—in the case of people who may be from different gender, race, income, and education backgrounds.

While that may have happened implicitly in the past, and with high levels of variation across different decision-makers, loan officers in this case, or banks, with AI-assisted decision-making processes, we are forced to define them explicitly and, ideally, consistently.

In my written testimony, I outline a series of steps to create AI systems that are likely to lead to equitable outcomes that range from coming up with the outcomes to building these systems to validating whether they achieve those outcomes, but it is important to note that these steps are not purely technical but involve understanding the existing social and decision-making processes, as well as require solutions that are collaborative in nature.

I think it is critical and urgent for policymakers to act and provide guidelines and regulations for both the public and private sector organizations, using AI-assisted decision-making processes in order to ensure that these systems are built in a transparent and accountable manner and result in fair and equitable outcomes for society.

As initial steps, we recommend, one, expanding the already existing regulatory frameworks in different policy areas to account for AI-assisted decision-making. A lot of these bodies already exist—SEC, FINRA, CFPB, FDA, FEC, you know, pick your favorite three-letter acronym. But these bodies typically regulate inputs that go into the process—race or gender may not be allowed—and sometimes the process, but rarely focus on the outcomes produced by these processes.

We recommend expanding these regulatory bodies to update their regulations to ensure they apply to AI-assisted decision-making.

We also recommend creating training programs, processes, and tools to support these regulatory agencies in their expanded responsibilities and roles. It is important to recognize that AI can have a massive positive social impact, but we need to make sure that we can put guidelines and regulations in place to maximize the chances of the positive impact, while protecting people who have been traditionally marginalized in society and may be affected negatively by these new AI systems. Thank you for this opportunity, and I look forward to your questions.

[The prepared statement of Dr. Ghani can be found on page 34 of the appendix.]

Chairman FOSTER. Thank you.

And I now recognize myself for 5 minutes for questions.

Dr. Thomas, the Equal Credit Opportunity Act, also known as ECOA, prohibits discrimination in lending based on the standard factors: race or color, religion, national origin, sex, marital status, age, and the applicant's receipt of income from any public assistance program. Today, is it technically possible to program these explicit constraints? If Congress gives exact guidance as to what we think is fair, are there still remaining technical problems? I would be interested in—yes, proceed.

Mr. THOMAS. Yes. We could program those into algorithms. For example, the Seldonian algorithms we have created, for most definitions of fairness, we could encode them now. The remaining technical challenge is just to recognize that often fairness guarantees are only with high probability, not certainty. So it may not be possible to create an algorithm that guarantees with certainty it will be fair with respect to the chosen definition of fairness, but we can create ones that will be fair with high probability, yes.

Chairman FOSTER. Any other comments on that general problem? Is it just a definitional question we are wrestling with, or are there technical issues that are—Dr. Kearns?

Mr. KEARNS. If I understood you correctly, as per my remarks, I think all of these definitions that try to get to fairness by restricting inputs to models are ill-formed. You should specify what behavior you want at the output. So, when you forbid the use of race, you forget the fact that unfortunately, in the United States, ZIP code is already a very good statistical proxy for race. So what you should just do is say, “Don’t have racial discrimination in the output behavior of this model,” and let the model use any inputs it wants.

Ms. HENRY-NICKIE. I would just add that in optimizing for one definition of fairness, sometimes we are actually creating a disparate treatment effect within the protected class group. One study showed that when they optimized for statistical parity, meaning the same outcome for both groups, no differences, they actually hurt qualified members of a protected class. And so, there is a very costly decision involved in constraining for one definition, and hurting people in the real world.

Chairman FOSTER. Mr. Ghani?

Mr. GHANI. To that point, you can always achieve some—whatever definition of fairness in terms of the outcomes you care about. The question is, at what cost? There are a lot of ways you can make fairly random decisions, and a lot of random decisions will be somewhat fair, but the cost will be, in terms of effectiveness of outcomes, you are not going to get to people who need the support, who need the help, who need the loans, who need the services.

So, the question is not whether the algorithms can achieve fairness. Yes, they can. But is the cost that comes with it acceptable to society and to the values that we care about?

Chairman FOSTER. Yes, Ms. Williams?

Ms. WILLIAMS. I would also add that this goes back to the point that I made about a narrative looking for facts. We want to be careful that, to Dr. Kearns’ point, I think solving for the outcome is actually probably most effective. The inputs are very important, yes, but also you are typically picking those inputs because there is a desired outcome that you want, and that is why you are choosing the data sets that you are choosing.

There also needs to be an element of making sure that you are examining and auditing the human behavior that is responsible for the decision-making based off of that output as well. It isn’t enough to simply look at just the model and the inputs, but it is looking at the output, choosing to solve for the desired output, and then looking at the human decision-making behind how that comes to be.



Chairman FOSTER. And the issue with black box testing, that you can look at the details of the algorithms, is that an appropriate stance for us to take in regulating this? This is something that we run into in things like regulating high frequency trading, where they are very protective of the source code for their trading, and they say: Just look at the trading tapes, and look at our behavior, and don't ask us how we come to that behavior.

Is that going to end up being sufficient here, or would the regulators have to look at the guts of the algorithm? Dr. Thomas?

Mr. THOMAS. That will depend on the chosen definition of fairness. If the definition of fairness is that you don't look at a feature like race, which is the kind that Professor Kearns is arguing against, if it was that kind of definition, you may need to look at the algorithm, because it could be looking at some other features that make it act as though it was looking at that protected attribute.

But if you are looking at a definition of fairness, like the ones Professor Kearns is promoting, things like equalized odds or demographic parity, which are requiring false positive and false negative rates to be bounded, those you could test in the black box way, looking at the behavior of the system and then determine if it is being fair or not, without looking at the code for the algorithm.

Chairman FOSTER. Yes, Dr. Kearns?

Mr. KEARNS. I think one could go a long way with black box testing. It is always better to be able to see source code. I think it is also important to remember that sometimes when we talk about algorithms or models, we are oversimplifying.

A good example is advertising results on Google. Underneath advertising results on Google is, indeed, a machine learning model that tries to predict the likelihood that you would click on an ad, and that goes into the process of placing ads. But there is also an auction being held for people's eyeballs and impressions, and these two things interact.

For instance, there have been studies showing that sometimes gender discrimination in the display of STEM advertising in Google is not due to the underlying machine learning models of Google but rather to the fact that there is a group of advertisers willing to outbid STEM advertisers for female impressions.

Chairman FOSTER. I will now have to bring the gavel down on myself for exceeding my time, and recognize the distinguished ranking member, Mr. Loudermilk, for 5 minutes for questions.

Mr. LOUDERMILK. Thank you, Mr. Chairman.

And thank you all for your incredible testimony. Spending 30 years in the information technology sector, I have learned one thing, which is, if you are going to take a scientific approach to anything, you can't use your own bias, but you have to suspect bias. Many times I have gone to dealing with cybersecurity issues, programming security on physical networks, and it doesn't work the way I thought it was supposed to work.

Several times, I went in to check myself, and found out that what I suspected was supposed to happen isn't what was happening. In other words, my own bias of, I program it for this outcome, but the machine was actually giving me the proper outcome.

The only reason I say that is, if you are going to take a scientific approach, you have to check your own bias as well. So, when I ask some questions here, don't interpret what I am trying to say. I just have—we have to understand that we all have bias. And we also have to look in—are there occasions when the output isn't what we expected, but it is the right output? And the only reason I am going down that because I want to ask some questions just to try to help us get to, where are we seeing the bias?

And I am not going into questioning—or making a statement that, yes, AI is perfect, and it is working the way it should be, that there is anything wrong with the testimonies. I think as a community, we have to come together and we realize that this is the future we are going to and we have to get things right. And so I just wanted to say that, that if I ask questions, don't take it that I am trying to question the validity of what you are telling me. I just need to dig a little deeper into some of this.

Ms. Henry-Nickie, as we are all concerned about potential bias in algorithms, we know from a scientific approach that humans have much more potential for bias than machines, if properly utilized and programmed. And I think that is what we are getting.

Ms. Williams said something in her testimony that just highlighted—I just kind of want to step through some things to see if we can really drive in to where the issue exists. In her testimony, she was talking about home mortgage disclosures, and it showed that—and I believe, if I am right, Ms. Williams, this was AI approving home mortgages, is that correct—and I think it was like only 81 percent of Blacks were approved, and 76 percent of Hispanics were approved.

So my question, Dr. Henry-Nickie, is, how do we know that those numbers weren't correct? In other words, was a 19 percent disapproval of Black borrowers and 24 percent of Latinos outside of what would normally we see if it wasn't through an algorithm?

Ms. HENRY-NICKIE. It is difficult to answer that question without looking at the algorithms, but I will tell you that it is not fair to assess what a proper outcome should be. The context matters.

Mr. LOUDERMILK. Right.

Ms. HENRY-NICKIE. And so, if the market bears an average denial rate of 19 percent, then that is the market. And if all groups—Hispanics, African Americans, and white borrowers—are being denied at systematically similar rates, then that is an outcome that I don't think we can argue with. What is troublesome or concerning in that kind of example would be a model that is systematically denying minority borrowers, and having that be based on their race or predicted by their race.

Mr. LOUDERMILK. Right.

Ms. HENRY-NICKIE. So I think it is—and we have all said it on the panel—looking squarely for computational technical solutions is part of the answer but it is not the complete answer. We need a systematic approach to making sure that we can understand what is going on in these algorithmic applications and also from there to monitor effects and most importantly processes.

Mr. LOUDERMILK. And so, when it comes to testing AI platforms, it is not just the algorithm. There is a whole lot of emphasis on the algorithm, which is a mathematical equation. That is one part of

a four-part testing that we need to do. The appropriateness of the data, the quality of the data, the availability of the data—you also have cognitive input systems that have to be considered if it is using facial identification for something. Is that actually operating?

The reason I am asking the questions is to say, are we focused on an algorithm when the problem may actually be in the data or the appropriateness of the data if there is—and we just will make the assumption for this argument—that the output of the AI system is wrong? But I also think we do have to have empirical data to prove that the output is wrong, and it is not in our own bias. And I am not suggesting that that is what it is, but from a scientific approach, we have to do that. In a forensic way, if we are going to find out where the problem is, we have to consider all of that.

If we have a second round, I will have more questions. Thank you, Mr. Chairman.

Chairman FOSTER. I anticipate that we will. The gentleman from Missouri, Mr. Cleaver, who is also the Chair of our Subcommittee on National Security, International Development and Monetary Policy, is recognized for 5 minutes.

Mr. CLEAVER. Thank you, Mr. Chairman, and thank you for holding this hearing. Dr. Thomas, what is AI? Can you, as quickly, as short a definition as you—

Mr. THOMAS. Unfortunately, it is a poor definition, but AI, I view as just a research field that contains a lot of different directions towards making machines more intelligent so that they can solve problems that we might associate with intelligent behavior.

Mr. CLEAVER. Machine intelligence?

Mr. THOMAS. Yes.

Mr. CLEAVER. Okay. So, if Netflix begins to have showings for certain viewers, customers, and they know what movies and shows that I would most likely enjoy, what determined that? How did they get that information? Is that AI?

Mr. THOMAS. Yes. Typically, that would be machine learning, which is a subfield of AI, that uses data collected from people, for example, to make decisions or predictions about what those people will like in the future.

Mr. CLEAVER. Okay. Thank you. For any of our witnesses, I was on the committee when we had the economic collapse in 2008, and witness after witness testified clearly, unambiguously, that there was great intentionality in the discrimination in mortgages with Black and Brown people. They admitted it. Can AI, Ms. Williams, eliminate that or confuse it even more?

Ms. WILLIAMS. It has the potential to do both. I'm sorry; I am giving you a very lawyerly answer, right? It depends. It literally can do both. My concern—the ranking member made a comment in regard to his question around, how do we know that this isn't the right answer based on the data that is received? Well, the answer to that, I would say, which is also analogous to your question, is if you are using historical data, the historical data already is biased.

So, if we are talking about something that is based on redlining or something that is based on income of women or income of Black people in particular, we know that we are historically underpaid,

even if we have the same credentials and qualifications and experience.

So, if you are using bad inputs, you are going to get bad outputs. It is very akin to what Congressman Foster said: Garbage in, garbage out. So it has the potential to solve for it if you are also being cognizant of the fact that not all biases are bad. There may be some ways to solve for it, particularly the human decision-making element at the end, of—when you get the output. But the inputs also need to be completely vetted and understood as well. So, again, if you are using something that is based off of old redlining data, that is already going to skew your results.

Mr. CLEAVER. And to any of you, one of the most dangerous things, I contend, having grown up in the deep South, is unconscious bias. There would be people who would, without any hesitation or reservation, declare that, I have designed this machine and the algorithms are completely unbiased. Is that even possible? Anybody? Yes, sir, Mr. Ghani?

Mr. GHANI. No. I don't think anybody is trained or certified today to the level where they can guarantee that an algorithm is unbiased. And I think, again, the focus on the algorithm is misleading. I think it is important to remember the algorithm doesn't do anything by itself.

Mr. CLEAVER. Yes.

Mr. GHANI. You tell it what to do. So, if you tell it to replicate the past, that is exactly what it will do. You can take bad data, but tell it, "Don't replicate the past, make it fair, here is what I mean by fair," even if that doesn't work, and as my fellow panelists were saying, the decisions we make based on the algorithm's recommendation, we don't have to do exactly what the algorithm says. We can override in certain cases when the algorithm gives us the right explanation, which we need that, and override it and/or reinforce what it is doing based on what our societal outcomes are.

So we need training for regulators to understand these nuances, because today we don't have that capacity inside agencies to understand this, implement it, and enforce these types of regulations that should exist regardless of AI. What we are talking about is not about AI. It is about societal values that should exist in every human decision-making process.

We are just talking about it today because the scale and the risks might be higher, but it is the same conversation that should have been happening continuously.

Mr. CLEAVER. My time has run out, but we had someone before this committee once who declared that he had never seen any discrimination and didn't know anything about it—and he was 60-years-old—but he said he knew some people who had. Thank you.

Chairman FOSTER. Thank you.

And the gentleman from Virginia, Mr. Rigglesman, is recognized for 5 minutes.

Mr. RIGGLESMAN. Thank you, Mr. Chairman.

Thank you, everybody, for being here. I had a whole list of questions, but now that I have heard you all, I am just going to just ask some cool things.

Dr. Thomas, I was really impressed by your thoughtful words about contextual bandits. When I did this, I had to worry about

technical or assumed bandits because we actually tried to template human behavior for node linking or information sharing and how they actually put that data together, and we had two or three people. And, by the way, when we templated each other's behavior, it was completely different. It was fantastic. But that is the algorithm we tried to do.

So, I have a question for you on these contextual bandits because, as soon as you said that, I thought, oh, goodness, I have never heard that term, specifically. We always just called them screw-ups.

Is there a list of contextual bandits that might be overlooked or not seen as egregious, and is there a prioritized set of rule set errors that you and your team or others have identified that we can point to and go look at, because, for instance, we had our huge list of errors that we had in our algorithmic rule sets that we were building through machine learning, but is there any—have you identified this list, or is there a list that we can see as far as those contextual bandits you are talking about?

Mr. THOMAS. I think we may have a miscommunication on the term, “contextual bandits.” By contextual bandit, I mean the machine learning paradigm where you make a decision based on a feature vector and then get a reward in return for it, and you optimize.

Is that the same usage of the phrase that you are using?

Mr. RIGGLEMAN. A little different, nope. You are right, because when you said, “contextual bandit,” I’m thinking about a bandit where you had a faulty piece of data put into your rule set, and that faulty piece of data came from somebody’s context and what that piece of data should do.

So, let me reframe the question. Is there any way to identify or is there a playbook or a technical order on how to remove some of those contextual bandits that, say, we as a committee can see or we can refer to?

Mr. THOMAS. Unfortunately, I am not particularly familiar with the specific definition of contextual bandit that you are using, so I apologize. “Bandits” in our setting refers to kind of like slot machines being called a one-armed bandit.

Mr. RIGGLEMAN. Oh, okay. I thought you were talking about pieces of data within it. I am sorry about that because I am using “contextual bandits” from now on. That is the greatest term I have heard in a long time.

And then, Dr. Kearns, I was listening to what you were saying. Where have improvements been in removing bias been most noticeable when you are looking at building these rule sets? Where have you seen that we have done the most improvement right now, and, again, is there something that I can go see, because I know our issues that we had in the DOD? Where can I go see where the most improvements are in removing bias and a way forward for us as we do this?

Mr. KEARNS. Yes. I guess, in my opinion, there is quite a bit of science on algorithmic fairness, and we sort of broadly know how to make things better right now, but it is, in my view, early days in terms of actual adoption, and I think one of the problems with adoption is that, for instance, even though many of the large tech

companies have small armies of Ph.D.'s who think specifically about fairness and privacy issues in machine learning, there have been relatively few actual deployments into kind of critical products at those companies, and I think that is because of the aforementioned costs that I and my fellow panelists have made, right?

If you impose a fairness constraint on Google advertising or in lending, that will inevitably come at a cost to overall accuracy. And so, in lending, a reduction in overall accuracy is either going to be more defaults or fewer loans granted to creditworthy people that would have given revenue.

I think the next important step is to sort of explain to companies, either by coercion or encouragement, that they need to think carefully about these tradeoffs, and that we need to start talking about making these tradeoffs quantitative and kind of acceptable to both the industry and to society.

Mr. RIGGLEMAN. And I think, Dr. Henry-Nickie, when you were talking about this, now that we went to tradeoffs, do you feel that—can it go the other way? Can we have too many tradeoffs when it comes to bias? And can we insert things in there that might not be real based on a political decision? I think that is the thing that everybody here wants to keep out of this, is that where is that line between making sure—do we have an algorithm writing on an algorithm for fairness, which is what we try to do, to write an algorithm to crosscheck our algorithms, or do we have to be very careful about what we identify as bias or fairness when we are making these rule sets, and where is that tradeoff, as far as can we go too political where it doesn't become fair based on the fact that we are too worried about what fairness looks like?

Ms. HENRY-NICKIE. I think it can become too political. When the CFPB tried to implement its BISG to make auto lending fair, it went extremely political and ended up screwing consumers.

Mr. RIGGLEMAN. Yes.

Ms. HENRY-NICKIE. And so, I think we have to step back collectively as regulators, on the scientific community, consumer advocates, technologists, and public policy scholars, and try to think about, how do we create collective gradations of fairness that we can all agree with? It is not a hard-and-fast issue, and, as Dr. Thomas said and Dr. Kearns, more fairness, but you hurt some groups in protected classes who we wanted better off anyway, before the algorithms were imposed.

Mr. RIGGLEMAN. I thank all of you for your thoughtfulness. I'm sorry. I know my time is almost up, but a little bit of time? I think it is up, right?

Chairman FOSTER. There is an unofficial 40 seconds of slot time. So—

Mr. RIGGLEMAN. Thank you.

Chairman FOSTER. —you now have 18 seconds.

Mr. RIGGLEMAN. You are a gracious man. Thank you, sir.

Mr. KEARNS. To just make one brief comment to make the political realities clear here: Pick your favorite specific mathematical definition of fairness and consider two different groups that we might want to protect by gender and by race. It really might be the case that it is inevitable that, when you ask for more fairness by

race, you must have less fairness by gender, and this is a mathematical truth that we need to get used to.

Mr. RIGGLEMAN. Thank you for that clarification. And thank you for your thoughtful answers. I appreciate it. And I yield back.

Chairman FOSTER. The gentleman from Illinois, Mr. Casten, is recognized for 5 minutes.

Mr. CASTEN. Thank you, Mr. Foster. I am just fascinated by this panel, and I find myself thinking that there is—I have deep philosophical and ethical questions right now that are really best answered in the context of a 5-minute congressional hearing, as all of our philosophers have taught us.

I do, though, think there are some seriously philosophical questions here, and so I would like you just to think as big picture as you can, and hopefully as briefly as you can.

First, Dr. Kearns, I was intrigued by your comment to Mr. Foster that we shouldn't define bias on the basis of inputs. I am just interested: Do any of the panelists disagree with that as a proposition?

Okay. So, then, Dr. Kearns, help me out with the second layer. Is it more useful to define the bias in terms of outputs or in terms of how the outputs are used? Because I can imagine an algorithm that predicts where that crime is likely to occur at point X. I can imagine using that for good to prevent the crime. I can imagine using that to trade against in advance of the crime and make money off of it.

How would you define the point of regulation or internal control where we should define that bias?

Mr. KEARNS. That is a great question. But it is not an easy one.

First of all, it can't possibly hurt to get the outputs right in the first place. Second, there are many situations in which the output is the decision. So, criminal sentencing is an example where, fortunately, still, the output of predictive models is given to human judges as an input to their decision-making process, but lots of things in lending and other parts of consumer finance are entirely automated now. So there is no human who is overseeing that the algorithm actually makes the lending decision. There, you need to get the outputs right because there is no second point of enforcement.

In general, I think, as per comments that people have made here already, it is true that we shouldn't become too myopically focused on algorithms and models only because there is generally a pipeline, right? There is a process to collect data from before, early in the pipeline, and there might be many steps that involve human reasoning down the line as well. But, to the extent that we can get the outputs fair and correct, that is better for the downstream process than not.

Mr. CASTEN. So then the point about these—hold on a second, because I have two more meaty questions, and, like I said, all of these are like Ph.D. theses questions.

Ms. Williams, you said in your comments that not all biases are bad. Do you have any really easy definition of how we would define good versus bad bias if we are going to go in and regulate this?

Ms. WILLIAMS. That is a good question. It is giving me a college throwback idea.

I guess it would be, if you have certain outputs that show disparity impact among groups or, let's say, certain housing decisions over the course of, let's say, three generations, if you somehow put that into your inputs or if you use that, if you are a human decision-maker who receives an output, and you decide that is something that you are going to try to correct for or solve for, then perhaps that is an example of bias for good.

Mr. CASTEN. Okay. So my last really meaty one—I am going to give you the really hard one, Dr. Henry-Nickie.

Let us assume, stipulate that people will make decisions based on bias, they will make money off of the decisions based on bias, because they already have. We already know that is going to happen.

From a regulatory perspective, what do you think is the appropriate thing to do after that has happened? Are they obligated to disclose? I know of cases where hedge funds have found that they were actually trading on horrible things in the world and the algorithm got out of control. Should they disclose that? Should they return the gains that they have had to that? Should they reveal the code?

If you are the philosopher king or queen, what is the right way for us to respond to something, having agreed that it should never have happened?

Ms. HENRY-NICKIE. Well, I think our current regulatory framework allows for that situation, and it allows us to revisit the issue, analyze, and understand who the population was that was hurt, what they look like, how much disgorgement we should go back and get in terms of redress for consumers. So I think it is completely appropriate to go back and ask—not ask, but right the harms for consumers who have been hurt.

Mr. CASTEN. But doesn't that assume that they have already disclosed it? In my scenario, where my algorithm is predicting a crime and I figure out how to short the crime—

Ms. HENRY-NICKIE. Disclosure does not absolve you of liability.

Mr. CASTEN. But if you are not obligated to disclose, how are we ever going to find out as regulators that it happened?

Ms. HENRY-NICKIE. I think that is a really good question. If you are not obligated to disclose, then we are in a Catch-22, and then how do we find and identify and detect, and how do we hold them accountable? I think it is important for the CFPB, the DOJ, the OCC, and the Federal Reserve to have their enforcement powers intact and strengthened to be able to hold bad actors, regardless of intent, accountable for their decisions.

Mr. CASTEN. Well, I am out of time. I yield back.

And I am sorry, Mr. Ghani. I know your hand was up. Feel free to submit comments. And, if any of you have thoughts on that, feel free to submit them. Thank you so much.

Chairman FOSTER. I believe it is likely that, if we don't have votes called, we will have a second round of questions.

The gentleman from North Carolina, Mr. Budd, is recognized for 5 minutes.

Mr. BUDD. Thank you, Mr. Chairman. This is a fascinating conversation. Professor Ghani, was there something that you were—



I think your hand was raised earlier. I have other questions for you, but if you wanted to clarify?

Mr. GHANI. Yes, I wanted to go back to the first Ph.D. thesis that was talked about: Is it enough to get the outputs right, or is it important how those outputs are going to be used? And I think that that is probably the most critical question that has been asked today, because it doesn't matter what your outputs are if you don't act on them appropriately, right?

Here is an example. If you are going to take the example of, you are not going to get all the outputs right, period. AI will never be good enough to get everything perfectly right. It is going to make mistakes. What mistakes are more important to guard against really depends on how those outputs are going to be used.

If we predict somebody might commit a crime and the intervention we have is going and arresting them, that is a punitive intervention. False positives are back, like disproportionate false positives; much, much, much worse than missing people.

If we predict that somebody is going to commit a crime, but they have a mental health need, and we are going to send out a mental health outreach team to help them, give them the support services they need, then missing people disproportionately, false negatives, are much, much, much worse than false positives.

And so the intervention is what really decides how we design these algorithms, and it is not the output; it is—we can have the same output. Different interventions will require different notions of what to optimize for and the impact of the bias in society. So, I want to make that distinction clear—

Mr. BUDD. Thank you.

Mr. GHANI. —because it does matter quite a bit.

Mr. BUDD. Thank you.

And, in terms of using this AI for giving people credit, I think we can agree that giving consumers access to credit can fundamentally change their lives, and this is one tool that we are using that can help them do so. It allows consumers to buy a home, a car, pay for college, or start a small business. Using alternative data such as education level, employment, status, rent, or utility payments has the potential to expand access to credit for all consumers, especially those on the fringes of the credit score range.

A recent national online survey shows that 61 percent of consumers believe that incorporating access to their payment history and their credit files will ultimately improve their scores. The same survey also found that more than half of consumers felt empowered when able to add their payment history into the credit files, and they cited the ability to access more favorable credit terms as one of the biggest benefits of sharing their financial information.

So can you further elaborate, Mr. Ghani, on how the use of alternative data expands access to credit for low- to moderate-income consumers who would otherwise be unable to access that same credit?

Mr. GHANI. Yes. I would go back to what Dr. Kearns was saying, that it is really not about the inputs, right? The sandbox we need to create is to enter those things in and then measure the outputs and then look at disparities in the rates at which you are going to offer loans or credits to people that you wouldn't have before.

So, imagine our societal goal is that the lending decisions we want to make should serve to reduce or eliminate disparities in home ownership rates across, let's say, Black and white individuals, or minorities and white individuals. If that is the societal goal we want to have, then these inputs may or may not help us achieve that, and what we want to be able to do is to test that out, have a framework for testing it, validating it, certifying that it is actually doing that, and then put this into place after we have done trials, just like other regulatory agencies do.

Starting with, if we put in these inputs, would it help? We don't know, but I think putting the right outcomes in place that you want to achieve and then testing it is the right approach to take.

Ms. HENRY-NICKIE. I would add to that.

Mr. BUDD. I want to add an open question here, and if you can comment on the same thing, but then answer the open question, and that is ways that we can be more encouraging to use tools like alternative data and AI to raise access to credit and lower the overall costs for consumers, if there are ways that we can encourage that here, so please?

Ms. HENRY-NICKIE. I will take that question first.

I think we have to be careful about experimenting with people's—consumers' financial lives. I think a healthy way to discover what our new products are out there might be through pilots, might be through continued active observation, and also vigilant oversight, as in the Upstart case.

To your question before, how do rental payments help to expand access to credit on alternative data? For example, in some markets, rental payments are as high as a mortgage or even higher, and, if you, as a first-time home buyer about to enter into this process have only had a rental payment history that is consistent, stable, not late, then taking that feature, substituting it for what a mortgage payment and standing in for mortgage payment—excuse me—could then push you above the margin to have the model predict that you were a good credit risk.

Mr. BUDD. Thank you, and I yield back.

Chairman FOSTER. Thank you.

The gentleman from Indiana, Mr. Hollingsworth, is recognized for 5 minutes.

Mr. HOLLINGSWORTH. Good afternoon. I appreciate everybody being here. Certainly, were my wife here, she would tell you that I am far outside my circle of competence. So, I am going to ask a lot of really stupid questions and let you all give me really intelligent answers to those stupid questions.

Can you clarify—the word “fairness” has been thrown around a lot. Can you clarify what you mean by fairness, the five of you? Have at it. Dr. Kearns, Ms. Williams, Dr. Thomas, everybody, anybody?

Ms. WILLIAMS. Okay. I will go first.

Mr. HOLLINGSWORTH. Okay.

Ms. WILLIAMS. For me, I look at fairness as ensuring that all groups have equal probability of being assigned favorable outcomes.

Mr. HOLLINGSWORTH. All groups have equal probability of being assigned outcomes irrespective of their current situations, or all in-

dividuals similarly situated are assigned the same outcome—the same probability of outcomes?

Ms. WILLIAMS. The latter.

Mr. HOLLINGSWORTH. The latter. Okay.

Dr. Kearns?

Mr. KEARNS. There are too many definitions of fairness, as we have already alluded to, but the vast majority of them begin with, the user has to identify what group or groups they wanted to protect and what would constitute harm to those groups. So, it is maybe a racial minority, and the harm is a false loan rejection, rejection for a loan that they would have repaid.

Mr. HOLLINGSWORTH. You have clearly short-circuited to what I was getting at, which is, we have a lot of senses of fairness and a lot of senses of what we want done, but the requirement in AI and algorithms is that we make explicit that which is right now implicit, right, and you have to be very good at making that explicit because the algorithm itself is going to optimize for what you tell it to optimize for, right? And so, you are going to have to make very clear what you were trying to optimize for in order to get that outcome, and then, to your point, what side you were unwilling to live with, right? I am unwilling to live with the extra risk on this side or perhaps that side depending on what situation you are in.

So, not only do you have to have a lot of awareness about exactly what you want to optimize for, but also a lot of awareness about, in the context, what you are really worried about and what you are concerned about, the false positives or the other side of it.

Dr. Thomas?

Mr. THOMAS. I absolutely agree. I missed what the precise question—

Mr. HOLLINGSWORTH. No. I saw you nodding your head and I didn't know if you had a comment to the previous question about fairness.

Mr. THOMAS. I am generally just in agreement that you are hitting on very good points, and—

Mr. HOLLINGSWORTH. I shall take that back to my wife. Maybe my circle of competence is bigger than I thought it was.

Mr. THOMAS. You are hitting on the point that there are many different definitions of fairness. The question of which one is right and nailing it down is very important.

Mr. HOLLINGSWORTH. Yes.

Mr. THOMAS. And something that I think you might be kind of dancing around is this idea that the negative outcomes that are consistent with different definitions of fairness can often all seem bad. There can be two different definitions of fairness, and, if we pick one, it means we are saying that the undesirable, unfair behavior of the other is necessarily okay.

Mr. HOLLINGSWORTH. Yes. And I think Dr. Kearns talked a little bit about this earlier, and it is something that puzzles me a lot because I think, in some places, the tradeoff in fairness for one group may mean less fairness in the other. Did you say that, Dr. Kearns?

Mr. KEARNS. I did.

Mr. HOLLINGSWORTH. Yes. And this is something that others have hit on as well, that we are going to have to grow comfortable saying to ourselves that we are going to trade fairness here for fair-

ness there—and not just more fairness for perhaps less accuracy in the model itself, which is something we have had more comfort in. But trading fairness and risk to a certain group is something we have been really uncomfortable with because we want fairness for everybody in every dimension, which seems—I don't want to say impractical, but it seems challenging inside an AI algorithm in optimization.

Mr. KEARNS. I would say—

Mr. HOLLINGSWORTH. Do you agree with this?

Mr. KEARNS. —that is, in fact, impractical. And let me just, while we are in the department of bad news, also point out that all of these definitions we are discussing are basically only aggregate definitions and only provide protections at the group level.

Mr. HOLLINGSWORTH. Right.

Mr. KEARNS. So, for instance, you can be fair, let's say, by race in lending. And if you are a Black person who was falsely rejected for a loan, your consolation is supposed to be the knowledge that white people are also being falsely rejected for loans at the same rates. There is literature on individual notions of fairness, definitions of fairness that try to make promises to particular people that are basically impractical and feasible. It is sort of a theoretical curiosity, but no more.

Mr. HOLLINGSWORTH. Yes. I appreciate that.

Each of you have talked a little bit about the pipeline that AI algorithms aren't birthed in the ether, right, that they rely on data, A; and, B, individuals craft these. I wonder if you might talk a little bit about the biases that we are talking about, are they more likely to arise from the algorithm itself, or are they likely to arise from the coder or the drafter of said algorithm, or are they likely to arise from the data that is being input into them? Where should we look first if we are going to look through that pipeline? Ms. Williams?

Ms. WILLIAMS. I would say look on the human level first, because a human is going to discern what is the narrative that they are actually solving for, and then therefore, what is the data that they are going to use, and they discern the quality of data that is used, and they then discern the training set that is created and how that is functional. I also want to be clear that I don't think that there are a bunch of mad coders sitting in a basement somewhere.

Mr. HOLLINGSWORTH. Yes. The fair expectations of society.

Ms. WILLIAMS. I don't think that is it. It is very—you don't know what you don't know.

Mr. HOLLINGSWORTH. Yes. I agree.

Ms. WILLIAMS. And I think, oftentimes, if people pick data that is available to them, they may not do a ton of due diligence to find additional data or data that may even offset some of the data that they already have. But I would say, start at the human level first, because that is where everything else sort of begins in terms of picking the data, the quality of data, and then actually doing the coding.

Mr. HOLLINGSWORTH. Yes. Thank you.

With that, I yield back, Mr. Chairman.

Chairman FOSTER. Thank you.

And now, I guess we have time for a quick second round of questions. Votes are at 3:30, and we have nerves of steel here, I'm learning, so we will give it a try here.

So, I will recognize myself for 5 minutes.

I would like to talk about the competitive situation that would happen when you have multiple companies, each running their own AI and, say, offering credit to groups of people.

If you just tell them, "Okay, maximize profits, that is a mathematically well-defined way to program your AI," and they would all do it identically, and the competition would work out in an understandable way.

Now, if you impose a fairness constraint on these, first off, that will reduce the profitability of any firm that you impose the fairness constraint on, so they are not simply maximizing profits, and then, if a new competitor may come in and, say, and say, oh, there is a profitable opportunity to cherry pick customers that you have—that your fairness constraint has caused you to exclude, and is that a mathematically stable competition? Has that been thought about? Do you understand the problem I am talking about?

Mr. KEARNS. If I understand you correctly, there is literature in economics on whether, for instance, racial discrimination in hiring can actually be a formal equilibrium of the Nash variety.

Chairman FOSTER. Maybe that is a good description.

Mr. KEARNS. Gary Becker was a very famous economist who did a lot of work in the 1960s and 1970s on exactly this topic, and it is complicated, but the top-level summary of his work is that the argument that you can't have discrimination in hiring at equilibrium because you wouldn't be competitive, because you are irrationally excluding some qualified sectors of the job market. He actually shows that, in fact, you can have discrimination even at equilibrium.

Chairman FOSTER. So, one of the questions would be whether you are better off actually having multiple players here. So, if someone is erroneously excluded because of some quirk in some model, then it would be to the advantage of society overall to have multiple players, so that person could go to a second credit provider?

Mr. KEARNS. You are asking kind of the reverse of Becker's question, which is, if you don't have sort of regulatory conditions on antidiscrimination, for instance, might there be arbitrage opportunities for new entrants? I don't know that that question has specifically been considered, but it is a good question.

Chairman FOSTER. Yes, Mr. Ghani?

Mr. GHANI. I think one thing I would point out is the premise that, if you put those constraints there, the profits will go down; that is not a guarantee. We don't know that, and here is why, right? I think it was Dr. Kearns who was talking about how there are a lot of people who just—we don't know what happens in somebody who was never—the type of person was never given loans before. What happens when you give them a loan, right?

So it could be that, when you start adding these fairness constraints, it turns out that you don't actually lose profits, and, in fact, you might increase profits. These are things called counterfactuals, where, because you have never given loans to peo-

ple like this, you don't know what the outcomes are. You might have just—the human decision-making process that existed before was only giving loans to people they thought were going to pay back loans.

Chairman FOSTER. That has to do with the exploratory phase of programming your neural—

Mr. GHANI. That is correct.

Chairman FOSTER. —to actually do random, crazy stuff because you may discover a pocket of consumers—

Mr. GHANI. Hopefully better than random, crazy stuff, but some smarter version of that, yes.

Chairman FOSTER. Yes. And so, this question of similarly situated people, that depends on the scope of the data that you are looking at, because two people can look similarly situated if you only look at their family and their personal history, and then, if you look at a wider set of things—I think this is what came up with Apple and Goldman where, if you just looked at one-half of a couple's credit information, you would give a different credit limit on a credit card, I think it was, whereas, if you look holistically at both halves of a couple, you get a different answer. And there is no obvious right answer to how wide you should spread your field of view here.

Is that an unresolvable problem that you are going to need Congress to weigh in on? Yes?

Mr. GHANI. I think this is exactly why these systems need people in the middle, but, also, these systems need collaborative processes upfront, including the people who are going to be impacted by them. If you start including those communities, they will tell you that there is actually really good work. There is a group in New Zealand who has been doing, how do we incorporate community input into designing these types of algorithms? What input attributes do we use that best represent the differences and similarities across these?

So it is inherently—it is going to be hard to automate that today, but I think that is the process we need, which is to include the community that is being impacted and humans in the loop, in the system, coming up with some of these things, and collaborating with the machines.

Chairman FOSTER. Well, okay. That sounds ambitious. I am just trying to think of assembling groups that are sufficiently knowledgeable about the nuts and bolts of this and to have—and where you are balancing the people who wind up winning and losing according to the tradeoffs you are going to be making.

Mr. GHANI. The challenge is some of the—the amount of data you have on people is also a function of who they are. Some people are less reluctant or more reluctant to give data about themselves. They may have less of a history. Immigrants are coming in who don't have a background, and credit history, so missing information. It is not just that you have the data, you can just get it and compare those. You might not have that data, and that is also biased in the data collection process.

Chairman FOSTER. Okay. I will gavel myself down and recognize the ranking member for 5 minutes.

Mr. LOUDERMILK. Thank you, Mr. Chairman.

Unfortunately, we only have a few more minutes. I think we could all be here all day discussing this.

Something Ms. Williams and Dr. Kearns said earlier has really been resonating: Not all bias is bad. We agree. In fact, if we take kind of the model we have been talking about, loan applications, whether there is a mortgage or not, the whole purpose of the AI platform is to be biased, right? That is the actual purpose, is to be biased, but what is the bias that we want? We want those who are likely to pay a loan back, really is what we are getting at. So I think I see what you are saying. There is some bias that you want in there.

What is the bias we want to eliminate, is really the question, and that goes to something Dr. Kearns said, well, if we reprogram it to make sure more of one racial group gets more approval, then you may see a gender impacted. And so, this is kind of a conundrum we are in until we figure out or define what bias do we want in a particular system, but, more importantly, what do we not want?

When I look at it as what do we not want, if Mr. Budd and myself are identical—I know that we are identical in income because I know what he does for a living, and that the law doesn't allow him to take any other income, right? But if he is Hispanic, I am white, and the chairman is Black, and we all have the same income, we all have the same assets, we all have basically the same biographical data, do we all get the same result, whether it is approval or disapproval? That is really what I think we are trying to get to.

It isn't that we weren't happy with the result that came out, but we have to go back and find out why. And that is what we are getting at.

Mr. Ghani, if it exists, and since algorithms are my mathematical equation, really, I think part of the problem is, when you get into the machine learning and the algorithm begins to rewrite itself, how do we track it?

We verify the data is good. I think most of the problems we have are probably in the data and the appropriateness of the data. Let me say not just in the raw data, but the appropriateness of the data.

But if we do want to check the algorithms, is there a way of running what I would call in the network world an audit trail in the development of the algorithm, throughout the operation of the algorithm, and each phase of decision it is making, and the actual coding, and is there a way to go back and do a forensic audit trail on these algorithms?

Mr. GHANI. Yes, absolutely, and I think that is the right approach, is you can audit the data, and that is great. But then you are going to—I think the starting point is you want to tell it what you want the system to achieve. Then, you want to turn those into technical requirements for the system to see what to tell it to do, and then you want to confirm that it did what you told it to do, and then you want to test it and see, does it continue to do what you—what it did yesterday, right?

When you answer, what should we ask a company to disclose, it is not the algorithm. It is not the code. It is not the data. It is this

entire audit trail, and that is what we need to look at to figure out where it is happening.

Mr. LOUDERMILK. Well, that's an interesting aspect, and take it a step further. The difference between software and artificial intelligence is we expect software to give us the same result every time, right? That is not the case with artificial intelligence, correct, because artificial intelligence is always looking for other data, and it may give us a different outcome the next day based on something that changed the day before, and it may rewrite itself to learn new things.

I think that is some of the challenge going forward is, if you tell the machine that is not the right answer, it is going to look for a different answer in the future. This is stuff we wrote science fiction about just 10 years ago, right? So, when we code the algorithms themselves, can we actually program in the artificial intelligence platform to do systematic reports throughout the process?

Mr. GHANI. Absolutely.

Mr. LOUDERMILK. Okay.

Mr. GHANI. That should be standard. That should be part of our training programs for people who are building these systems. It should be part of training for auditors who are doing compliance. Absolutely, that is the right approach.

Mr. LOUDERMILK. Okay. And the last part is probably more a statement than a question. In my opening statement, I talked about different analytical models. I think the one that concerns us the most is what I call the execution model. We have presentation of data. We have predictive analysis. We have prescriptive analysis that prescribes, okay, approve or don't approve. And we can do that, but, yet, there is a human element making the same decision.

It is like the backup warning on my car that beeps and it says something is behind me. It doesn't stop the car. I still make the decision. But if you watch the Super Bowl, the Smart Park, right, it is actually making the decisions. In this case, it is the machine making the decision of go/no-go on the loan. It is executing on that, and I think, until we get this fixed, we may need to look at, is there an appeal process for that go/no-go that a human element can go in and work?

So, thank you, Mr. Chairman. It sounds like our warning bell is going off, and my time has expired.

Chairman FOSTER. Thank you.

I would like to also thank our witnesses for their testimony today.

Without objection, the following letters will be submitted for the record: the Student Borrower Protection Center; Cathy O'Neil of O'Neil Risk Consulting & Algorithmic Auditing; the BSA Software Alliance; The Upstart Network, Incorporated; and Zest AI.



The Chair notes that some Members may have additional questions for this panel, which they may wish to submit in writing. Without objection, the hearing record will remain open for 5 legislative days for Members to submit written questions to these witnesses and to place their responses in the record. Also, without objection, Members will have 5 legislative days to submit extraneous materials to the Chair for inclusion in the record.

This hearing is now adjourned.

[Whereupon, at 3:33 p.m., the hearing was adjourned.]



# **A P P E N D I X**

February 12, 2020

**Equitable Algorithms: Examining Ways to Reduce AI Bias in Financial Services**

Testimony by  
 Rayid Ghani,  
 Distinguished Career Professor  
 Machine Learning Department and the Heinz College of Information Systems and  
 Public Policy  
 Carnegie Mellon University

Before the  
 Committee on Financial Services  
 Artificial Intelligence Task Force  
 U.S. House of Representatives

The Honorable Maxine Waters, Chairwoman  
 The Honorable Patrick McHenry, Ranking Member

Wednesday, February 12, 2020

Chairwoman Waters, Ranking Member McHenry, Members of the Committee, thank you for hosting this important hearing today, and for giving me the opportunity to submit this testimony.

My name is Rayid Ghani and I am a Distinguished Career Professor in the Machine Learning Department and the Heinz College of Information Systems and Public Policy at Carnegie Mellon University. I've worked in the private sector, in academia, and extensively with government agencies and non-profits in the US and globally on developing and using Machine Learning and AI systems to tackle social and public policy problems across health, criminal justice, education, public safety, human services, and workforce development in a fair and equitable manner.

Artificial Intelligence (or Machine Learning)<sup>1</sup> has a lot of potential in helping tackle critical problems we face in society today, ranging from improving the health of our children by reducing their risk of lead poisoning<sup>2</sup>, to reducing recidivism rates for people in need of mental health services<sup>3</sup>, to

---

<sup>1</sup> I will use the terms AI and Machine Learning interchangeably in this testimony

<sup>2</sup> Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning. Potash et al. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2015)

<sup>3</sup> Reducing Incarceration through Prioritized Interventions. Bauman et al.. ACM SIGCAS Conference on Computing and Sustainable Societies, 2018.

improving educational outcomes for students at risk of not graduating from school on time<sup>4,5</sup>, to improving police-community relations by identifying officers at risk of adverse incidents<sup>6</sup>, to improving health and safety conditions in workplaces<sup>7</sup> and in rental housing<sup>8</sup>. AI systems have the potential to help improve outcomes for everyone and result in a better and more equitable society. At the same time, any AI (or otherwise developed) system that is affecting people's lives has to be explicitly built to focus on increasing equity and not just optimizing for efficiency. It is important to recognize that AI can have a massive, positive social impact but we need to make sure that we put guidelines in place to maximize the chances of the positive impact while protecting people who have been traditionally marginalized in society and may be affected negatively by the new AI systems.

An AI system, designed to explicitly optimize for efficiency, has the potential to result in leaving “more difficult or costly to help” people behind, resulting in an increase in inequities. **It is critical for government agencies and policymakers to ensure that AI systems are developed in a responsible and collaborative manner**, including and incorporating input from all groups of stakeholders including: developers who build and deploy AI systems, decision-makers who implement the systems in their workflows, and the community being impacted by these systems. **Integrating input from these diverse voices is a critical element of ensuring that new AI systems result in equitable outcomes for everyone.**

#### **Equitable outcomes and not “just” unbiased or equitable algorithms**

Contrary to a lot of work in this area today, I believe that “simply” developing AI algorithms that better account for fairness and bias is generally not sufficient to achieve more equitable decisions or outcomes. Rather, the goal of these efforts should be to make entire systems and their outcomes equitable. Since algorithms are typically not (and should not be) making autonomous decisions in critical situations, the entire decision-making system includes the AI algorithms, the decisions that are being taken by humans using input from those algorithms, and the impact of those decisions. It is entirely possible to have a perfectly fair and equitable algorithm providing fair and equitable recommendations but the human decisions following them may be biased or the interventions allocated as a result of that human decision are not as effective for certain people as they are for others, resulting in inequity in outcomes.

At the same time, it is possible to design a system that contains an algorithm that is not fair but coupled with the appropriate bias mitigation and intervention plan, can result in increasing equity in

---

<sup>4</sup> A Machine Learning Framework to Identify Students at Risk of Adverse Academic Outcomes. Lakkaraju et al. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

<sup>5</sup> <http://www.dssgfellowship.org/project/identifying-factors-driving-school-dropout-and-improving-the-impact-of-social-programs-in-el-salvador/>

<sup>6</sup> Early Intervention Systems – Predicting Adverse Interactions Between Police and the Public. Helsby et al. Criminal Justice Policy Review, 2017.

<sup>7</sup> <http://www.dssgfellowship.org/project/improving-workplace-safety-through-proactive-inspections/>

<sup>8</sup> <http://www.datasciencepublicpolicy.org/projects/public-safety/san-jose-housing/>

outcomes. In some recent preliminary work we did with Los Angeles City Attorney’s office, we found that by careful consideration and analysis, we can mitigate the disparities that a potentially biased algorithm may create and coupled with a tailored intervention strategy, the system has the potential to result in equitable criminal justice outcomes across racial groups<sup>9</sup>.

### **AI systems optimize for what the developers tell them to optimize for**

AI algorithms are neither inherently biased nor unbiased (in the societal sense). They typically work by taking historical data and attempting to build a “model” that replicates some outcome that is specified in that historical data while attempting to also generalize in to the future. The developers of such a system often have to specify how to manage the two tradeoffs – how much of the past to replicate and how much to generalize to the (unseen) future. When such a system is built, the developers of the system also specify what metric(s) to optimize for. If the system is asked to correctly predict as many of the past decisions that were provided for it to “learn” from as possible, that is exactly what it attempts to do, regardless of the race, gender, age, or income of the people who these decisions were about. That is one step where a lot of bias may come in to the decisions recommended by this system.

The AI developer can, in fact, tell the algorithm to balance replicating as many human decisions correctly as possible with ensuring fairness and equity across certain protected attributes of people that we care about. Sometimes, the developers fail to incorporate equity considerations in building their AI models, which is of course equivalent to choosing a metric that attempts to replicate as many human decisions as possible, possibly resulting in re-creating and reinforcing historically biased decision processes. In these cases, it is important to remember that the human processes that designed the AI system should own the blame rather than passing it off to an AI algorithm that is being guided and optimized incorrectly, for the wrong goals.

### **AI is forcing us to make societal (and public policy) values explicit**

Because an AI system requires us to define exactly 1) what we want to optimize it for, 2) which mistakes are costlier (financially or socially) than others, and 3) by how much, it forces us to make these ethical and societal values explicit. It is important to know that these values are of course implied in any decision-making process, including all the human decision-making processes that exist today, but are not necessarily made explicit. These implicit values coded in humans making decisions when biased and unfair, result in inequitable outcomes. For an AI system to function, these values need to be provided as a critical input. For example, for a system that is recommending lending decisions, we may have to 1) specify the differential costs of flagging someone as unlikely to pay back a loan and being wrong about it versus predicting that someone will pay back a loan and being wrong about it, and 2) specify those costs explicitly in the case of people who may be from

---

<sup>9</sup> Predictive Fairness to Reduce Misdemeanor Recidivism Through Social Service Interventions. Rodolfa et al. Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (ACM FAT\*) 2020.

different gender, race, income, or education level groups. While that may have happened implicitly in the past and with high levels of variation across different human decision makers (loan officers in this case), with AI-assisted decision-making processes, we are forced to define them explicitly.

One key question we have to answer here is who and how should we come up with these sets of values for a given problem setting. Unfortunately, today, these decisions are too often left essentially by default to the AI system developer or an arbitrary set of individuals who define those values in an AI algorithm (explicitly or implicitly). The recommendations at the end of this testimony go into more detail on what I recommend should be done but it certainly should not be left to the AI system developer making those choices alone; the team and process should include all stakeholders including policymakers and the community being impacted by this system.

### All Types of Biases are Not Equal

An AI system (or human) can be unfair in a variety of ways and there is no universally-accepted definition of what it means for an AI system to be fair. Take the example of a system being used to make loan determinations. Different people might consider it “fair” if:

- It makes mistakes about denying loans to an equal number of white and black individuals
- The chances that a given black or white person will be wrongly denied a loan is equal, regardless of race
- Among the population who were denied loans, the probability of having been wrongly denied a loan is independent of race
- For people who should be given loans, the chances that a given black or white person will be denied a loan is equal
- The lending decisions serve to reduce or eliminate disparities in home ownership rates across black and white individuals

These different notions of fairness have formal names and definitions in research literature<sup>10</sup> and a great deal of research has been done describing these fairness notions in different fields. In different contexts, reasonable arguments can be made for each of these potential definitions, but unfortunately, not all of them can hold at the same time<sup>11,12</sup>. In general, understanding which type of bias should be prioritized and weighted more than others requires consideration of the societal goals and a detailed discussion between decision makers, AI developers, and most importantly those who will be affected by the application of the model. Each perspective may have a different concept of fairness and a different understanding of harm involved in making different types of errors, both at individual and societal levels. Practically speaking, coming to an agreement on how fairness should

<sup>10</sup> Sahil Verma and Julia Rubin. 2018. Fairness Definitions Explained. In Fair- Ware’18: IEEE/ACM International Workshop on Software Fairness, May 29, 2018, Gothenburg, Sweden. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3194770.3194776>

<sup>11</sup> Alexandra Chouldechova. 2017. Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. *Big Data* 5, 2 (6 2017), 153–163. <https://doi.org/10.1089/big.2016.0047>

<sup>12</sup> Moritz Hardt, Eric Price, ecprice, and Nati Srebro. 2016. Equality of Opportunity in Supervised Learning. In *Advances in Neural Information Processing Systems* 29. Neural Information Processing Systems Foundation, Barcelona, Spain, 3315– 3323.

be measured in a purely abstract manner is likely to be difficult. Often it can be instructive instead to explore different options and metrics based on preliminary results, providing tangible context for potential trade-offs between overall performance and different definitions of equity and helping guide stakeholders through the process of deciding what to optimize<sup>13</sup>. While we have a more comprehensive set of guidelines we term the Fairness Tree<sup>14</sup>, some of the guidelines we have developed and use in our work include:

- If the intervention is punitive in nature (e.g., determining whom to deny loan), individuals may be harmed by intervening on them in error so we may care more about metrics that focus on false positives.
- If the intervention is assistive in nature (e.g., determining who should receive loan forgiveness), individuals may be harmed by failing to intervene on them when they have need, so we may care more about metrics that focus on false negatives.
- If the available resources are significantly constrained such that we can only intervene on a small fraction of the population at need, a different set of metrics may be of more use (see Fairness Tree<sup>10</sup> for more details).

### **Bias in AI systems can come from a lot of sources and it's important to separate them out**

Bias may be introduced into an AI system at any step along the way and it is important to carefully think through each potential source and how it may affect the results. In many cases, some sources may be difficult to measure precisely (or even at all), but this doesn't mean these potential biases can be readily ignored when developing interventions or performing analyses. These sources include

**1. Biased data sources:** due to either data being used to build an AI system not being representative of the population it will be used to make decisions for, or having incorrect/biased outcomes for certain people (based on historical biases in society and/or human decision-making such as over-policing black communities), or the unknowability of certain outcomes from past decisions (for instance, you can't know whether or not an individual who was denied a loan would actually have repaid it had it been granted).

**2. Bias due to decisions made by AI developers when designing the system:** I will not go into detail here but would refer to other literature<sup>15</sup> that describes different analytical decisions that are made when developing an AI system that can lead to biases.

**3. Application Bias:** This is often not due to the AI algorithm being biased but because of the way the results of an AI algorithm are applied. One way this might happen is through heterogeneity in the effectiveness of an intervention across groups. For instance, imagine an AI system built to identify individuals most at risk for developing diabetes in the near future for a particular preventive

<sup>13</sup> Predictive Fairness to Reduce Misdemeanor Recidivism Through Social Service Interventions. Rodolfa et al. Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (ACM FAT\*) 2020.

<sup>14</sup> <http://www.datasciencepublicpolicy.org/projects/aequitas/>

<sup>15</sup> <https://textbook.coleridgeinitiative.org/chap-bias.html>



treatment. If the treatment is much more effective for individuals with a certain genetic background relative to others, the overall outcome of the effort might be to exacerbate disparities in diabetes rates even if the AI algorithm itself is unbiased.

### What does it take to create AI systems that lead to equitable outcomes for society?

The following steps need to be taken to attempt to create AI systems that are likely to lead to equitable outcomes for society:

1. **Defining** the (equitable) outcomes we want to achieve in society (which includes the societal values and a collaborative, multi-stakeholder process).
2. **Translating/Mapping** those desired societal outcomes into analytical requirements that the AI system should optimize for.
3. **Building** an AI system that fulfills those analytical requirements and releasing documentation on how it was built to achieve those goals. This step includes
  - A. **Detecting** biases in intermediate/iterative versions of the system
  - B. **Understanding** the root causes of the biases
  - C. **Improving** the system by reducing the biases (if possible) or selecting tradeoffs across competing objectives
  - D. **Mitigating** the impact (and coming up with an overall mitigation plan) of the residual biases of the system
4. **Validating** through a trial (and providing evidence) that the AI system did, in fact, fulfil those requirements and achieve the initial outcomes defined in step 1 before deploying the system.
5. **Continuous Monitoring & Evaluation** of the entire system (AI algorithm followed by human decisions) during its lifetime to ensure that it continues to achieve equitable outcomes from step 1.

It is important to note that the steps above are **not purely technical**, but rather involve understanding existing **social and decision-making processes and systems** as well as **collaboratively coming up with solutions for each step**. These steps may require new data to be collected and existing processes to be modified in order to ensure equitable outcomes. For example, if data about race or gender was not being collected in the past, and the goal is to monitor and achieve equity across different groups of gender and race, it will require new data collection processes. Likewise, goals surrounding fairness and equity must be actively integrated into the

modeling, evaluation, and decision-making processes. A considerable body of work<sup>16 17 18</sup> has demonstrated that notions of “fairness through unawareness” (e.g., simply excluding or ignoring these protected attributes in AI systems) is insufficient for achieving equitable results, both because these attributes are often highly correlated with other predictors and due to historical disparities in outcomes themselves.

### **Moving Forward to a More Equitable Society: Our Recommendations**

It is critical and urgent for policymakers to act and provide guidelines and/or regulations for both the public and private sector organizations using AI-assisted decision-making processes in order to ensure that these systems are built in a transparent and accountable manner and result in fair and equitable outcomes for society. As initial steps, we recommend:

#### **1. Expanding the existing regulatory environment to account for AI-assisted decision-making**

The potential risks and benefits of AI to society are as wide and varied as the contexts to which it can be applied. A model or algorithm that yields beneficial and equitable outcomes in one context might yield just the opposite in another. While AI algorithms across different areas have a lot in common, developing a unified regulatory framework for AI that works well across all possible applications is likely to be an unrealistic proposal. Rather, the need for regulatory oversight is inherent in the application of this tool to achieving societal goals across different policy domains.

**Instead of creating a Federal AI regulatory agency across policy areas, we should expand the already existing regulatory frameworks in different policy areas, building on their domain-specific expertise while updating them to account for AI-assisted decision-making.**

The regulatory bodies already exist — including SEC, FINRA, CFPB, FDA, FEC, FTC, and FCC — and are well-positioned with the responsibility and policy area knowledge for ensuring compliance with existing regulations, but will need to account for new challenges in applying that oversight introduced by the growing application of AI to their domains. These bodies typically regulate the inputs that go into a decision-making process (for example, what attributes cannot be used such as

---

<sup>16</sup> Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*. ACM Press, New York, New York, USA, 214–226. <https://doi.org/10.1145/2090236.2090255>.

<sup>17</sup> R. G. Fryer, G. C. Loury, and T. Yurek. 2007. An Economic Analysis of Color-Blind Affirmative Action. *Journal of Law, Economics, and Organization* 24, 2 (11 2007), 319–355. <https://doi.org/10.1093/jleo/ewn053>

<sup>18</sup> Toon Calders and Indre Žilobaitė. 2013. Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures. In *Discrimination and Privacy in the Information Society. Studies in Applied Philosophy, Epistemology and Rational Ethics, Volume 3*, Springer Press, Berlin, Germany, 43–57.

race or place of residence) and often the processes themselves, but do not always focus on the outcomes produced by these processes. We recommend expanding these regulatory bodies to:

1. Update the regulations to make them outcome-focused.
2. Update the regulations to ensure they apply to AI-assisted decision making.
3. Define the set of artifacts an organization (government or industry) should publicly release before deploying (and ideally during the development phases of) an AI system. This includes information on how the system was built, what it was designed to optimize for, what tests were run to check if it did, what types of people it is effective for, who does it fail for, how long was it in trials for, and how did the effectiveness change over time. Ideally this should be put in place for any process involving decision making of any kind, whether human decisions or AI-assisted decisions but becomes critical in cases where the scale of deployed AI systems increases the risk. This set will need to vary based on the impact this system can have on people's lives.
4. Define a set of risks that could lead to inequities that need be considered when building an AI system and a mitigation plan for each of these risks.
5. Set up an extended data collection process and infrastructure to collect additional data attributes (such as race, gender, or income) that may not already be collected but are necessary to measure equity outcomes (to deal with the "fairness through unawareness" issue described earlier).
6. Set up evaluation standards to compare the performance of these systems to the human decision-making processes currently being used.
7. Define standards around the explainability of the AI systems in order to provide recourse to individuals who may be adversely impacted by the decisions made using the system.

These expanded bodies should be responsible for defining standards as well as for continuous monitoring, audits, and compliance with the standards and regulations.

## **2. Creating Trainings, Processes, and Tools to Support Regulatory Agencies in their Expanded Roles**

As these agencies expand their role, they will need to be supported by increasing their internal capacity to fulfil this role and ensure that regulations are being effectively complied with. We recommend creating trainings, processes, and tools to help them

1. Understand where existing regulations may and may not be well-adapted to applications involving AI-assisted decision-making.
2. Understand and define what equitable outcomes standards to set.
3. Understand how to evaluate whether the requirements created for an AI system were in fact aligned with the identified societal equitable outcomes.
4. Understand how to evaluate whether the AI system did in fact do what it was designed to do.

5. Develop a continuous monitoring and audit process and tools (such as Aequitas<sup>19</sup>) to support the audit process.
6. Create standards for when a system should “expire” and a corresponding renewal process.

### **3. Procuring AI systems should include Key Requirements in the Request for Proposals (RFP) Process**

Government agencies and corporations putting out RFPs for AI systems that are making critical decisions and affecting people should require proposers/bidders to include:

- An explicit initial project phase to gather requirements for what it would mean to have equitable outcomes and what they should be. This process should include a diverse team and work with stakeholders including: developers who build and deploy AI systems, decision-makers who implement the systems in their workflows, and the community being impacted by these systems.
- A detailed plan and methodology for Steps 1-5 in the previous section of this testimony titled “What does it take to create AI systems that lead to equitable outcomes for society?”
- A continuous improvement plan to ensure that the system continues to not only be evaluated but also improved upon to achieve equitable outcomes.

---

<sup>19</sup> <http://www.datasciencepublicpolicy.org/projects/aequitas/>

**Testimony of Makada Henry-Nickie,  
Fellow, Race, Prosperity, and Inclusion  
Brookings Institution**

**Before the  
Task Force on Artificial Intelligence United States House Committee on Financial Services**

**Hearing on  
“Equitable Algorithms: Examining Ways to Reduce AI Bias in Financial Services”**

February 12<sup>th</sup>, 2020

Chairman Foster, Ranking Member Hill, and distinguished members of the Task Force on Artificial Intelligence. Thank you for the invitation and opportunity to testify before the Committee on equitable algorithms and algorithmic bias. I am Makada Henry-Nickie, Fellow at The Brookings Institution, my research covers consumer financial protection and labor market impacts of new technologies. My comments today will focus on AI-driven benefits, algorithmic bias in financial services, and algorithmic oversight, my comments are my own and do not reflect any official Brookings position. I hope my contribution furthers the Committee’s understanding of the opportunities and challenges facing consumers, regulators, financial institutions, and the scientific community as the integration of AI into financial products and services accelerates. Forward-leaning congressional leadership on the implications of artificial intelligence is critically important to ensuring that emerging technologies interact with consumers responsibly and deliver inclusive benefits to all members of our society.

**Market Interest in Artificial Intelligence is Soaring**

Artificial intelligence has permanently reshaped the financial services marketplace and altered consumers’ behaviors, preferences, and their expectations of financial institutions. Consumers, digital natives in particular, are increasingly open to engaging with their financial institutions through AI-based interfaces. According to Adobe Analytics, 44% of GenZ and 31% of millennials have interacted with a conversational interface (chatbot); surprisingly, these generational segments overwhelmingly preferred interacting with a conversational bot to a human representative.<sup>1</sup>

Shifts in consumer preferences underpin the business case for further AI investments. A two-year study of Bank of the West customers revealed that once customers opened digital bank accounts, they were nearly 60% more likely to embrace other products including, credit cards, mortgage

---

<sup>1</sup> Adobe Analytics: How Different Generations Bank (2019). Available at: <https://theblog.adobe.com/adobe-analytics-research-how-different-generations-bank/>.

refinancing, and personal loans and increase annual revenues.<sup>2</sup> While systematic evidence of the industry's return on investment is sparse, periodic surveys such as the Narrative Sciences reported that the financial services industry's investment in AI technologies grew by a remarkable 60% between 2016 and 2017.<sup>3</sup> It's clear from the sector's increasing investments in artificial intelligence that deep learning and machine learning technologies will continue to have a substantial influence on the consumer financial market.

Beyond the success of conversational interfaces, incumbent banks, nonbanks, fintechs, and insurance companies employ deep learning across a diverse array of business verticals. Nascent AI anomaly detection models have been successfully deployed to detect various types of fraud, including payments, transactions, and loans. Tech startups, such as TrueAccord and Collectly, are experimenting with machine learning software to reinvent the debt collection experience.<sup>4</sup>

AI has similarly penetrated traditional domains such as target marketing, lead generation, and credit decisioning. Though, loan underwriting algorithms, in particular, have received intense public scrutiny fueled in large part by egregious cases of algorithmic discrimination. Most recently, a software developer revealed that Apple's branded credit card, issued by Goldman Sachs, was gender-biased after Apple denied his spouse's application for a credit line increase, despite her higher credit score and similar income.<sup>5</sup> The tech company's misstep is not altogether uncommon. In March 2019, the Department of Housing Urban Development (HUD) sued Facebook for discriminatory marketing practices. HUD alleged that Facebook's allowed advertisers marketing housing services, including mortgage lenders and rental agents, to selectively curate target audiences and exclude protected groups. The tech giant's Custom Audiences and Lookalike tools allowed advertisers to explicitly exclude certain groups based on protected characteristics such as "women in the workforce" or "foreigners" in violation of the Fair Housing Act (FHA).<sup>6</sup>

Instances of algorithmic discrimination transcend financial services, as notable and disconcerting examples can be found in other domains from Amazon's biased hiring algorithm to Google Photo's image classifier that associates blacks with images of gorillas.<sup>7</sup> Discrimination or bias that systematically disadvantages minorities was a recurrent theme in each successive instance. What's clear from these illustrative cases is that artificial intelligence algorithms can adversely impact minority groups and exacerbate disparities. Consequently, it is of paramount importance that policymakers, regulators, financial institutions, and technologists critically examine the benefits, risks, and limitations of artificial intelligence and proactively design safeguards against algorithmic harm, in keeping with societal standards, expectations, and legal protections.

<sup>2</sup> Panno, Kelsey, S&P Global, "Study Finds Digital Banking Adoption Leads to More Valuable Customers." Jun. 22, 2016. Available at: <https://www.spglobal.com/en/research-insights/articles/study-finds-digital-banking-adoption-leads-to-more-valuable-customers>

<sup>3</sup> Narrative Science. (2018). Research Brief: The Rise of AI in Financial Services. Retrieved from <https://narrativescience.com/wp-content/uploads/2018/11/Research-Report-The-Rise-of-AI-in-Financial-Services-2018.pdf>

<sup>4</sup> Chin, C. (2018, September 4). Silicon Valley Wants to Use Algorithms for Debt Collection. *Wired*, Retrieved from <https://www.wired.com/story/silicon-valley-algorithms-for-debt-collection/>

<sup>5</sup> Vigdor, N. (2019, November 4). Apple Card Investigated After Gender Discrimination Complaints. *New York Times*, Retrieved from <https://www.nytimes.com/2019/11/10/business/apple-credit-card-investigation.html>

<sup>6</sup> The United States Department of Housing and Urban Development, on behalf of the Assistant Secretary for Fair Housing and Equal Opportunity v. Facebook. (March 28, 2019). Retrieved from [https://www.hud.gov/sites/dfiles/Main/documents/HUD\\_v\\_Facebook.pdf](https://www.hud.gov/sites/dfiles/Main/documents/HUD_v_Facebook.pdf)

<sup>7</sup> Madonik, R. (2018, January 11). When it Comes to Gorillas, Google Photos Remain Blind. *Wired*. Retrieved from <https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/>



### **Artificial intelligence can improve the financial lives of consumers**

The story of AI in financial institutions is not all bad; innovative fintechs have made salient contributions that make financial services more inclusive and more accessible for consumers. AI-enabled financial applications that enable consumers to accumulate savings incrementally or micro-invest are often overlooked in financial inclusion dialogues. Still, these emerging tech-enabled solutions can embody valuable financial inclusion tools. Through micro-savings apps, fintechs and legacy financial institutions have empowered millions of consumers to save more and to do so automatically. For example, Digit—an AI-enabled micro-savings app—uses machine learning algorithms to analyze checking account transactions and identify micro-savings patterns that enables its users to save consistently. Since its launch, Digit's users have saved over \$2.5 billion; the company reports that its "auto-save" algorithm empowers users to save an average of \$2,000 annually.<sup>8</sup>

In credit markets, a combination of machine learning models and high-dimensional alternative data is generating cautious interest in the ability of algorithms and Big Data to expand access to credit and decrease historical disparities. According to a 2018 study of German consumers, digital factors such as phone operating system, type of internet service or format of an email address can reliably predict loan default.<sup>9</sup> In the U.S., practical applications of alternative data do not routinely include such controversial factors. Instead, alternative data are more likely to include variables not traditionally incorporated into conventional credit underwriting models. For instance, rental payments, utility bills, or deposit transaction histories can fill critical gaps in assessing an applicant's ability to repay a loan and predict the likelihood of default.<sup>10</sup> The Consumer Financial Protection Bureau's (CFPB) publicly released findings from its No Action Letter (NAL) review of Upstart Network Inc's lending algorithm and use of alternative data. Upstart is the first CFPB-approved fintech lender sanctioned to use alternative data in its underwriting models. According to CFPB's audit, loan approval rates increased by nearly 30% for some customer segments, and the price of credit was dramatically lowered as APRs decreased, on average, between 15 and 17%. What's more, the Bureau reported that Upstart's data showed no evidence of fair lending disparities for members of protected classes.<sup>11</sup>

Additionally, findings from two recent studies show that algorithmic lending, directly and indirectly, expands access to credit. On the one hand, algorithmic peer-to-peer (P2P) lending indirectly enhanced access to credit by creating positive credit report signals resulting in increased loans from traditional banks to P2P borrowers.<sup>12</sup> Meanwhile, a UC Berkeley study found that algorithmic lending substantially decreased pricing disparities and eliminated underwriting discrimination for African-American and Hispanic borrowers.<sup>13</sup> Crucially, these emerging research

<sup>8</sup> Celebrating Digit's Impact and Funding. (2019, September 30). Retrieved from <https://blog.digit.co/>

<sup>9</sup> On the Rise of FinTechs – Credit Scoring using Digital Footprints Tobias Berg, Valentin Burg, Ana Gombovi, and Manju Puri NBER Working Paper No. 24551 April 2018, Revised July 2018 JEL No. D12,G20,O33

<sup>10</sup> FinReg Lab (2019) The Use of Cash-Flow Data in Underwriting Credit. Available at: [https://finreglab.org/wp-content/uploads/2019/07/FRL\\_Research-Report\\_Final.pdf](https://finreglab.org/wp-content/uploads/2019/07/FRL_Research-Report_Final.pdf)

<sup>11</sup> Patrice Ficklin and Paul Watkins, *An Update on Credit Access and the Bureau's First No-Action Letter*, Consumer Financial Protection Bureau (Aug. 6, 2019). Available at: <https://www.consumerfinance.gov/about-us/blog/update-credit-access-and-no-action-letter/>

<sup>12</sup> Balyuk, Tetayana (2019) Financial Innovation and Borrowers: Evidence from Peer-to-Peer Lending. Available at: <https://www.fdic.gov/bank/analytical/fintech/papers/balyuk-paper.pdf>

<sup>13</sup> Bartlett, Robert et. al (2019) Consumer Lending Discrimination in the Fintech Era. Available at: <https://faculty.haas.berkeley.edu/morse/research/papers/discrim.pdf>

studies show that despite the risks, algorithmic models have the potential to provide benefits to consumers.

#### **Algorithms propagate bias**

While artificial intelligence has the potential to deliver tremendous benefits that improve consumers' financial lives, algorithmic decision-making is inherently risky and susceptible to bias. The question of bias is less than straightforward and can be frustratingly complex and difficult to disentangle. Bias from a technical perspective affects an algorithm's accuracy or ability to make correct predictions about the real world based on its experience with training or example data.<sup>14</sup> But the definition, though precise, does not capture an intuitive, societal interpretation of the biased exclusions and unfair outcomes described in news stories about Google, Facebook, and Amazon that or discriminatory conduct that violates civil rights and equal opportunity statutes.

Tracing the sources and transmission mechanisms of bias is critical to informing the design of technological and policy solutions to reduce biased outcomes. Machine learning research has established an unambiguous link between biased outcomes and flawed training data. Class imbalance bias stemming from minority underrepresentation or selecting sample data with distorted distributions, such as systematic discrimination, can introduce selection bias into the modeling process.<sup>15</sup> These sources of bias might explain why the UC Berkeley study found that while algorithmic lenders did not discriminate against minority applicants in their underwriting decisions, they systematically charged them higher interest rates. This result is inconsistent with CFPB's NAL fair lending conclusions that algorithmic lending was associated with more equitable pricing. On the contrary, the Berkeley study confirmed that algorithmic lending perpetuated discriminatory pricing practices—Hispanic and African American borrowers paid 5.3 basis points more in interest than their white counterparts.<sup>16</sup>

In the final analysis, machine learning algorithms capable of producing equitable outcomes were not sophisticated enough break the logically flawed statistical correlation between race and credit denials or supplant the biased effects of decades of explicit racist housing policies. Algorithmic bias has tangible opportunity costs, by the researchers' estimation, minority borrowers pay an estimated \$765.0 million in excess interest payments annually, instead of saving or paying down student loan debt.

Machine learning bias is fluid and can shift in response to changes in underlying data or design processes and hence requires a flexible and vigilant ecosystem of safeguards to ensure that artificial intelligence delivers on its full potential. CFPB's declaration of Upstart's success as a potential equitable algorithm should not be regarded as an endorsement of a bias-free endeavor. Dataset shifts from non-stationary data distributions or changes in a neural network's activation function can potentially bias an algorithm over time. This hypothetical outcome is not entirely implausible, experimental alternative data such as educational background variables are early in their deployment and need critical ground-truth datasets to benchmark accuracy. Public education datasets have documented coverage gaps in certain variables such as college major.

---

<sup>14</sup> Jindong Gu: "Understanding Bias in Machine Learning", 2019, 1st Workshop on Visualization for AI Explainability in 2018 IEEE Vis; [<http://arxiv.org/abs/1909.01866> arXiv:1909.01866].

<sup>15</sup> Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rodriguez: "Fairness Constraints: Mechanisms for Fair Classification", 2015; [<http://arxiv.org/abs/1507.05259> arXiv:1507.05259]

<sup>16</sup> Ibid



Machine learning bias is neither inevitable nor final. And algorithmic bias is not benign. Algorithmic decision-making has enormous systemic social and economic consequences for affected racial, gender, and sexual minorities; these effects should not be ignored or trivialized.

**Algorithmic Oversight is Indispensable for Consumer Financial Protection.**

AI-enabled financial technologies are relatively nascent and primarily involve weak or narrow forms of AI. However, financial institutions are increasingly experimenting with advanced deep learning neural networks that are second to none in fitting high volumes of data to extraordinarily accurate predictive functions. Lamentably, deep learning's opaque "Black Box" effect even challenges AI experts when asked two fundamental questions: How? And. Why? These questions are central to human intuition and our cognitive ability to understand and negotiate our environment.

Beyond philosophical musings, our regulatory system rests firmly on a framework that assesses accountability through a causal lens, to which the answers to the questions of *how* and *why* are crucial for the system to function and effectively serve and protect American consumers. In a causal system, explanations have semantic significance and assist in making connections between reckless judgments or honest mistakes and unfair outcomes. Regulators need to be clear-eyed about an institutional agent's intent to assess the extent of its liability; without clear, rational explanations and clear causal connections between discriminatory outcomes and decision processes, the accountability framework becomes unstable and dysfunctional.

Understandably, AI's black-box effect underpins a growing chorus of calls for intuitive AI explanations between model correlations and biased outcomes. However, explainable AI is not equivalent to the type of transparency we need to redress harms caused by algorithms or identify positive lessons to inform the development of equitable algorithms. Achieving an unbiased and impartial algorithm is improbable because machine learning forces the system designer to choose a tolerable balance, based on her preferences or optimization goals.

A systemic solution that mitigates the harms of biased algorithms continues to escape the legions of AI researchers are aggressively exploring technical solutions to the challenge. Instead, Congress should focus on strengthening, maintaining, and growing the resiliency of the federal consumer oversight framework. Specifically, this task force should take action to strengthen and improve the model governance architecture. Recently, the Government Accountability Office (GAO) concluded that SR 11-7 a mission-critical model risk management framework is subject to review under the CRA.<sup>17</sup> While the effect of GAO's opinion is not immediately apparent, CFPB's precedent makes it clear that a critical regulatory gap will emerge and potentially weaken regulators' capacity to supervise financial institutions adequately. The task force should encourage CFPB to develop a parallel consumer-focused model governance framework, in light of the proliferation of algorithmic decision-making and marketing tools. Finally, the taskforce should vigilantly monitor the progress of HUD's proposed rule changes to amend the disparate impact standard.<sup>18</sup> The proposed rule

<sup>17</sup> GAO (2019, October 24) opinion on the "Board of Governors of the Federal Reserve System—Applicability of the Congressional Review Act to Supervision and Regulation Letter 11-7", Retrieved from <https://www.gao.gov/assets/710/702190.pdf>

<sup>18</sup> Housing and Urban Development Department Proposed Rule, "HUD's Implementation of the Fair Housing Act's Disparate Impact Standard" Aug. 19, 2019 <https://www.federalregister.gov/documents/2019/08/19/2019-17542/huds-implementation-of-the-fair-housing-acts-disparate-impact-standard>

introduced five new criteria for establishing disparate impact burdens that, in principle, serve to provide a safe harbor to institutions using algorithms to exploit vulnerable consumers and exacerbate historical disparities.

**Testimony of Prof. Michael Kearns  
House Financial Services Committee  
Task Force on Artificial Intelligence  
February 12, 2020**

My name is Michael Kearns, and I am a professor in the Computer and Information Science Department at the University of Pennsylvania. I hold a PhD in computer science from Harvard University, and for more than three decades my research has focused on machine learning and related topics. I have consulted extensively in the technology and finance sectors, including on legal and regulatory matters. I discuss the topics in these remarks at greater length in the recent book *The Ethical Algorithm: The Science of Socially Aware Algorithm Design* [1].

The use of machine learning for algorithmic decision-making has become ubiquitous in the finance industry and beyond. It is applied in consequential decisions for individual consumers (such as lending or credit scoring), in the optimization of electronic trading algorithms at large brokerages, and in making forecasts of directional movement or volatility in markets and individual assets. With major exchanges now being almost entirely electronic, and with the speed and convenience of the consumer Internet, the benefits of being able to leverage large-scale, fine-grained historical data sets via machine learning have become apparent.

The dangers and harms of machine learning have also recently alarmed both scientists and the general public. These include violations of fairness (such as racial or gender discrimination in lending or credit decisions) and privacy (such as leaks of sensitive personal information). It is important to realize that these harms are generally not the result of human malfeasance, such as racist or incompetent software developers. Rather, they are the unintended consequences of the very scientific principles underlying machine learning.

Machine learning proceeds by fitting a statistical model to a training data set. In a consumer lending application, such a data set might contain demographic and financial information derived from past loan applicants, along with the outcomes of granted loans. Machine learning is applied to find a model that can predict loan default probabilities from the properties of applicants, and to make lending decisions accordingly. Because the usual goal or objective is exclusively the accuracy of the model, discriminatory behavior can be inadvertently introduced. For example, if the most accurate model overall has a significantly higher false rejection rate on black applicants than on white applicants, the standard methodology of machine learning will indeed incorporate this bias. Minority groups often bear the brunt of such discrimination since by definition they are less represented in the training data.

Note that such biases routinely occur even if the training data itself is collected in an unbiased fashion, which is rarely the case. Truly unbiased data collection requires a period of what is known as *exploration* in machine learning, which is rarely applied in practice because it involves (for instance) granting loans randomly, without regard for the properties of applicants. When the training data is already biased, and the basic principles of machine learning can amplify

such biases or introduce new ones, we should expect discriminatory behavior of various kinds to be the norm and not the exception.

Fortunately, there is help on the horizon. There is now a large community of machine learning researchers who explicitly seek to modify the classical principles of machine learning in a way that avoids or reduces sources of discriminatory behavior. For instance, rather than simply finding the model that maximizes predictive accuracy, we can add the constraint that our model must not have significantly different false rejection rates across different racial groups. This constraint can be seen as forcing a balance between accuracy and a particular notion of algorithmic fairness. The modified methodology generally requires us to specify what groups or attributes we wish to protect (such as racial or gender), and what harms we wish to protect them from (such as high false rejection rates). These choices will always be specific to the context under consideration, and should be made by key stakeholders. The algorithms required to enforce fairness constraints are often more complex than the standard ones of machine learning, but not excessively so.

There are some important caveats to this agenda. First of all, there are “bad” definitions of fairness that should be avoided. One example is forbidding the use of race in lending decisions in the hope that it will prevent racial discrimination. It doesn’t, largely because there are so many other variables strongly correlated with race that machine learning can discover as proxies. Even worse, one can show simple examples where such restrictions will in fact harm the very group we sought to protect [1]. Unfortunately, to the extent that consumer finance law incorporates fairness considerations, they are usually of this flawed form that restricts model inputs. It is usually far better to explicitly constrain the model’s output behavior (as in the example of equalizing false rejection rates in lending).

It is also inevitable that constraining models to be fair will cause them to be less accurate, because we are specifying additional conditions to be met beyond just accuracy. Such trade-offs can and should be made quantitative --- for instance, by varying how much disparity we allow in false rejection rates across racial groups (from 0 percent disparity to 100 percent disparity), we can trace out the numerical curve of accuracies that can be achieved for each disparity. This is as far as science can take us --- again, stakeholders must decide what is the right accuracy-fairness balance. We must also be cognizant of the fact that different notions of fairness may be in competition with each other as well. For example, it is entirely possible that by asking for more fairness by race, we must suffer less fairness by gender. These are painful but unavoidable scientific truths.

I note in closing that while my remarks have focused on the potential for designing algorithms that are better behaved, they also point the way to regulatory reform, since most notions of algorithmic fairness (as well as other social norms such as privacy) can also be algorithmically audited. If we are concerned over false rejection disparities by race, we can systematically test models for such behaviors and measure the violations. I believe that the consideration of such algorithmic regulatory mechanisms is both timely and necessary, and I have elaborated on this in other recent writings [1,2].

Citations and Further Reading:

[1] *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. Michael Kearns and Aaron Roth. Oxford University Press, 2019.

[2] *Ethical Algorithm Design Should Guide Technology Regulation*. Michael Kearns and Aaron Roth. Brookings Institution Policy Brief, 2020. Available at <https://www.brookings.edu/research/ethical-algorithm-design-should-guide-technology-regulation/>

**Testimony to the House Committee on Financial Services Task Force on Artificial Intelligence**

**Hearing: “Equitable Algorithms: Examining Ways to Reduce AI Bias in Financial Services”**

**February 12, 2020**

**Submitted by Dr. Philip S. Thomas**

**Assistant Professor, University of Massachusetts Amherst**

Chairman Foster, Ranking Member Loudermilk, and members of this task force, thank you for the opportunity to testify today.

I am Philip Thomas, an assistant professor at the University of Massachusetts Amherst. My goal as a machine learning researcher is to ensure that machine learning algorithms are safe and fair – properties that may be critical for the responsible use of AI in finance.

Towards this goal, in a recent *Science* paper, my co-authors and I proposed a new type of machine learning algorithm, which we call a *Seldonian* algorithm. Seldonian algorithms make it easier for the people using AI to ensure that the systems they create are safe and fair. We have shown how Seldonian algorithms can avoid unfair behavior when applied to a variety of applications including optimizing online tutorials to improve student performance, influencing criminal sentencing, and deciding which loan applications should be approved.

While our work with loan application data may appear most relevant to this task force, that work was in a subfield of machine learning called *contextual bandits*. The added complexity of the contextual bandit setting would not benefit this discussion, and so I will instead focus on an example in a more common and straightforward setting called *regression*. In this example, we used entrance exam scores to predict what the GPAs of new university applicants would be if they were accepted. This GPA prediction problem resembles many problems in finance, for example rating applications for a job or loan. The fairness issues that I will discuss are the same across all these applications.

In the GPA prediction study, we found that three standard machine learning algorithms over-predicted the GPAs of male applicants on average and under-predicted the GPAs of female applicants on average, with a total bias of around 0.3 GPA points in favor of male applicants. A Seldonian algorithm successfully limited this bias to below 0.05 GPA points with only a small reduction in predictive accuracy.

The rapidly growing community of machine learning researchers studying issues related to fairness has produced many similar AI systems that can effectively preclude a variety of types of unfair behavior across a variety of applications. With the development of these fair algorithms, machine learning is reaching the point where it can be applied responsibly to financial applications, including influencing hiring and loan approval decisions.

I will now discuss technical issues related to ensuring the fairness of algorithms, which might inform future regulations aimed at ensuring the responsible use of AI in finance. First, there are many definitions of fairness. Consider our GPA-prediction example:

- One definition of fairness requires the average predictions to be the same for each gender. Under this definition, a system that tends to predict a lower GPA if you are of a particular gender would be deemed unfair.
- Another definition requires the average error of predictions to be the same for each gender. Under this definition, a system that tends to over-predict GPAs for one gender and under predict for another would be deemed unfair.

Although both of these might appear to be desirable requirements for a fair system, for this problem it is not possible to satisfy both simultaneously. Any system, human or machine, that produces the same average prediction for each gender necessarily over-predicts more for one gender, and vice versa. The machine learning community has generated more than twenty possible definitions of fairness, many of which are known to be incompatible in this way.

In any effort to regulate the use of machine learning to ensure fairness, a critical first step is to define precisely what fairness means. This may require recognizing that certain behaviors that appear to be unfair may necessarily be permissible, in order to enable enforcement of a conflicting and more appropriate notion of fairness. Although the task of selecting the appropriate definition of fairness should likely fall to regulators and social scientists, machine learning researchers can inform this decision by providing guidance with regard to which definitions are possible to enforce simultaneously, what unexpected behavior might result from a particular definition of fairness, and how much or little different definitions of fairness might impact profitability.

Regulations could also protect companies. Fintech companies that make every attempt to be fair, using AI systems that satisfy a reasonable definition of fairness, may still be accused of racist or sexist behavior for failing to enforce a conflicting definition of fairness. Regulation could protect these companies by providing an agreed-upon, appropriate, and satisfiable definition of what it means for their systems to be fair.

Once a definition of fairness has been selected, machine learning researchers can work on developing algorithms that will enforce the chosen definition. For example, our latest Seldonian algorithms are already compatible with an extremely broad class of fairness definitions and might be immediately applicable. Still, there is no “silver bullet” algorithm for remedying bias and discrimination in AI. The creation of fair AI systems may require use-specific considerations across the entire AI pipeline, from the initial collection of data through to monitoring the final deployed system.

Another observation that might inform efforts at regulation is that, for many reasonable definitions of fairness, it is not possible to ensure with certainty that any system, human or

machine, is fair. Any data used to evaluate the fairness of a system might not be representative of the actual population that the system will be applied to in the future. So, a system that appears to be fair based on the available data may not actually be fair. However, as we obtain more data, we can become increasingly confident that the data resembles the larger population, and hence that the system will be fair when used. In this way, when fairness cannot be guaranteed with certainty, it can usually be guaranteed with high probability. While this motivated my research into creating systems that are safe and fair with high probability, this observation might also inform how AI systems are regulated. Requiring companies using AI to ensure that their systems are fair with certainty may be asking the impossible. Hence, one might regulate the process rather than the outcome – to require the use of algorithms that are fair with high probability and the use of mechanisms to quickly identify and repair unfair behavior when it inevitably occurs.

Several other questions must be answered for regulations to be effective and fair. For example: Will fairness requirements that appear reasonable in the short-term have the long-term impact of reinforcing existing social inequalities? How should fairness requirements account for the fact that changing demographics can result in a system that was fair last month being unfair today? When unfair behavior occurs, how can regulators determine whether this is due to the aforementioned inevitability of unfair behavior, or the improper use of machine learning?

Thank you again for the opportunity to testify today. I look forward to your questions.



**Bari A. Williams – Proposed Testimony of use of Artificial Intelligence in Financial Services**

To Chairwoman Maxine Waters, The Task Force on Artificial Intelligence of the House Financial Services Committee:

February 12, 2020

I am Bari A. Williams, an attorney and startup advisor, born and raised, and still live in Oakland, CA, working in technology transactions, with a focus on artificial intelligence (“AI”), privacy, and commercial contracts. My educational background includes a BA from UC Berkeley, an MBA from St. Mary’s College of California, and a Masters in African-American Studies from UCLA. In my career, I’ve worked for Facebook, Stubhub, and All Turtles, which is an AI startup studio, akin to an incubator.

In my work in the tech sector, I’ve been exposed to many interesting use cases for technology that provide convenience, efficiency, and optimization to our lives. But one nagging question always lingers – at whose expense are these gains made, and how do we solve for the negative impacts of some of the most pervasive uses of technology?

AI provides a unique example of this. To begin – what is AI? It is essentially someone’s bias, via datasets, baked into code that can determine the ads one sees, one’s credit worthiness, employment prospects, school admissions, housing opportunities, and criminal justice implications (i.e. facial recognition technology, gunshot locaters such as ShotSpotter, predictive policing such as Hunchlab, and predictive sentencing technology).

- (1) How are data sets, proprietary algorithms, and models are deployed and used within financial services, and what are ways to improve their deployment in financial services?

There are five main issues with AI, particularly in financial services: (1) what data sets are being used – who fact checks the fact checkers; (2) what hypotheses are set out to proven using this data – has the narrative that is being written been adequately vetted; (3) how inclusive is the team creating and testing the product – who are you building products with; (4) what conclusions are drawn from the pattern recognition and data that the AI provides – who are you building products for, and who may be harmed or receive benefit, and; (5) how do we ensure bias neutrality, and what is the benefit of neutrality.

Data sets in financial services are used to determine home ownership and mortgage, savings and student loan rates; the outcomes of credit card and loan applications; credit scores and credit worthiness, and insurance policy terms. It affects other outcomes, such as credit card fraud prediction. The danger in this is a repeat of redlining, the discriminatory practice of ensuring Black homeowners were confined to specific areas of a city and that their credit worthiness led to higher interest rates. The problem is that the data sets that are being used by these companies are “stale,” meaning they are dated and old. The older data has these remnants of credit worthiness during redlining, including income earned (which is already a huge disparity for people of color), and additional debt incurred.

The largest segment of AI use in financial services is anti-fraud. We see this not just in banks, but in any company that deals in consumer transactions, like StubHub, for instance. There are tech companies that make this software, such as Sift, to identify potential fraud risks. The problem is that if you are using a stale data set that skewed in favor of a certain demographic for “fraud potential,” it is already flawed, and the pattern recognition will be a self-fulfilling prophecy. Additionally, data is which is already coded based on someone’s personal bias, as data by itself doesn’t discern any conclusions. Data sets are often \*chosen\* to support a specific hypothesis, not to be neutral. This isn’t just a notion. In 2017, per [data from the Home Mortgage Disclosure Act](#) showed that 19% of Black borrowers and almost 14% of Latinx/Hispanic borrowers were turned down for a conventional loan.

There are several ways to improve the deployment of this technology in financial services. It starts with companies owning their power in implementing these technologies, and being deliberate about auditing the systems. Companies must proactively look for and identify bias in their AI by asking themselves these five questions:

1. Ensure all data groups have equal probability of being assigned to favorable outcomes.
2. Ensure all groups of a protected class have equal positive predictive value.
3. Ensure all groups of a protected class have predictive equality for false positive and false negative rates.
4. Maintain equalized odds ratio, opportunity ratio and treatment equality.
5. Minimize average odds difference and error rate difference.

Additionally, two techniques that can also drive fair outcomes include leveraging statistical techniques to resample or reweigh data to reduce bias, which is like a visual of giving someone a box to stand on if they are short to make them the same height and have the vantage point of someone privileged with more height. The second technique includes adding a “fairness regulator,” which is a mathematical constraint to ensure fairness in the model, to existing algorithms. The fairness regulator is akin to my fifth question about what does it mean to be bias natural. It is important to remember that not all biases are bad – some are actually beneficial and seek to right wrongs. There may be ways to award additional “points” or level the playing field with historically discriminated against marginalized groups to ensure parity with interest rates, financial advice, and credit assessment.

That said, nothing beats having diverse teams make and test these algorithms prior to use in the market. That will help ensure data sets are also diverse, and there is no disparate impact when testing. Lack of diversity in tech is an ethical issue, not just one about ‘doing the right thing.’

## (2) What emerging methods or ways AI can be used to decrease discrimination and bias?

There are elements of AI that can be used for good. The concepts of transparency and fairness are not mutually exclusive, but to the contrary, are closely related. One solution uses mathematical methods in two separate products that provide explanations – the ability to identify what’s driving the disparity, and fairness – thus, provides the ability to reduce the disparity.

Some emerging methods that AI is used for productive outcomes includes AI that actually identifies bias via identifying pattern recognition with disparate impact. An example of this is seen with Zest, a tech financial services company, which has created a product, ZAML Fair, that reduces bias in credit assessment by ranking an algorithm's credit variables by how much they lead to biased outcomes, and then muffles the influence of those variables to produce a better model with less biased outcomes.

If more banks, and even consumer facing companies that use credit (i.e. retailers, credit reporting bureaus, etc.) would utilize a tool like this, we would see greater impartiality in financial decision making, which would save consumers billions of dollars, and may actually benefit companies by demonstrating efforts to be equitable, thus encouraging more business with consumers.

- (3) What ways existing laws and regulations can be applied to provide more transparency while still preserving data privacy and maintaining strong cybersecurity standards?

As I tell my clients, and my kids... a rule is only as good as its enforcement. To that end, it is imperative that the government use the laws already enacted to create greater parity and transparency.

The Fair Housing Act can be applied to ensure more fairness in the use of AI in financial services. For example, if a [mortgage lending model](#) finds that older individuals have a higher likelihood of defaulting on their loans then decides to reduce lending based on age, there is a legal claim for this to constitute illegal age discrimination and housing discrimination.

Additionally, greater enforcement of discrimination based on disparate impact on the basis of any protected class is illegal under the US Equal Credit Opportunity Act of 1974. The excuse that "the model did it, not a human" won't work, when humans are coding the models. Per a [2018 study conducted at UC Berkeley](#) found that both traditional face-to-face loan decisions and those made by machine learning systems charged Latinx/Hispanic and Black borrowers interest rates that were 6-9 basis points higher. Lending discrimination costs these US minority borrowers \$250-\$500 million per year in extra mortgage interest. The study concluded that algorithms have not removed discrimination, but may have shifted the mode, and also made it more efficient.

Using greater enforcement of the laws on the books, calling for transparency into the data sets used to train these algorithms, as well as understanding the technique utilized by any human involvement in decision making after assessing an AI output could be very effective.

- (4) Are there any regulatory and legislative proposals to strengthen federal oversight of algorithmic decision-making and AI technologies utilized by financial institutions?

In addition to the suggestion of greater enforcement of the US Equal Credit Opportunity and The Fair Housing Act, I have submitted, along with this written testimony, an attachment as Exhibit

A, that is a proposed “AI Bill of Rights,” which would be guidelines that companies must meet in order to deploy this technology. No longer should it be “ship it fast,” which is a Silicon Valley ethos to get a product to market as quickly as possible, oftentimes to usurp a competitor, but instead the industry should adopt the medical ethos of “do no harm.” The premise of my recommendations is attached in Exhibit A, but of note, they include guidelines for requiring transparency and auditing of data sets being available to consumers and/or the govt. to discern equitable inclusion for unbiased results, and for all terms of service and information on data collected and its use be written in *plain English*, not legalese.

Additionally, the larger problem has been that technology is constantly iterating and improvising and improving, while law has not kept pace at the same rate. To that extent, the expertise needed to understand the tech production process is likely not going to happen in government, but perhaps creation of a hybrid model where there’s an institution that enables expertise to be cultivated, while also understanding the process of what it takes to turn a proposed bill into law. So, when companies are planning to implement AI systems, it’s not just “ship it fast, and we’ll see what happens and fix it on the backend.” You actually have to verify the claims of your system, there is transparency around data sets used, and a company can answer five key questions: (1) what are you building and how, (2) what information are you using as the foundation of your system, (3) who are you building it for, (4) who are you building it with (diversity in tech and testing matters), and (5) is there any disparate impact when testing your system. Similar to the FDA model, it’s acknowledged that not all drugs work for everybody. Not all technology is inclusive. One size does NOT fit all, so there are limitations.

The impact of technology has been a gift and a curse. The advent of this fourth industrial revolution has brought great convenience, but at a great expense – privacy, data collection and use, and less human interaction at the behest of automated decision making. In our quest to provide greater efficiency and convenience, we have been lax to look at who is left behind and how. If we aren’t careful, we will automate greater discrimination into the tools that we use everyday, and further exacerbate the legacy of lack those in marginalized communities. I implore the committee to do a deeper dive into how they can both enforce and enhance the US Equal Credit Opportunity and The Fair Housing Act, in addition to adopting the AI Bill of Rights as attached as Exhibit A.

Thank you for the opportunity to speak with the Committee.

### **Exhibit A – Proposed AI/Technology Bill of Rights**

1. Make all terms of service and privacy policies in plain English (or whatever the applicable language may be). Even better if they can be written in a Q&A format.

The rationale for this rule is the ability to easily read and understand how information is collected, why, and how it will be used, and any ability for a consumer to delete or opt out of said collection.

2. Transparency and auditing of data sets should be made available to consumers and/or the govt. to discern equitable inclusion for unbiased results.

The rationale for this rule is to ensure that data sets are vetted for accuracy, any bias, or to make suggestions for other data sets that may complement or correct for errors in the data sets used to train the AI algorithms. It is important to do this to ensure there is fairness from the start.

3. No product should be embedded/integrated into products without testing for inclusion (i.e. differently abled, LGBTQ, employment/housing decisions, rural/city implications, economic ramifications) and no disparate impact. If a product shows a disparate impact upon a marginalized population within [TBD%] range, the product should not be sent to market.

The rationale for this rule is to ensure that products are not rushed to market without testing for adverse effects on one group of people of a protected characteristic more than another. It ensures parity of the ability of usage of the product, and that there will not be negative impact on already marginalized groups.

4. To ensure that AI products do not have a disparate impact on marginalized populations, companies shall do beta testing with people from marginalized groups, though also ensuring their privacy and protection v. limited data collection, anonymized data, and encryption at any and all points in the process where available.

The rationale for this rule is to ensure that testing is done with marginalized communities before a product is shipped. It is a complimentary rule to #2. Additionally, this supports more diversity in tech. If a company does not have sufficient employee population to engage in this testing, it encourages outsourcing to diverse suppliers who can assist with this beta testing with focus group organization, or doing it in in-house. This solves for both the company testing issue, and supports more diversity in tech, which produces an ethical and inclusive product.

5. AI For Good - The ability to "opt-in" to data collection, and sufficient notification (again, in plain English) before data is collected and used. Again, this information should be written in plain English.

The rationale for this rule is to ensure that consumers and individuals have the ability to control their own information. Currently, people are passively giving away what would be deemed “proprietary information” about themselves, and this affords the ability to control what is learned about a person and how it is applied. This is especially important regarding financial services when credit reports and scores are part of an employment applications and background checks.

6. If there are certain devices that an app needs access to, which wouldn't be intuitive (i.e. camera and gallery access for a food delivery app), a plain English explanation of why this access is necessary, and an example of how it will be used shall be provided.  
Example: "We need access to your camera/contacts to facilitate \_\_\_\_\_, and it will be used for [XX] duration in the following ways: \_\_\_\_\_."

The rationale for this rule is to give people all the information they need to make an informed decision before they decide to download and/or use an app. Oftentimes, people do not read the required permissions necessary to use an app, and instead just “scroll and accept.” Unfortunately, that means that an app may have access to features and attributes of your phone that it doesn't actually need to effectuate the service, but instead to just surveil you and collect information. By having these notifications written in plain English, it becomes clear what is needed, and what isn't, and a consumer can make a more informed decision about using that app, another, or none at all.

7. Optionality of Features - Provide the ability for consumers to opt-in to SOME features of an app, but not all. As the previous example notes, if we do not want to provide access to our cameras, contacts, or some requested access, though not necessary to deliver services consumer wants, then the app may still work, but will not provide full functionality possible had all permissions been given.

\*At the moment, a consumer's choice is to provide all access requested, or not have use of the app. Depending on model of phone, alternative options, etc., this could demonstrate disparate impact.

The rationale for this rule is to ensure that consumers have choice when using an app, and can decide what permissions or access the app will have to their phone and its features. Currently, consumers do not have that option, and it means in order to use an app, one often has to consent to overbroad uses and access to features of a phone, such as the camera, contacts, and voice data. These access rights are often unnecessary to effectuate services, but are just a means of data collection and to surveil a consumer, often used to market new services, sold to another company or data broker, or to assist with building a new product. Customers shouldn't help create wealth and IP for a company without their explicit knowledge and consent, not just for convenience and lack of knowledge.

8. Data Portability and Right to be Forgotten - Consumers should have the right to retrieve, correct, or delete personal data controlled by any company that has access to such data (with correct permissions and rules around certain categories of data - i.e. medical records in emergency, criminal records that are not expunged, etc.). Along with this right,

much like GDPR, if a consumer decides to delete all of their records from an app or platform, that should ensure their information is wiped clean.

\* This is especially true if algorithms are being created to look at credit worthiness, if there are insurance or medical decisions being made off of erroneous or dated historical data, etc.

The rationale for this rule is akin to the California Consumer Privacy Act, which affords a consumer the ability to audit and remove their data, or to opt-out of certain features. This would be very helpful when dealing with financial services, should excessive credit show up on a report which is requested for employment, or information that may bias decision making, such as rental history and locations. By giving a consumer power over their data collection and use, it restores trust in the companies that consumers decide to do business with, as they have the information needed to make an informed decision, but not enough to negatively impact decision making.

9. Affirmative Action for AI - After the 2008 housing crisis, 44% of Black Americans have a credit score below 650, after being steered to sub-prime loans. That said, some functions of AI that are currently used as using historical data that has bias baked into the code (i.e. housing data that incorporates redlining features, or predictive policing that includes historical crime data that disproportionately target Black and brown people, primarily in low-income areas as seen in Ferguson and San Francisco) should find a way to seek equity and parity in analysis of marginalized groups to not further negatively impact them.

The rationale for this rule is to correct for past wrongs, and to ensure greater equity. AI has the ability to build in extra scripts to award additional inputs and "points" to people of certain profiles for parity when looking at financial services, health disparities, or housing. This should be considered when making enterprise to consumer-based AI technology that could have far reaching, lasting effects on an entire community's wealth prospects.

The pace of law lags that of technology, as the latter drives innovation and the former waits to see the results before passing legislation or creating policies. If companies are forward-thinking in their application of predictive analytics, AI, and machine learning they can make these technologies inclusive without the need for new laws or regulations.





February 12, 2020

Dear Chairman Foster and Ranking Member Loudermilk,

I am writing on behalf of BSA | The Software Alliance to thank you for leading the Task Force on Artificial Intelligence and convening today's critically important hearing. BSA is an association of the world's leading enterprise software companies that provide businesses in every sector of the economy with tools to operate more competitively and innovate more responsibly.<sup>1</sup> BSA members are at the forefront of developing AI-enabled solutions that empower their customers to transform raw data into actionable intelligence.

While the benefits of AI will reverberate throughout the economy, it will have a particularly profound impact on data-intensive industries, such as the financial services sector. As this Task Force's hearings have demonstrated, AI is already being deployed across the financial services industry in ways that help consumers, including to improve the accuracy of financial forecasting, to reduce the risk of fraudulent transactions, and to deliver a more personalized customer relations experience. Of course, as AI is integrated into business processes that could impact the public's access to housing and credit, this Task Force (and the House Financial Services Committee more generally) has an important oversight role to play. BSA stands ready to assist you in that effort.

BSA recognizes that public trust is an essential component of a thriving digital economy. The focus of today's hearing – examining how to reduce the risk of AI bias – is one critical element of ensuring the public's trust and confidence in AI. That trust depends on ensuring that existing protections for consumers will not be undercut by the use of AI. Simply put, existing laws should apply to the use of new technologies, and decisions that would otherwise be unlawful should not avoid liability simply because they may now involve the use of an AI system.

We have already seen government agencies grappling with how existing laws apply to new technologies like AI—in ways that may undermine confidence in AI technologies. The clearest example is the recent proposal by the Department of Housing and Urban Development to create a safe harbor for defendant's who use AI systems to make lending decisions that result in a disparate impact. As detailed in our attached submission to HUD, we have significant concerns that the proposal could discourage institutions from closely monitoring their own use of AI systems for unintended impacts. As a result, the proposed rule could exacerbate the risk of bias and thereby undermine public trust in AI. HUD's proposal is

---

<sup>1</sup> BSA's members include: Adobe, Atlassian, Autodesk, Bentley Systems, Box, Cadence, CNC/Mastercam, IBM, Informatica, Intel, MathWorks, Microsoft, Okta, Oracle, PTC, Salesforce, ServiceNow, Siemens Industry Software Inc., Sitecore, Slack, Splunk, Trend Micro, Trimble Solutions Corporation, Twilio, and Workday.



Chairman Foster and Ranking Member Loudermilk  
 February 12, 2020  
 Page 2

the first intervention by a US government agency to define how civil rights protections will apply to the use of AI, but it will not be the last. This Task Force can play an important role in ensuring that the precedent set by HUD is one that avoids risks for the public.

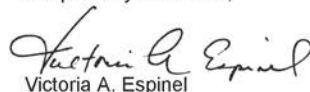
These issues also deserve more focused attention, from both government and industry. HUD's proposed safe harbors strike at the heart of one of the most active areas of AI research and will arise again as other government agencies evaluate how the use of AI will impact their missions. For these reasons, BSA has called on the National Institute of Standards and Technology ("NIST") to convene a multistakeholder process to develop an AI lifecycle risk management framework. Such a process would bring together experts from government, industry, and academia to develop of a framework for identifying and mitigating the risks of bias that can emerge as AI is designed, developed, and deployed. An AI risk management framework would be valuable not only for government agencies that are creating AI policy (including HUD), but also the companies that are developing and using AI technologies with the potential to impact the public. Enlisting NIST for this important effort would also build on NIST's successes in creating frameworks that address cybersecurity and privacy risks.

BSA also supports continued research and investment – both public and private – on ways to mitigate bias. Developing mechanisms for identify and mitigate the risks of AI bias has emerged as an area of intense focus for experts in industry, academia, and government. In just the past few years, a vast body of research has identified a range of organizational best practices, governance safeguards, and technical tools that can help manage risks of bias throughout the AI lifecycle. Such efforts are only one element of the industry's approach to addressing bias.

BSA members are committed to ensuring that their technologies enhance fairness and mitigate the potential for discrimination. In the long term, we recognize that requires systemic commitment to nurturing a diverse technological workforce – to ensure a diverse array of individuals are involved in developing, using, and deploying AI technologies. BSA and its members therefore support initiatives that empower and expand the technological workforce. To help drive those efforts, BSA launched [Software.org](https://www.software.org/), an educational foundation that highlights and directly engages in efforts to expand opportunities in computer science for girls and other underrepresented groups. One of Software.org's most exciting partnerships is with Girls Who Code. This year, Software.org and Georgetown University Law Center's Institute for Technology Law & Policy will host a summer immersion program to teach a class of young girls coding skills that will help them pursue a career in STEM. Those DC efforts are among over 75 Girls Who Code programs across the country, including classrooms sponsored by some of Software.org's supporting companies: Adobe, Autodesk, IBM, and Microsoft.

BSA appreciates the opportunity to provide these comments to the Task Force. We welcome an opportunity to further engage with you on these important issues going forward.

Respectfully submitted,

  
 Victoria A. Espinel

cc: Chairwoman Waters  
 Ranking Member McHenry



October 18, 2019

Office of General Counsel  
 Department of Housing and Urban Development  
 451 7<sup>th</sup> St. SW  
 Room 10276  
 Washington, DC 20410

Re: ***HUD's Consideration of the Fair Housing Act's Disparate Impact Standard***  
**Docket No. FR-611-P-02; RIN 2529-AA98**

Dear Assistant Secretary Farias:

BSA | The Software Alliance (BSA) is the leading advocate for the global software industry before governments and in the international marketplace.<sup>1</sup> Our members are at the forefront of software-enabled innovation that is fueling economic growth in every industry sector. As global leaders in the development of data-driven enterprise software solutions, BSA's members have a keen interest in working with policymakers to establish a legal environment that helps engender the public's trust and confidence in the technologies that are driving today's digital economy. We therefore welcome this opportunity to provide comments to the Department of Housing and Urban Development's (HUD) proposed rule concerning the interpretation of the Fair Housing Act's disparate impact standard.<sup>2</sup>

Given BSA's focus on the intersection of technology and policy, these comments focus narrowly on aspects of the Proposed Rule bearing on the use of Artificial Intelligence (AI) and the creation of potential safe harbors in circumstances where a "plaintiff identifies an offending policy or practice that relies on an algorithmic model."<sup>3</sup> We are concerned that

---

<sup>1</sup> BSA's members include: Adobe, Akamai, Apple, Autodesk, Bentley Systems, Box, Cadence, CNC/Mastercam, DataStax, DocuSign, IBM, Informatica, Intel, MathWorks, Microsoft, Okta, Oracle, PTC, Salesforce, ServiceNow, Siemens PLM Software, Sitecore, Slack, Splunk, Symantec, Trend Micro, Trimble Solutions Corporation, Twilio, and Workday.

<sup>2</sup> 84 Fed. Reg. 42854 (August 19, 2019) [hereinafter "Proposed Rule"].

<sup>3</sup> Proposed Rule at 42859.

the proposed safe harbors – as currently drafted – could undermine trust in digital technologies that increasingly are involved in high-stakes decisions that impact people’s lives.

BSA members are firmly committed to ensuring that their technologies enhance fairness and mitigate the potential for discrimination. As digital technologies are deployed in ways that implicate the public’s ability to obtain access to housing and finance, it is critical that HUD has the resources and authorities it needs to robustly enforce the Fair Housing Act’s prohibitions on discrimination. As a matter of principle, the public must be confident that the Fair Housing Act (FHA) will continue to afford the same level of protection irrespective of whether a lending or housing decision was made by a person, or a person assisted by a machine. One objective of this proceeding should therefore be to ensure that the use of technology will not hinder the enforcement of legitimate FHA claims. Simply put, existing laws should apply to the use of new technologies, and decisions that would otherwise incur liability under the FHA’s disparate impact standard should not benefit from a safe harbor merely because they involve the use of an AI system.

The use of advanced technologies in connection with housing and lending decisions presents both opportunities and risks. On the one hand, the adoption of AI by financial institutions has the potential to reduce discrimination and promote fairness by facilitating a data-driven approach to decision-making that is less vulnerable to human biases.<sup>4</sup> For instance, the use of AI can improve access to credit and housing to historically marginalized communities by enabling lenders to evaluate a greater array of data than is ordinarily accounted for in traditional credit reports. At the same time, researchers caution that flaws in the design, development and/or deployment of AI systems have the potential to perpetuate existing social biases.<sup>5</sup> Such biases can arise in a variety of ways, including circumstances in which an AI system is “trained” using data that reflects historical biases or when AI systems are deployed in populations that do not reflect the demographics of the data upon which they were trained.

Developing mechanisms for identifying and mitigating the risks of AI bias has emerged as an area of intense focus for experts in industry, academia, and government. In just the past few years, a vast body of research has identified a range of organizational best practices, governance safeguards, and technical tools that can help manage risks of bias throughout

---

<sup>4</sup> See, e.g., Jennifer Sukis, *The origins of bias and how AI may be the answer to ending its reign*, Medium (Jan. 13, 2019), <https://medium.com/design-ibm/the-origins-of-bias-and-how-ai-might-be-our-answer-to-ending-it-acc3610d6354>.

<sup>5</sup> See, e.g., Nicol Turner Lee, Paul Resnick, and Genie Barton, *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms*, Brookings (May 22, 2019), <https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/>.

the AI lifecycle. Because static evaluations of AI models cannot account for all potential issues that may arise when AI systems are deployed in the field, experts agree that mitigating risks of AI bias requires a lifecycle approach, including ongoing monitoring by end-users to ensure that the system is operating as intended.

In light of the continuing evolution of this field of research, we urge HUD to take a cautious approach as it considers potential safe harbors for disparate impact claims arising from a defendant's use of algorithmic models. We appreciate HUD's clarification that the safe harbors are "not intended to provide a special exemption for parties who use algorithmic models" and instead are aimed at providing defendants with guidance about how they "can show their models achieve 'legitimate objectives.'"<sup>6</sup> However, for the reasons outlined below, we are concerned that the proposed safe harbors may ultimately create greater uncertainty for entities that use and/or develop algorithmic systems and potentially exacerbate the risks associated with AI bias. We outline below the basis of our concerns.

#### I. The Proposed Rule's Inconsistent Use of Terminology Creates Uncertainty

HUD's explanation of the Proposed Rule describes the "first defense" (hereinafter Safe Harbor #1) and "third defense" (hereinafter Safe Harbor #3) as functionally "similar." HUD indicates that Safe Harbor #1 enables a defendant to prevail if it shows that the "model is not the actual cause of the disparate impact" through a "piece-by-piece" examination to determine whether "a factor used in the model is correlated with a protected class."<sup>7</sup> HUD likewise characterizes Safe Harbor #3 as enabling a defendant to prevail if it proves (through the use of a qualified expert) that the "model is not the actual cause of the disparate impact."

Notwithstanding HUD's characterization of these defenses as functionally "similar," the proposed text for the defenses seems to employ terminology differently:

- Safe Harbor #1 can be invoked if a defendant shows that the "material factors that make up the inputs used in the challenged model...do not rely in any material part on factors that are substitutes or close proxies for protected classes under the Fair Housing Act."<sup>8</sup>
- Safe Harbor #3 can be invoked if a neutral third party validates that "none of the factors used in the algorithm rely in any material part on factors that are substitutes or close proxies for protected classes under the Fair Housing Act."<sup>9</sup>

---

<sup>6</sup> Proposed Rule at 42859.

<sup>7</sup> Id.

<sup>8</sup> Id. at 42862 (§ 100.500 (c)(2)(i)).

<sup>9</sup> Id. (§ 100.500 (c)(2)(iii)).



To avoid confusion, HUD should clarify whether the use of different terminology in Safe Harbor #1 and Safe Harbor #3 is intentional. To the extent the inquiries under Safe Harbor #1 and Safe Harbor #3 are intended to focus on different aspects of a challenged model, HUD should provide additional guidance in the final rule.

## **II. The Proposed Rule's Focus on Individual Inputs is Both Over- and Under-Inclusive**

Safe Harbor #1 and Safe Harbor #3 appear to create a bright line rule that would excuse disparate impacts that arise from a defendant's use of an AI system that does not rely on individual inputs that are "substitutes or close proxies for protected classes under the Fair Housing Act." Although the Proposed Rule lacks specific guidance about how HUD will assess whether an input to an algorithmic model is a "substitute" or "close proxy" to a protected class, HUD notes that the defenses would be unavailable if a plaintiff is able to demonstrate "that a factor used in the model is correlated with a protected class."<sup>10</sup> Conditioning eligibility for the safe harbor on an analysis that focuses on individual inputs would result in a range of unintended outcomes.

On the one hand, such a safe harbor would be unduly narrow. As a practical matter, it could have the effect of preventing lending institutions from relying on data inputs, such as income, that bear a close nexus to creditworthiness, but which may also be correlated to protected classes. Such a safe harbor could also preclude AI systems from containing features that have the effect of mitigating potential biases. Precluding the use of variables that are correlated to protected classes could deter lenders from using AI systems that leverage such variables for the explicit purpose of *preventing* disparate impacts.<sup>11</sup> Foreclosing the use of AI models that use protected classes (or proxies thereof) for the express purpose of de-biasing the model would of course be counterintuitive to the purpose of the Proposed Rule.

On the other hand, Safe Harbors #1 and #3 would also be overly broad, potentially privileging systems that produce discriminatory results based on inputs that bear no reasonably intuitive relationship to credit risk. The focus on individual inputs misapprehends the risk that a model may rely on a *combination* of facially neutral inputs that amount to a proxy for a protected class. By focusing only on the individual inputs to a model, the safe

---

<sup>10</sup> *Id.* at 42859.

<sup>11</sup> See, e.g., Nicol Turner Lee, Paul Resnick, and Genie Barton, *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms*, Brookings (May 22, 2019), <https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/>

harbors could theoretically be invoked in circumstances where a model relies on facially neutral variables that produce extremely discriminatory outcomes. The risk is particularly pronounced if the facially neutral variables do not bear a reasonably explainable relationship to the target variable (e.g., credit risk) that the model is intended to measure.<sup>12</sup>

### III. The Proposed Rule's Reference to Industry Standards is Unclear

Safe Harbor #2 can be invoked by a defendant who uses an algorithmic model that is “produced, maintained, or distributed by a recognized third party that determines industry standards.”<sup>13</sup> We seek additional clarity about the types of “industry standards” and “recognized” third parties to which this provision refers. Under a narrow reading, Safe Harbor #2 may only apply in circumstances where an international standard-setting body, such as the International Organization for Standards, both develops a standard and then distributes an associated algorithmic model that implements the standard. Under a broader reading, HUD may be referring to widely deployed technologies produced by an individual company. Alternatively, HUD may be referring to automated underwriting systems built on algorithmic models that are produced by government-sponsored entities (e.g., Fannie Mae and Freddie Mac).

### IV. The Proposed Rule Creates Perverse Incentives that Exacerbate Risks of Bias

Safe Harbor #2 could also have the perverse effect of discouraging institutions from closely monitoring their own use of AI systems for unintended impacts. In the explanation of Safe Harbor #2, HUD suggests that liability for bias caused by algorithmic models that are “standard in the industry” should be borne by “the party that is actually responsible for the

---

<sup>12</sup> Such concerns prompted the Federal Reserve Board to issue a 2017 advisory bulletin cautioning against the use of facially neutral data inputs that do not bear a reasonably intuitive connection to creditworthiness. See Carol A. Evans, *Keeping Fintech Fair: Thinking About Fair Lending and UDAP Risks*, Consumer Compliance Outlook (Fed. Res. Sys., Phila, Pa.), 2017, <https://www.consumercomplianceoutlook.org/2017/second-issue/keeping-fintech-fair-thinking-about-fair-lending-and-udap-risks/> (“Careful analysis is particularly warranted when data may not only be correlated with race or national origin but may also closely reflect the effects of historical discrimination, such as redlining and segregation. For example, it’s been reported that some lenders consider whether a consumer’s online social network includes people with poor credit histories, which can raise concerns about discrimination against those living in disadvantaged areas. Instead of expanding access to responsible credit, the use of data correlated with race or national origin could serve to entrench or even worsen existing inequities in financial access. *Finally, it is important to consider that some data may not appear correlated with race or national origin when used alone but may be highly correlated with prohibited characteristics when evaluated in conjunction with other fields.*”) (Emphasis added.)

<sup>13</sup> Id. at 42862 (§ 100.500 (c)(2)(ii)).

creation and design of the model.”<sup>14</sup> Setting aside the uncertainty (noted above) about the type of industry standards this refers to, such a bright line rule overlooks the complexity of the AI ecosystem and threatens to establish a one-size-fits-all policy that may deter end-users from monitoring their own usage of an algorithmic model to ensure that is not creating a disparate impact. As noted above, the risk of bias must be continuously monitored because the performance of a model can be impacted if it is deployed into an environment in which the demographics differ from the data upon which it was trained. In many circumstances, only the entity that has deployed the model will be in a position to monitor its operation. However, Safe Harbor #2 could create a disincentive to perform such monitoring if doing so could increase their exposure to liability from which they would otherwise be shielded.

### Conclusion

The growing ubiquity of AI has the potential to improve the delivery of services that will impact almost every facet of our daily lives. As AI is integrated into business processes that have consequential impacts on people – such as their ability to obtain access to credit or housing – it is imperative to ensure that existing legal protections apply even as technologies evolve. The public must be confident that these protections apply regardless of whether a decision is made by a person or by a machine. The safe harbors in the Proposed Rule would undermine that confidence, create uncertainty, and ultimately exacerbate the risks associated with AI bias.

The Proposed Rule’s safe harbors constitute the first intervention by a US government agency to define how civil rights protections will apply to the use of Artificial Intelligence. The complex issues that are implicated by the safe harbors strike at the heart of one of the most active areas of AI research and will arise again as other government agencies evaluate how the use of AI will impact their missions. Accordingly, we urge HUD to be very cautious and to consider whether these issues might benefit from a coordinated interagency consultation process.

The Executive Order on Maintaining American Leadership in AI tasked the Office of Science Technology and Policy and the Office of Management with the development of guidance for the heads of all agencies that is intended to “reduce barriers to the use of AI technologies in order promote their innovative application while protecting civil liberties.” Given that this Proposed Rule bears squarely on uses of AI that implicate core civil liberties protections, we urge HUD to consult closely with OSTP and OMB before issuing a final rule. We would likewise urge HUD to consult with the National Institute of Standards and Technology about the potential for convening a multistakeholder process for the purpose of developing an AI

---

<sup>14</sup> Id. at 42859.

Page 7

lifecycle risk management framework. Such a process would enable experts from government, industry, and academia to collaborate on the development of a framework for identifying and mitigating the risks of bias that can emerge during the various phases of the AI lifecycle. The development of an AI risk management framework would be valuable not only for government agencies – such as HUD – that are developing AI policy, but also the companies that are developing and deploying AI technologies.

\* \* \* \* \*

Thank you again for the opportunity to share our views on these important issues.

Sincerely,



Christian Troncoso  
Director, Policy



**Written Statement of Brenda Leong**  
 U.S. House of Representatives  
 Committee on Financial Services, Task Force on AI  
 Rayburn House Office Building  
 Washington, D.C. 20515  
 February 12, 2020

Thank you for the opportunity to provide a written statement for the record of the hearing on Equitable Algorithms: Examining Ways to Reduce AI Bias in Financial Services with The Task Force on Artificial Intelligence. My name is Brenda Leong, and I am Senior Counsel and Director of AI and Ethics at the Future of Privacy Forum (FPF). FPF thanks the Task Force Chair and Ranking Member for convening this hearing, and for working to address the privacy and civil liberties challenges of the use of artificial intelligence and machine learning-based applications in financial services products and services, and specifically how to protect those systems from the impacts of undesired or unintended bias.

We submit this statement to:

- Observe that automated decision-making is not new in the financial services sector, and that AI-powered programs and services remain subject to the regulatory and compliance structures in place to protect consumers,
- Describe beneficial ways that financial institutions are using AI to gain efficiencies or add capabilities: to combat fraud, extend credit to traditionally underserved individuals, improve internal research and analysis and customer service functions,
- Identify several factors that can present fairness and equity concerns that are unique or heightened by processing within an AI or Machine Learning-based system, and to
- Identify the technical, policy, regulatory and legislative actions that can help mitigate risk and bias from the use of these systems.

**About Future of Privacy Forum:**

FPF is a nonprofit organization that serves as a catalyst for privacy leadership and scholarship, advancing principled data practices in support of emerging technologies. We believe that the power of information technology is a net benefit to society, and that it can be well-managed to control risks and offer the best protections and empowerment to consumers and individuals.

FPF has a substantial portfolio of work regarding the privacy, bias, and fairness issues surrounding Artificial Intelligence (AI), across many industry applications. We analyze policy proposals and provide feedback to policymakers. We speak with stakeholders – including leaders from the corporate, public sector, and non-profit communities – to exchange best practices and knowledge regarding machine learning models. After an extensive development process, we published Privacy Expert’s Guide to AI and Machine Learning,<sup>1</sup> and created a continuously updated set of resources for Ethics, Governance and Compliance news and guides,<sup>2</sup> and Artificial Intelligence and Robotics Publications.<sup>3</sup> These references comprise a compendium of information for those seeking guidance and updated analysis of the various challenges of machine learning applications in a variety of contexts, focusing on the challenges in common across industries.

---

<sup>1</sup> Future of Privacy Forum, Privacy Expert’s Guide to AI and Machine Learning, September 2018, <https://fpf.org/2018/10/18/fpf-release-the-privacy-experts-guide-to-ai-and-machine-learning/>

<sup>2</sup> Future of Privacy Forum, Ethics, Governance and Compliance Resources, <https://sites.google.com/fpf.org/futureofprivacyforumresources/ethics-governance-and-compliance-resources?authuser=1>

<sup>3</sup> Future of Privacy Forum, AI and Robotics Academic Publications, <https://sites.google.com/fpf.org/futureofprivacyforumresources/artificial-intelligence-and-robotics-academic-publications?authuser=1>

## **I. Introduction**

Artificial Intelligence technology continues to evolve and appear in new contexts in the financial services sector. There are several main uses and functions that benefit from AI, including: Trading Algorithms, Digital Identity Verification, Credit Scoring, Process Automation, Fraud Detection, and Anti-Money Laundering. New applications are being considered all the time for both “back office” functions and in consumer-facing opportunities.

There are, however, specific concerns about the privacy protections needed for the responsible use of this expanding technology, particularly in a highly regulated area such as financial services companies. In this sector more than any other, trust in the fair and equitable impacts of AI is critical to creating a foundation of protections for personal data. Concerns around bias must be carefully understood and managed to ensure appropriate policy and regulatory controls.

## **II. AI and Machine Learning Are Being Used and Considered for a Variety of Beneficial Applications in Financial Services**

“Artificial Intelligence” has become a catch-all phrase used to describe automated systems of all kinds. But it is important when considering consumer risks, as well as regulatory approaches, that the technology be specific and defined.<sup>4</sup> Machine Learning (ML) is the primary type of AI in use or being considered for Financial Services Applications, but not every form of AI is based on Machine Learning. AI includes natural language processing, much robotic process automation, machine learning, and within ML, the use of neural networks.

In the financial services industry – including commercial banks, retail banks, stock brokers, insurance companies, and others – AI is being incorporated in a variety of products and

---

<sup>4</sup> G. Zhe Jin, Artificial Intelligence and Consumer Privacy, January 2018, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3112040](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3112040)

services. These may include setting interest rates for mortgages, savings accounts, and student loans; recommendations for approving or rejecting credit card and loan applications; and offering or setting the terms for insurance policies. One common use across various parts of the Financial Services industry is fraud prediction and prevention. AI powers the “RegTech” or regulatory technology that allows banking firms to stay in compliance with “Know Your Customer” requirements and Anti-Money Laundering regulations.<sup>5</sup> AI is also used for identity verification (device fingerprinting; personal logins), and interacting with virtual assistants<sup>6</sup>, such as “chat bots” that help consumers set up an account, access help, or even provide long term investment strategy advice.<sup>7</sup>

Despite this extensive list, however, AI is used to a more limited degree in the Financial Services sector than many people might expect. Even in back office functions (predicting server down times; tracking data usage and flow; and staff management) where AI is used for improved process, efficiency, and accuracy, financial institutions most commonly keep people in the loop, using the system recommendations as an input to a human’s final review or decision. These organizations broadly realize that there is still much uncertainty as to impact of these models, in both practical, and legislative compliance related aspects. Many have determined that the

---

<sup>5</sup> Sudipto Ghosh, AiThORITY.com Primer on What is RegTech: Definitions, Stats and Tools, AiThORITY, February 3, 2020, <https://www.aithority.com/ait-featured-posts/aithority-com-primer-on-what-is-regtech-definitions-stats-and-tools/>

<sup>6</sup> Capitol One Finds Ways to Make Its Digital Assistants More Proactive, Donna Fuscaldo, Forbes, January 2020, [https://www.google.com/url?q=https://www.forbes.com/sites/donnafuscaldo/2020/01/31/capitol-one-finds-ways-to-make-its-digital-assistant-more-proactive/%235e7d97211706&sa=D&ust=1581457597308000&usg=AFQjCNEcXhD1Vr3u0DXi0Du2Gv\\_dRYB81g](https://www.google.com/url?q=https://www.forbes.com/sites/donnafuscaldo/2020/01/31/capitol-one-finds-ways-to-make-its-digital-assistant-more-proactive/%235e7d97211706&sa=D&ust=1581457597308000&usg=AFQjCNEcXhD1Vr3u0DXi0Du2Gv_dRYB81g)

<sup>7</sup> This \$11 Billion Tech Investment Could Disrupt Banking, JP Morgan Chase, <https://www.google.com/url?q=https://www.jpmorganchase.com/corporate/news/stories/tech-investment-could-disrupt-banking.htm&sa=D&ust=1581457597307000&usg=AFQjCNHR-qJ91Q2SLSnGH-zqnorScMT-ZQ>

maturity of these systems is such that while much may be implemented internally,<sup>8</sup> client facing features must be adopted slowly and carefully.

Other fintech areas where AI is being tested or considered include:

- Character recognition systems for medium term note issuance – this allows analysis and sharing of output data faster, more reliably
- Wire transfer processing
- Contract review (language search and analysis) – where the system is trained for targeted language and then processed for faster review and more consistent products

And like any business in any industry, fintech organizations may employ ML-based systems for HR processing; employee monitoring; machine monitoring; facility access; and cyber security.

Much of this is not new. Financial service providers have long engaged statistical and probability models as well as predictive analytics to forecast performance and evaluate risk. Now, with the inclusion of larger and more complex databases, and the availability of new methods of analysis, many fintech firms deploy extremely complex algorithms to predict the ROI, profitability, and repayment risks. Automation may be able to provide objective analysis using model-based assessments of a borrower's creditworthiness with the ability to better control for bias than traditional reviews subject to the limits of the human reviewer(s). At scale, the application of learning algorithms in credit markets may allow firms to consider nontraditional data in assessing creditworthiness and potentially integrate historically excluded individuals, expanding access to credit to the unbanked in the United States, as well as individuals globally who lack access to financial services.<sup>9</sup>

---

<sup>8</sup> DerivativePath, <https://www.derivativepath.com/> (as an example, the use of AI for foreign exchange and derivative management)

<sup>9</sup> Kristin Johnson et al., Artificial Intelligence, Machine Learning, and Bias in Finance: Toward Responsible Innovation, Fordham Law Review, Vol. 88, Issue 2, 2019, <https://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=5629&context=flr>

Some early examples of fintech firms promising to better integrate underserved communities were those who introduced digital money transfer services, the equivalent of cash exchanges (most familiarly using app platforms such as Venmo) as well as platforms that offered digitally distributed credit application functions. Facilitating cash exchanges provides opportunities for those who lack access to conventional banks with personal checking and savings accounts. And expanding credit markets by using sophisticated algorithms may increase the opportunities to offer credit – a necessary step for financial growth.<sup>10</sup> These services do also raise related concerns, including transparency and accountability on the part of the fintech organizations, along with the social impacts of determining “fairness” in credit markets and interest rates, marketing techniques, and structuring of credit products.<sup>11</sup> Many consumer advocates remain cautious. Even though “exclusionary and predatory” credit market practices are legally prohibited, discriminatory processes and inequitable outcomes persist.<sup>12</sup> Given the fears of exploitation and abuse of unbanked communities and higher risk credit applicants, plans for market expansion based on automated decision-making should be carefully considered.<sup>13</sup>

Automated decision-making processes in the financial services sector are built upon the combination of massive data built on the past and the freshest data from today. This means that decision-making algorithms can be “adversely trained” or taught to make sub-optimal decisions

---

<sup>10</sup> Duke University, FinReg Blog, November 2018, <https://www.google.com/url?q=https://sites.duke.edu/thefinregblog/2018/11/14/fintech-lending-risks-and-benefits/&sa=D&ust=1581457597364000&usq=AFQjCNH5184hUXE3c4r99iBd3ohjv0oGdA>

<sup>11</sup> Kristin Johnson et al., Artificial Intelligence, Machine Learning, and Bias in Finance: Toward Responsible Innovation, Fordham Law Review, Vol. 88, Issue 2, 2019, <https://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=5629&context=flr>

<sup>12</sup> Matthew Adam Bruckner, The Promise and Perils of Algorithmic Lenders’ Use of Big Data, Chicago Kent Law Review, Volume 93, Issue 1 FinTech’s Promises and Perils, 2018, <https://scholarship.kentlaw.iit.edu/cgi/viewcontent.cgi?article=4192&context=cklawreview>

<sup>13</sup> Tanaya Macheel, PayPal-TIO Deal Could Increase Venmo Revenue, Utility, Tearsheet, February 20, 2017, <https://tearsheet.co/modern-banking-experience/paypal-tio-deal-could-increase-venmo-revenue-utility/>



based upon short term variations in macro-economic forecasts, micro-economic trends, and local consumer consumption patterns. Without constant and effective monitoring of the performance of automated decision systems, such a system for approval of mortgage applications could be "adversely trained" to recommend approvals for applicants from areas with a higher income than other areas, even if that area has not historically been an area of high wealth or credit potential. Offering differential mortgage approvals based upon trends that adversely train AI systems is one form of undesirable biases resulting in disparate impacts.

### **III. Bias in Machine Learning Algorithms is a Complicated Problem, Implicating Fairness and Equality of Opportunities and Outcomes**

Systems historically run and managed by people demonstrate biases that are well documented. As recently as 2017, data from the Home Mortgage Disclosure Act showed that:

- 10.1 percent of Asian applicants were denied a conventional loan. By comparison, just 7.9 percent of white applicants were denied.
- 19.3 percent of Black borrowers and 13.5 percent of Hispanic borrowers were turned down for a conventional loan.<sup>14</sup>

Loan denial rates for some ethnic groups are far higher than the average denial rate of 9.6 percent. These results are from processes that did not rely on the use of AI.

For financial services institutions transitioning to digital systems, bias is a concern in almost every application, including algorithms to review loan applications, trade securities, predict financial markets, identify prospective employees, and assess potential customers. Addressing sources of system bias – that is, inequalities in either inputs, outputs, analysis processes, or settings and error rates that result in “unfair” recommendations – are an on-going

---

<sup>14</sup> Sray Agarwal et al., Fair AI: How to Detect and Remove Bias from Financial Services AI Models, Finextra, September 11, 2019, <https://www.finextra.com/blogposting/17864/fair-ai-how-to-detect-and-remove-bias-from-financial-services-ai-models>.

challenge in ML-based models in general.<sup>15</sup> The technology to evaluate models for system bias is advancing at the same time, but not always at the same pace, and so constant review and oversight is essential for any automated decision-making system, particularly those with legal or personally impactful outcomes.<sup>16</sup>

Discrimination regarding the impacts on any legally protected class is prohibited under the US Equal Credit Opportunity Act of 1974.<sup>17</sup> But bias clearly impacts outcomes in many systems, both those based on human decisions, and those relying on algorithmic recommendations. A 2018 study conducted at UC Berkeley found that both traditional face-to-face decisions and those made by machine learning systems charged Latinx/African-American borrowers interest rates that were 6-9 basis points higher.<sup>18</sup> The higher rate equates to these borrowers paying \$250-\$500 million per year in extra mortgage interest. However, the automated system did offer recommendations for loan approval to a broader percentage of minority applicants. The study concluded that algorithms had not fixed existing discrimination, but may have shifted the mode in the sense that more applicants were able to find financing at all.

Part of the challenge of automating these systems is that the biases from the patterns of the past are all too easily embedded in the automation of the present and future. While ML and AI are technologies thought of as completely “other” from human thinking, they are so far still always based on algorithms and models created by people. Thus, these algorithms are prone to

<sup>15</sup> S. Corbett-Davies and S. Goel, The Measure and Mismeasure of Fairness, Aug 2018, <https://arxiv.org/abs/1808.00023>

<sup>16</sup> Sarah Tan et al., Distill-and-Compare: Auditing Black-Box Models Using Transparent Model Distillation, 2017, <https://www.semanticscholar.org/paper/Distill-and-Compare%3A-Auditing-Black-Box-Models-Tan-Caruna/752fd6f73c0840e5919180441c3c575da4a41124>.

<sup>17</sup> 15 U.S. Code § 1691, available at: <https://www.law.cornell.edu/uscode/text/15/1691>.

<sup>18</sup> Robert Bartlett et al., Consumer-Lending Discrimination in the FinTech Era, November 2019, NBER Working Paper No. 25943, <http://faculty.haas.berkeley.edu/morse/research/papers/discrim.pdf>.



incorporating the biases of their designers, as well as the biases of the systems they're designed to serve, because the only data available to train them already reflects decades or even centuries of inequality. Because AI algorithms learn from data, any historical partiality in an organization's data can quickly create biased AI that bases decisions on inherently unfair datasets.<sup>19</sup>

These human biases exist in all industries and fields. Research has shown that judges' decisions are influenced by their own personal characteristics, while employers grant interviews at different rates to candidates with identical resumes but with names perceived to reflect different races.<sup>20</sup> Humans also routinely misinterpret information that they may identify as representing patterns of correlation.<sup>21</sup> Employment applications are sometimes reviewed to consider credit histories in ways that unfairly disadvantage minority groups, even though a link between credit history and job performance has not been established. Human-run processes are also difficult to review for consistency or reliability. People who self-report are frequently imprecise, whether deliberately or not, about the factors they considered, or may not even be aware of the various influences on their thinking.

Thus, training data is a part of the problem. Huge amounts of training data are required to train ML-based systems to any usable degree, but if this data comes from existing biased processes, the datasets created will reflect those inequities, and it will train the model such that

---

<sup>19</sup> AI Now, Race, Gender and Power in AI, April 2019, <https://medium.com/@AINowInstitute/gender-race-and-power-in-ai-a-playlist-2d3a44e43d3b>

<sup>20</sup> Marianne Bertrand & Sendhil Mullainathan, Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination, *American Economic Review*, Vol. 94, No 4, September 2004, <https://www.aeaweb.org/articles?id=10.1257/0002828042002561>

<sup>21</sup> Roel Dobbe et al., A broader view on bias in automated decision-making: Reflecting on epistemology and dynamics, July 6, 2018, arXiv preprint, <https://arxiv.org/pdf/1807.00553.pdf>

its recommendations will reflect those historical biases.<sup>22</sup> Consider if AI might be used by a company to set starting salaries for new hires. One of the inputs would certainly be salary history, but given the well-documented history of sexism in corporate compensation levels, that data could import gender bias into the calculations.<sup>23</sup>

A key problem for resolving the challenge presented by biased algorithms is identifying where the sources of bias arise. For simple algorithms based upon linear models, outcomes suggesting a disparate impact could be traced back to the sources of bias in the data, or the model components or the computations that led to that outcome. However, the greater complexity of ML models, which reflect thousands of variables and complex programming techniques like neural networks, make it unlikely that even the original programmers can say assuredly what the factors or interactions at fault might be. This is the problem of “explainability” and relates to the transparency of ML systems for review and evaluation.<sup>24</sup>

A further complication is that AI algorithms are by definition evolving. Unlike a static computer program, they “learn” and change over time. Initially, an algorithm creates recommendations using the process as refined on the training and testing datasets available at launch. Then based on the application of the model to real data, the system will continue to adapt its functioning, reflecting the continued processing of the increasing amounts of data. As the system gains experience in the form of more and more data, it further refines its connections and pattern analysis. These changes do not require human intervention to edit the code, but are

<sup>22</sup> Sally Ward-Foxton, Reducing Bias in AI Models for Credit and Loan Decisions, EE Times, April 30, 2019, <https://www.eetimes.com/reducing-bias-in-ai-models-for-credit-and-loan-decisions/#>

<sup>23</sup> John Billasenor, Artificial Intelligence and Bias: Four Key Challenges, Brookings Institute, January 3, 2019, <https://www.brookings.edu/blog/techtank/2019/01/03/artificial-intelligence-and-bias-four-key-challenges/>

<sup>24</sup> Andrew Burt et al., Beyond Explainability: A Practical Guide to Managing Risk in Machine Learning Models, July 2018, <https://fpf.org/2018/06/26/beyond-explainability-a-practical-guide-to-managing-risk-in-machine-learning-models/>

modifications made by the model to its own programming. In some cases, this evolution can introduce or strongly reinforce an undesired bias.<sup>25</sup>

Even in systems that do not collect or use data that includes sensitive or protected class fields such as race or gender, bias relative to these traits can occur. This is due to data “proxies” – fields that strongly correlate with other factors such that the patterns identified using them will result in outcomes that impact along those protected categories. The most commonly used example is the fact that zip codes frequently turn out to be a proxy for socio-economic status, race, and sometimes even general employment categories. Thus, if the system at issue is searching for patterns to define fraud scoring risk levels, it can end up scoring some racial groups at higher levels, despite never having had access to data about their race.

Organizations using these systems must continuously test the adoption of proxies within the model, that is, outputs that align along discriminatory lines, regardless of original design or intent. Not only test for but also be willing to discard models that exhibit proxies with disparate outcomes.

However, the way AI works for analysis and pattern recognition means algorithms can also be part of the solution. In some cases, AI can be applied to identify, and then reduce, humans’ misinterpretation of patterns. Some experiments show that algorithms can impact decision making in a way that causes it to become fairer when measured across identified classes.<sup>26</sup> One study resulted in automated financial underwriting systems benefitting historically

<sup>25</sup> J. Dunkelau, et al, Fairness Aware Machine Learning, October 2019, [https://www.phil-fak.uni-duesseldorf.de/fileadmin/Redaktion/Institute/Sozialwissenschaften/Kommunikations-und\\_Medienwissenschaft/KMW\\_I/Working\\_Paper/Dunkelau\\_Leuschel\\_2019\\_Fairness-Aware\\_Machine\\_Learning.pdf](https://www.phil-fak.uni-duesseldorf.de/fileadmin/Redaktion/Institute/Sozialwissenschaften/Kommunikations-und_Medienwissenschaft/KMW_I/Working_Paper/Dunkelau_Leuschel_2019_Fairness-Aware_Machine_Learning.pdf)

<sup>26</sup> Jon Kleinberg et al., Human Decisions and Machine Predictions Quarterly Journal of Economics, Volume 133, Issue 1, February 2018, <https://academic.oup.com/qje/article-abstract/133/1/237/4095198>

underserved applicants.<sup>27</sup> Recommendations made by AI could be analyzed and audited by other AI systems for more accurate understanding of their consistency, reliability, and potential bias.

#### **IV. How Can Bias Be Managed in Financial Services AI Using Technological and Governance Solutions?**

Artificial Intelligence can be a pain point for the Financial Services sector, but it can be managed through both conventional governance tools and by using other technological tools to expose and mitigate algorithmically driven biases. For example, governance tools can include careful use of contractors' expertise and managerial attention to employee's attitudes towards uses of AI.

A survey of professionals in the financial services industry sought to identify the primary areas they felt could be, or had been, improved with AI systems. Higher accuracy, greater consistency, and reduced processing times were some of the most significant benefits of AI technology across the backoffice applications.<sup>28</sup> In the same study, most individuals preferred a contracted model where the financial services provider would partner with an outside provider to manage their AI technology systems. This reflects the recognition that AI technology requires particular expertise to implement and manage. In-house capabilities are unlikely to be sufficiently sophisticated for the maturity of increasingly complex AI platforms and models.

This perspective is likely to be correct, as any AI algorithm can have bias: in the data, in the model design, or creeping into it "in the wild" (i.e. in applications with real life data and situations). Trained AI programmers and designers are likely necessary to proactively look for

---

<sup>27</sup> AnnaMaria Andriotis, Freddie Mac Tests Underwriting Software That Could Boost Mortgage Approvals, Wall Street Journal, September 24, 2019, <https://www.wsj.com/articles/freddie-mac-tests-underwriting-software-that-could-boost-mortgage-approvals-11569333848>

<sup>28</sup> Marcel Deer, AI Adoption in the Financial Sector, Marcel Deer, Medium, February 7, 2019 <https://medium.com/towards-artificial-intelligence/ai-adoption-in-the-financial-sector-77c6bb81cfd3>

and identify bias, correct for it, then ensure future processing outputs are fairer. Unfortunately, there is no one best or proven way to do this evaluation for every case. Research is progressing by academics and industry research and development to find ways to accomplish this analysis.

These questions are an example of a broad framework for ways to check for systematic bias:<sup>29</sup>

1. Are any identifiable groups suffering from systematic data error?
2. Has any group been ignored, or underrepresented?
3. Are groups represented proportionally, particularly along protected class categories?
4. Are there enough features to sufficiently include minority groups?
5. Is the model using or creating factors that are proxies?
6. Are there stereotyping features?
7. Is the model appropriate for the underlying use case?
8. Is the output accuracy similar for all groups? (Are predictions skewed any identifiable subsets or groups?)
9. Is the model optimizing all required metrics?

There are other ways to design or describe useful frameworks, with similar considerations for the analysis and reviews. This set of recommendations is another way to consider what the model impact is.<sup>30</sup> By creating alternative groups to simulate protected classes, and reviewing factors to ensure these groups have equal predictive values and equality across false positive and false negative rates, it is possible to detect and potentially measure bias in your AI.

1. Ensure all data groups have an equal probability of being assigned to the favorable outcome for a protected/sensitive class.
2. Ensure all groups of a protected/sensitive class have equal positive predictive value.
3. Ensure all groups of a protected/sensitive class have predictive equality for false positive and false negative rates.
4. Maintain an equalized odds ratio, opportunity ratio and treatment equality.
5. Minimize the average odds difference and error rate difference.

<sup>29</sup> Aarwal, Fair AI: How to Detect and Remove Bias from Financial Services AI Models <https://www.finextra.com/blogposting/17864/fair-ai-how-to-detect-and-remove-bias-from-financial-services-ai-models>

<sup>30</sup> Id.



Explainability of these systems in a way that is sufficient to satisfy everyone from regulators, to consumers, and industry experts will remain challenging.<sup>31</sup>

There are other tools and solutions that can be applied to the datasets as well. Statistical analysis tools like aggregation, masking records or fields, injecting “noise” into datasets, blurring and perturbations are all ways to manipulate the data to both provide protection for individual data, and to improve the evaluative accuracy of the dataset as a whole.

Differential privacy and synthetic data are also options. Synthetic data, while still in its early stages, shows promise for many of the challenges for correcting historically biased data. Synthetic data is a generated dataset of fake individual records that sufficiently represent the scale and scope of actual data to be useful for many of the analysis functions that do not reflect upon specific individuals or impact individual accounts. The synthetic datasets can be optimized for accuracy, to mirror as closely as possible the details of actual data, but they can also be optimized for less bias. There are always tradeoffs for these types of optimization, in this case, a likely loss of some accuracy or functionality. However, the balance of accuracy and fairness can be managed to ensure that the resulting dataset is sufficient for internal sharing, access management, research, and modeling – this keeps risk lower with minimal numbers of individuals having access to “real” customer data. This type of artificial data might be sufficient for designing user interfaces or testing for accessibility from third party platforms.

## **V. What Actions Could be Taken in the Legal and Regulatory Environment?**

There are times when discussing legal and regulatory standards for AI and ML-based systems when the concerns and arguments expressed imply we are starting from some sort of

---

<sup>31</sup> P. Hall, et al, Proposed Guidelines for the Responsible Use of Explainable Machine Learning, November 2019, <https://arxiv.org/pdf/1906.03533.pdf>

blank slate, and that when the challenges of bias in these systems become apparent, we must immediately take targeted action to prevent harm. But in fact, AI systems operate in the same regulated world that exists for other technology platforms.

Since the civil rights movement of the 1960s, there have been claims against the financial services industry that institutions treated some individuals less favorably than others. Once the civil rights laws established “protected classes” for particular oversight, the focus was on discrimination affecting individuals in those classes. In 1971, the term “disparate impact” was first used in the Supreme Court case *Griggs v. Duke Power Company*. The Court ruled that it was illegal for a company to rely on factors which were shown to unfairly favor white applicants to make hiring or promotion decisions, whether or not the discrimination was intentional.<sup>32</sup> This lack of intent is still applicable – and any AI systems that yield recommendations that demonstrate a disparate impact on protected classes would still be illegal.

In addition to intent, more recent cases have made disparate impact claims that focus on the effect, instead of the intention, of lending policies. The Supreme Court ruling in *Texas Department of Housing and Community Affairs v. Inclusive Communities Project* affirmed the use of the disparate impact theory based on outcomes. In this case a statistical analysis of housing patterns showed that a tax credit program resulted in effective segregation by race.<sup>33</sup> Affirming disparate impact should be a flag for technology and compliance managers in financial services. An algorithm that inadvertently disadvantages a protected class continues to be unacceptable under existing laws.

---

<sup>32</sup> Davis, Wright, Tremaine, Discrimination and Algorithms in Financial Services: Unintended Consequences of AI, March 6, 2018, <https://www.dwt.com/blogs/payment-law-advisor/2018/03/discrimination-and-algorithms-in-financial-service>

<sup>33</sup> *Id.*

Other current laws and regulations still apply as well, including the general laws against unfair or misleading trade practices, labor and employment laws, applicable privacy laws, as well as the entire regulatory structure around financial services in particular. Therefore, taking new action to legislate AI specifically should be approached with caution.

As discussed in earlier sections, there are developing best practices for overall AI fairness implications. These emergent AI governance practices and standards should be the baseline of any further guidance. However, AI risk-benefit comparisons are vastly different depending on context and application, and it is impossible to consider that any one rule could successfully address bias concerns across the entire range of use cases. It is possible that some level of legislative guidance would be appropriate in the new digital environment of automated decision making using these complex systems, but if so, the most effective would likely be based on protecting the underlying values and principles<sup>34</sup> at issue rather than seeking to set detailed technical standards or create performance rules that could easily be avoided or outdated in a short time.

#### **VI. Conclusion:**

Financial services organizations have the responsibility, both legally and ethically, to treat their customers, whether other businesses or individuals, fairly and equally. As more players in this industry employ AI systems in more use cases, it is incumbent on them to ensure that their algorithms are fair and explainable.

Similar challenges regarding new technology applications have been faced before. From wiretapping phones, to accessing the contents of emails, consumer protection laws have had to address the issues around particular technology platforms and determine how best to provide

---

<sup>34</sup> D. Mulligan, et al, This Thing Called Fairness, September 2019, <https://arxiv.org/pdf/1909.11869.pdf>



appropriate levels of privacy and security for individuals, protect their interests as consumers, and also facilitate business models that provide useful features and services.<sup>35</sup> These historical examples reflect the ongoing need to determine the appropriate balance of technological, legal, and policy standards and protections, along with the underlying threshold question of whether some applications, or some use cases, are simply too high risk to implement regardless of perceived benefits.

AI systems offer many potential benefits, including the opportunity to improve on biased human systems, and to increase fairness and equality at scale, but to do so there must be appropriate accountability across developers and users for their impacts, and clear evaluations of how these models are applied or used in ways that affect individuals. How we face these challenges will determine how we move further into the conveniences of a digital world, while continuing to embrace our fundamental ideals of personal liberty and freedom.

---

<sup>35</sup> J. Black, A. Murray, *Regulating AI and Machine Learning: Setting the Regulatory Agenda*, 2019, <http://ejlt.org/article/view/722/980>



Cathy O'Neil, CEO  
O'Neil Risk Consulting and Algorithmic Auditing  
15 Claremont Ave, 91  
New York NY 10027  
(617)780-1051  
cathy@orcaarisk.com

February 7, 2020

Chair Bill Foster  
Ranking Member Barry Loudermilk  
Honorable Members  
House Financial Services Committee Task Force on Artificial Intelligence  
2129 Rayburn House Office Bldg.  
Washington, DC 20515

Dear Chair, Ranking Member, and Members:

I'd like to thank you for this opportunity to offer an opinion on your important work on understanding and reducing artificial intelligence (AI) bias in financial services. I am the author of the 2016 book *Weapons of Math Destruction: how big data increases inequality and threatens democracy*, which delved into this very question of algorithmic bias for different financial industries such as insurance, hiring, and credit. In 2016 I founded the company O'Neil Risk Consulting and Algorithmic Auditing (ORCAA), an algorithmic auditing company that helps companies and governmental agencies deploy algorithms equitably and legally.

As an expert, I have good news and bad news. The bad news is that we should expect all algorithms to start out biased, simply because they are trained on data that echoes our imperfect society. The AI that automates human processes are just as problematic as the human processes they replace, at least at the beginning. The good news, though, is that we do have the ability to measure the bias in AI, account for it, and often modify the AI to be less problematic. In other words, AI can and should be scrutinized for bias, we should expect to find it, and we should demand that illegal bias is removed.

Along these lines, I would like to submit an essay I recently wrote about my concerns regarding the United States Department of Housing and Urban Development (HUD) approach to the legal theory of disparate impact. In my opinion, HUD is not asking for enough oversight on AI in



housing and gives too much leeway to landlords to use algorithms as a mechanism of discrimination.

This is just one essay of many that I could write regarding how we can and must hold AI to high legal standards. Thank you for your continuing work.

Sincerely,

A handwritten signature in black ink that reads "Cathy O'Neil". The signature is written in a cursive, flowing style.

Cathy O'Neil



---

# EDUCATIONAL REDLINING

Student Borrower Protection Center

February 2020

---

[PROTECTBORROWERS.ORG](https://PROTECTBORROWERS.ORG)

---

With new advances in financial products and services come age-old risks of discrimination. Without caution, the fintech revolution could perpetuate a system that has historically locked communities of color out of mainstream credit markets.

## Table of Contents

Executive Summary	04
About this Report	06
Introduction	08
The Community College Penalty	11
The HBCU/HSI Penalty	15
Recommendations	20
Conclusion	25

## Executive Summary

- Across the financial services sector, “alternative data” has been touted by established consumer lenders and new entrants alike as a tool to expand access to credit for historically underserved communities, including people of color. This report examines one subset of this data—education data, an umbrella term describing information related to a consumers’ higher education—when determining access to credit and the price of consumer financial products.
- The use of education data in underwriting raises significant fair lending concerns, and its widespread adoption could reinforce systemic barriers to financial inclusion for Black and Latinx consumers. Further, the use of education data can exacerbate inequality across the American economy. Where the effects of these practices have negative economic consequences for borrowers from historically marginalized communities, these practices are known as “Educational Redlining.”
- The following report, *Educational Redlining*, includes a detailed discussion of these practices and describes the specific risks posed to borrowers, communities, and the economy when consumer lenders rely on education data when determining access to credit and the cost of credit.
- This report features two case studies that examine the effects of these practices on hypothetical, similarly situated consumers using publicly available information about the lending practices at two consumer lenders—Wells Fargo and the financial technology company Upstart. These case studies show:
  - **Borrowers who take out private loans to pay for college may pay a penalty for attending a community college.** Wells Fargo charges a hypothetical community college borrower an additional \$1,134 on a \$10,000 loan when compared to a similarly situated borrower enrolled at a four-year college.
  - **Borrowers who refinance their student loans through a company using education data may pay a penalty for having attended an HBCU.** When refinancing with Upstart, a hypothetical Howard University graduate is charged nearly \$3,499 more over the life of a five-year loan than a similarly situated NYU graduate.

- **Borrowers who refinance student loans may pay a penalty for having attended an Hispanic-Serving Institution (HSI).** When refinancing with Upstart, a hypothetical graduate who receives a Bachelor's Degree from New Mexico State University, an HSI, is charged at least \$1,724 more over the life of a five-year loan when compared to a similarly situated NYU graduate.
- Based on this analysis, SBPC has issued the following recommendations to Congress, federal and state regulators, and the consumer lending industry to address potential violations of federal and state fair lending laws and to mitigate the effects of these practices on economic inequality:
  - **Congress must enhance oversight.** Congress should examine the use of education data by consumer lenders, including monitoring for potential disparities caused by this practice and its effects on economic inequality. Further, Congress should investigate regulators' oversight over the companies engaged in these practices. This should include scrutiny of the Consumer Financial Protection Bureau's handling of the No-Action Letter awarded to Upstart—a regulatory safe harbor that may be shielding the company from violations of federal fair lending laws.
  - **Federal and state regulators must take immediate action to halt abuses.** Federal and state regulators should prioritize oversight over lenders that use education data when underwriting or pricing consumer loans and take immediate action where industry practices violate fair lending laws.
  - **The financial services industry must strengthen transparency when lending based on education data.** Firms in the financial services industry that use alternative data should immediately publish data demonstrating the effects of such practices on individual borrowers, empowering lawmakers, regulators, and the public to understand the effects of these practices on consumers.



## About this Report

Credit is a key ingredient in the generation of economic opportunity, and it plays a “remarkably consequential” role in the expansion of economic mobility among marginalized populations.<sup>1</sup> And yet, consumers of color continue to face obstacles when seeking access to affordable credit. Research shows that African American and Latinx consumers at every income bracket are more likely to either be offered

“As more financial services companies look to adopt this approach, policymakers, regulators, and fintech companies must heed caution. The use of alternative data may further marginalize the very communities it purports to help.”

less credit than requested or denied credit outright than their similarly situated white peers.<sup>2</sup> While racial disparities in credit can be traced back to systemic discrimination underlying American society and the U.S. financial system,<sup>3</sup> evidence suggests that traditional credit scoring models perpetuate these disparities because “even the most basic lending standards . . . ‘impact’ racial and ethnic groups differently.”<sup>4</sup>

Financial technology (fintech) firms have touted the use of “alternative data” as a method for overcoming biases entrenched in traditional credit underwriting models that often exclude consumers with limited credit profiles.<sup>5</sup>

These companies assert that creditworthiness can be gauged through factors like social media use, educational attainment, and work history.<sup>6</sup> After including these alternative inputs in underwriting models, companies market their products as providing expanded access to credit to marginalized communities.<sup>7</sup> However, as this report demonstrates, such statements fail to present policymakers, regulators, and law enforcement officials with full context for the potential risks associated with using alternative data.

As more financial services companies look to adopt this approach, policymakers, regulators, and fintech companies must heed caution. The use of alternative data may further marginalize the very communities it purports to help.

In 2019, Student Borrower Protection Center (SBPC) fellow Aryn Bussey documented the risks associated

with one category of alternative variables for credit underwriting: education data.<sup>8</sup> Companies using education data have looked to SAT scores, sector of the institution of higher education attended (*e.g.*, for-profit, private nonprofit, public), college majors, and more as proxies for likelihood of repayment.<sup>9</sup> Bussey's analysis reviewed the myriad of concerns of policymakers, academics, advocates, and law enforcement related to the use of education criteria in underwriting.<sup>10</sup> This report builds on Bussey's work, further examining those risks, and provides two case studies highlighting disparities in outcomes when companies use education data in underwriting decisions.

Specifically, in this report, we examine the extent to which a consumer's choice of college, including attendance at a community college or Minority-Serving Institution (MSI), impacts their cost of credit. We analyze sample rate quotes from lenders that advertise the use of education criteria in credit decisions and provide case studies for two lending products: a newly originated private student loan from Wells Fargo and private student loan refinancing products offered by Upstart. Offered rates were compared across postsecondary institutions with all other inputs held constant.<sup>11</sup> Our findings from our broader analysis and the highlighted case studies are consistent: holding all else constant, borrowers who attend community colleges, Historically Black Colleges and Universities (HBCUs), and Hispanic-Serving Institutions (HSIs) will pay significantly more for credit because of people's assumptions and prejudices regarding those who sit next to them in the classroom.

“ Our findings from our broader analysis and the highlighted case studies are consistent: holding all else constant, borrowers who attend community colleges, Historically Black Colleges and Universities (HBCUs), and Hispanic-Serving Institutions (HSIs) will pay significantly more for credit, because of people's prejudices regarding those who sit next to them in the classroom. ”

## Introduction

The fintech industry is rapidly changing the way that consumers participate in credit markets. Researchers estimate that the credit market excludes 45 million consumers because classic underwriting models deny credit to those with little or no scorable credit history.<sup>12</sup> Fintech companies increasingly seek to serve this population by incorporating new forms of data into underwriting models. In doing so, these companies claim they can offer lower cost products that are more widely available.<sup>13</sup>

Should this claim be realized, this approach would be encouraging, as expanded access to affordable credit is critical to improving economic opportunity and creating fairer financial markets for traditionally marginalized consumers. However, as this report shows, the use of alternative data in underwriting to predict credit risk may ultimately do just the opposite—disparately affecting marginalized consumers and exacerbating economic inequality.

Traditional underwriting algorithms use a consumer's past payment performance to predict repayment behavior and determine creditworthiness.<sup>14</sup> As a result, these models are somewhat limited in their ability to assess the creditworthiness of young consumers and others who lack extended payment histories.<sup>15</sup> Additionally, critics contend that classical score-based credit models overlook consumers with repayment histories concentrated outside of mainstream credit products.<sup>16</sup> Fintech companies have sought to fill this gap and expand their base of potential customers by looking beyond these extant input variables. Fintech lenders use new input variables—commonly referred to as alternative data—in underwriting algorithms to process data “in ways that reveal correlations between seemingly irrelevant data points about a borrower and that borrower's ability to repay.”<sup>17</sup>

This report focuses on one specific class of input variables increasingly used by fintech lenders—education data. Education data includes a range of variables tied to a consumer's postsecondary education, including institutional sector and selectivity, college major, and even assessment scores. As University of Oklahoma College of Law professor Christopher Odinet explains, fintech firms “are ever-expanding their online lending activities to help students finance or refinance educational expenses. These online companies are using a wide array of alternative, education-based data points—ranging from applicants' chosen majors, assessment scores, the college or university they attend, job history, and cohort default rates—to determine

creditworthiness.<sup>18</sup>

However, while the fintech industry argues that education data allows for expanded and more inclusive underwriting, this report illustrates how its use may lead to disparate outcomes for certain consumers.<sup>19</sup> Specifically, the use of education data in underwriting risks discriminating against borrowers of color and exacerbating income equality across the population at large. As National Consumer Law Center staff attorney Chi Chi Wu testified before Congress:

The use of education and occupational attainment reinforces inequality, given that a consumer's educational attainment is most strongly linked with the educational level of his or her parents. Use of educational or occupational attainment would probably top the list of mobility-impeding data, and would ossify the gaping racial and economic inequality in our country.<sup>20</sup>

With new advances in financial products and services come age-old risks of discrimination, thereby perpetuating a system that has historically locked communities of color out of mainstream credit markets. Accordingly, non-individualized input variables that risk reinforcing systemic disparities and discrimination demand greater scrutiny from policymakers and law enforcement. Education data is no exception.

For example, people of color have historically been and continue to be denied equitable access to higher education, particularly at elite institutions.<sup>21</sup> By considering the college or university attended by the consumer, a lender may capture disparate patterns in college attendance across class and race, thereby introducing bias in the underwriting process.<sup>22</sup> The resulting credit decision risks producing discriminatory results. As Bussey explains:

[A]lthough degree attainment is on the rise for many racial and ethnic groups, research shows there is a shortage of minority students, particularly African-American and Latino students, at selective institutions of higher education. Only nine percent of Black students, eight percent of Indigenous American students, and twelve percent of Latino students attend America's most elite public universities. When credit terms are tied to attendance at supposedly "elite" institutions, it can unfairly impact borrowers of color. Widespread adoption of educational criteria to determine creditworthiness will further stratify socioeconomic barriers to economic opportunity and mobility for Black and Brown consumers.<sup>23</sup>



## Education Data Use Risks Redlining

Discrimination resulting from the use of education data in underwriting is not new. For the last century, borrowers of color have been subjected to discriminatory credit terms simply because of where they live.<sup>24</sup> Despite fair lending laws prohibiting this type of practice, modern-day redlining based on geography continues to stymie economic opportunity for consumers of color.<sup>25</sup> Similar to the effects of discrimination based on geography, the use of educational data in underwriting risks redlining people of color out of the American Dream once again.

For example, in 2007, then-New York Attorney General Andrew Cuomo launched an inquiry to determine whether lenders' use of certain criteria discriminated against student loan borrowers based on their enrollment at a specific institution of higher education.<sup>26</sup> Cuomo noted the potential for educational redlining when warning that students attending minority-serving institutions (MSIs), such as historically black colleges and universities (HBCUs), may pay much higher interest rates.<sup>27</sup> Cuomo's investigation into one large lender found that its use of education data in underwriting led to interest rate spreads of up to six percent when compared to similarly situated borrowers simply because of the school attended by the applicant.<sup>28</sup>

Since Cuomo's inquiry, regulators and researchers have further documented how the use of education criteria in underwriting decisions is likely to disproportionately affect protected classes.<sup>29</sup> This outcome is particularly troublesome where lenders consider the selectivity of an institution in underwriting. First, despite perceptions of institutional prestige and future earnings, researchers have repeatedly found that institutional selectivity does not broadly correspond with increased earnings, finding only a "slight effect, if any at all."<sup>30</sup> Second, as previously discussed, the use of education data risks perpetuating the deep-rooted discrimination that pervades America's higher education system. And finally, potentially discriminatory factors are unjustified where "nondiscriminatory [factors] . . . are already highly predictive of likelihood of repayment."<sup>31</sup>

Accordingly, it is imperative to understand and protect against the potential for discrimination against subsets of borrowers.<sup>32</sup>

## The Community College Penalty

Community colleges play a critical role in the higher education ecosystem by providing a local pathway to postsecondary learning for a broad range of students, particularly low-income, first generation, and underrepresented minority students.<sup>33</sup> For example, while 37 percent of Latinx college students attend a public four-year or private nonprofit four-year institution, 56 percent of Latinx students attend public two-year institutions.<sup>34</sup> Similarly while only 39 percent of white students attend a two-year public college and 56 percent attend a four-year institution, 44 percent of black students attend a two-year public college, a proportion larger than the percent of black students attending a four-year institution.<sup>35</sup>

In theory, affordable, accessible post-secondary education should help mitigate the racial wealth gap and improve economic mobility. However, the increased use of education data in underwriting models threatens to do the opposite. As the following case study illustrates, rather than providing community college students with affordable credit, consumer lenders instead enforce a community college penalty. Our case study shows that, in one example of a private student loan product marketed by a large bank, borrowers attending community colleges might be charged higher interest rates and offered shorter repayment terms than otherwise identical peers at four-year schools. This penalty risks disparately impacting borrowers of color and necessarily involves judging people's individual creditworthiness based on nonindividualized factors.

In the following case study, we use publicly available information about the terms and conditions of Wells Fargo's private student loan offerings, comparing hypothetical Wells Fargo customers enrolled at select community colleges with similarly situated Wells Fargo customers enrolled at select four-year institutions. The findings of this case study highlight how this approach to pricing can adversely affect students at community colleges, and in turn, students of color.

## Case Study: Wells Fargo

Wells Fargo Bank offers a series of private student loan products for higher education financing.<sup>36</sup> The following study analyzes two of these product offerings: the *Wells Fargo Collegiate* student loan, a private student loan available to all undergraduate students attending four-year schools,<sup>37</sup> and the *Wells Fargo Student Loan for Career & Community College*, a private student loan available specifically to students attending two-year schools, career-training programs, and other non-traditional schools.<sup>38</sup>

### Methodology

To determine how community college attendance affects private student loan product pricing, we modeled hypothetical applicants attending community colleges and four-year colleges. Applicants are identical in every respect, except for the institution of higher education attended.

Using input information for each hypothetical applicant, we submitted inquiries for private student loan product offers using Wells Fargo's publicly available "Today's Rates" tool.<sup>39</sup> We then compared the terms presented in the respective outputs from Wells Fargo. Because Wells Fargo reports a range of interest rates for each of its various student loans, we based our analysis on the average of the interest rates quoted for each credit product. We applied those averages to a model paydown sequence for a \$10,000 loan to find implied monthly payments and total payments across the loan term. We assumed that the loan has no origination fee, that the loan was disbursed in equal halves in August and January of the student's final year of study, and that a six-month grace period followed the student's graduation.

In the example below, we highlight the outputs for hypothetical applicants attending two institutions: Chapman University, a four-year university in Orange, California, and Los Angeles ORT College, a community college in Los Angeles, California. We opted to highlight these two institutions based on their proximity,<sup>40</sup> but note that the findings were consistent across hypothetical applicants.

### Findings

This section explores the rate and cost variation offered to borrowers of a *Wells Fargo Collegiate Loan* and *Wells Fargo Career & Community College Loan*.

EDUCATIONAL REDLINING

2020

**Bank Lender: Wells Fargo**  
Product: Private Student Loan

**LOAN  
AMOUNT  
\$10K**

### Borrower Profile

#### Chapman University

(Private 4-Year University)

Major: Computer science  
Occupation: Financial analyst  
Annual income: \$50,000

#### LOAN OFFERS

Loan Interest Rate:  
8.22%

**Total Cost:  
\$19,171**

#### Los Angeles ORT College

(Community College)

Major: Computer science  
Occupation: Financial analyst  
Annual income: \$50,000

#### LOAN OFFERS

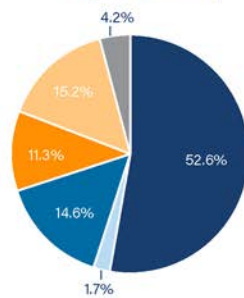
Loan Interest Rate:  
10.87%

**Total Cost:  
\$20,305**

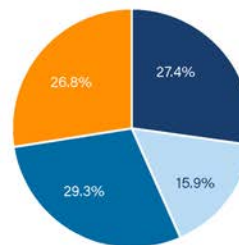
**Community College Penalty: +\$1,134**

### Student Populations

Chapman University



Los Angeles ORT College



■ White ■ Black/African American ■ Latinx/Hispanic ■ Asian ■ Other/Unknown ■ Non-Resident Alien

Demographic data from the U.S. Dept't of Education



- **Wells Fargo charges higher interest rates on its community college loan than its four-year undergraduate loan for similarly situated borrowers.** Using the average of reported rates, a borrower with a community college loan would pay \$1,134 more on a \$10,000 loan than a borrower with the four-year undergraduate loan. Over the life of a \$10,000 loan, a community college borrower would pay approximately \$16,829 with the lowest rate offering and \$24,200 with the highest rate offering. In comparison, a four-year undergraduate loan borrower would pay \$14,749.40 with the lowest rate offering and \$24,335 with the highest rate offering. Even with identical credit profiles, community college borrowers would pay a higher price for credit than students at four-year institutions.
- **Wells Fargo offers shorter loan repayment terms, regardless of the borrower's creditworthiness, for its community college loans.** Wells Fargo offers a 12-year repayment term on its Career & Community College Loan. In contrast, Wells Fargo offers a 15-year repayment terms on its Collegiate Loan. However, a borrower with the community college loan would still pay more overall due to the higher interest rates they face. Both loan products offer the same terms for in-school deferment and grace periods.

## The HBCU/HSI Penalty

Minority-Serving Institutions (MSIs), including Historically Black Colleges and Universities (HBCUs) and Hispanic-Serving Institutions (HSIs), play a significant role in expanding access to higher education. For example, in addition to serving underrepresented minorities, HBCUs and HSIs are also more likely to enroll women and older students.<sup>41</sup> However, as one researcher notes, these institutions “exist at the intersection where the American Dream of unbridled possibilities meets the American Nightmare of persistent racial-ethnic subordination.”<sup>42</sup>

HBCUs, HSIs, and the students they serve face obstacles that make student debt almost an inevitability for attendees. For example, these institutions notably receive less funding than non-minority serving institutions.<sup>43</sup> Additionally, students attending HBCUs and HSIs take on more student debt, on average.<sup>44</sup>

As the following case study illustrates, fintech lenders’ use of education data may impose an “HBCU/HSI penalty” on borrowers—a financial burden that has measurable, immediate economic consequences even for graduates who have already managed to overcome the obstacles described above. Our case study shows that borrowers who graduated from HBCUs or HSIs may be charged higher interest rates and origination fees than borrowers who graduated from non-minority serving institutions, thereby risking disparately impacting borrowers of color.

In the following case study, we use publicly available information about the rates offered to applicants seeking to refinance student loan debt with Upstart Network (Upstart), comparing hypothetical Upstart customers who graduated from HBCUs or HSIs, with similarly situated Upstart customers who graduated from select four-year institutions and non-minority serving institutions. The findings of this case study highlight how the use of alternative data in underwriting can adversely affect certain consumers of color in the education finance market even after they have already graduated.

## Case Study: Upstart

Upstart is an online lending platform that provides financing for a range of personal loans.<sup>45</sup> According to the company, its platform is intended to “improve access to affordable credit while reducing the risk and cost of lending” to its partners.<sup>46</sup> In addition to using traditional underwriting criteria, Upstart also incorporates nontraditional factors such as educational attainment and employment history.<sup>47</sup> As with most fintech lenders, Upstart’s underwriting algorithm is proprietary, but Upstart has publicized its use of alternative data in lending decisions.<sup>48</sup>

In September 2017, the Consumer Financial Protection Bureau (CFPB) issued its first No-Action Letter (NAL) to Upstart.<sup>49</sup> The NAL “signifies that [the CFPB] has no present intent to recommend initiation of supervisory or enforcement action against Upstart with respect to the Equal Credit Opportunity Act.”<sup>50</sup> In accordance with the NAL, Upstart has reported lending and compliance information to the CFPB, such as approval decisions, mitigation of consumer harm, and expansion of access to credit for underserved populations.<sup>51</sup>

## Methodology

To determine how the choice of institution attended affects the pricing of private student loan refinancing products, we modeled hypothetical applicants with degrees from schools across various institutional sectors, including two- and four-year colleges with HBCU, HSI, and non-MSI designations. Inputs for prospective applicants were identical in every respect, except for the institution attended by the applicant.

Each hypothetical applicant is a 24-year-old New York City resident with a bachelor’s degree.<sup>52</sup> Each applicant works as a salaried analyst at a company not listed among those offered by Upstart. Applicants have been employed by their current employer for five months, earn \$50,000 annually, and have \$5,000 in savings. Applicants have no investment accounts or additional compensation and have not taken out any new loans in the past three months. Each applicant requested a \$30,000 student loan refinancing product.

Using the above input information for each hypothetical applicant, we submitted inquiries for a private student loan refinancing product using Upstart’s publicly available rate comparison tool.<sup>53</sup> We then compared the terms presented in the respective outputs.

In the example below, we highlight the outputs for hypothetical applicants attending three institutions: New York University (NYU), a non-MSI; Howard University, an HBCU; and New Mexico State University-Las Cruces (NMSU), an HSI. We opted to highlight these three institutions based on their varied MSI designations,<sup>54</sup> but note that the findings were consistent across hypotheticals.

### Findings

This section explores the rate and cost variation offered for private student loan refinancing products to otherwise identical borrowers who attended different colleges. Results are based on applicants seeking \$30,000 to refinance student loans, to be repaid over three- or five-year terms.

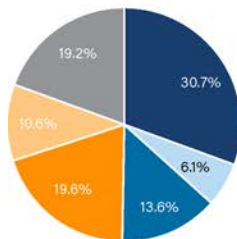
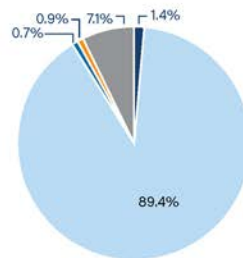
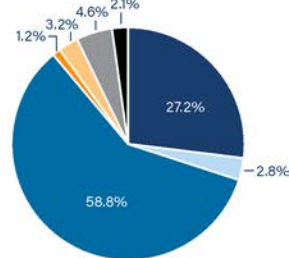
Holding all other inputs for prospective applicants constant, we find that a hypothetical refinancing applicant who attended Howard University, an HBCU, would pay more than an applicant who happened to have attended NYU. In this example, borrowers who attended the HBCU pay higher origination fees and higher interest rates over the life of their loans. Similar results are observed for applicants who attended NMSU, an HSI. In effect, borrowers who attend certain MSIs are penalized simply because of where they went to college.

EDUCATIONAL REDLINING

2020

**Fintech Lender: Upstart Network, Inc.**

Product: Private Student Loan Refinance

**LOAN  
AMOUNT  
\$30K****Borrower Profile****New York  
University***(Non-MSI)*Major: Computer science  
Occupation: Financial analyst  
Annual income: \$50,000**LOAN OFFERS**Loan interest rate: 16.34% APR  
Origination fee: \$1,231**Total Cost:  
\$42,288****Howard  
University***(HBCU)*Major: Computer science  
Occupation: Financial analyst  
Annual income: \$50,000**LOAN OFFERS**Loan interest rate: 21.29% APR  
Origination fee: \$1,960**Total Cost:  
\$45,785****New Mexico  
State University***(HSI)*Major: Computer science  
Occupation: Financial analyst  
Annual income: \$50,000**LOAN OFFERS**Loan interest rate: 19.23% APR  
Origination fee: \$1,862**Total Cost:  
\$44,011****HBCU Penalty: +\$3,499****HSI Penalty: +\$1,724****Student Populations****New York  
University****Howard  
University****New Mexico State  
University**

■ White 
 ■ Black/African American 
 ■ Latinx/Hispanic 
 ■ Asian 
 ■ Other/Unknown 
 ■ Non-Resident Alien 
 ■ Native American

Demographic data from the U.S. Dept of Education



- **Howard University graduates are charged \$3,499 more than similarly situated NYU graduates.** Over a three-year repayment term, the NYU graduate would pay \$35,093, while the Howard graduate would pay \$35,676. The disparity increases over a five-year repayment term (another repayment term offered by Upstart), with the NYU and Howard borrowers paying \$42,287 and \$45,785, respectively.
- **Howard University graduates are charged an additional \$729 in origination fees than similarly situated borrowers who attended NYU.<sup>1</sup>** In this example, Howard borrowers would pay \$1,960 to originate a loan with a five-year repayment term, whereas the NYU borrowers would pay \$1,231 to originate a loan for the same repayment term. Likewise, for a three-year loan term, Howard borrowers would pay \$1,624 in origination fees, as compared to \$1,292 for NYU borrowers.
- **New Mexico State University (NMSU) graduates are charged nearly \$1,724 more than otherwise identical NYU graduates.** Over a five-year repayment term, a NMSU graduate with a \$30,000 student loan refinancing product would pay \$44,011 in lifetime loan costs, while the otherwise identical NYU graduate would pay \$42,287. This includes the NMSU graduate being charged \$632 more in origination fees.

<sup>1</sup> Note that all loan applicants are modeled as requesting a \$30,000 loan refinancing product, which includes all relevant origination fees already added to the loan amount. These origination fees vary across applicants, with Upstart quoting different fee amounts for different applicants. This variance implies that while the overall loan amounts compared here are the same, the proportion of the refinancing product actually applied to underlying student loans differs, with borrowers who face higher origination fees applying less of their \$30,000 refinancing product to their outstanding student loans. The present estimates of disparities in the cost of refinancing are floor estimates, and students charged higher origination fees (that is, borrowers at HBCUs and HSIs) would need to take out larger loans to refinance the same dollar value of student loans.

## Recommendations

The following recommendations to Congress, regulators, and industry highlight opportunities to address the issues outlined in this report. The industry practices discussed in detail above potentially violate a range of federal and state fair lending and consumer protection laws. More broadly, these practices may further perpetuate inequality, creating new barriers to building wealth for families across the country.

By taking immediate action, stakeholders can address the serious legal issues and far-reaching economic consequences presented by the use of education data in consumer lending.

### **Recommendation 1: Congress should scrutinize the use of education data in consumer lending and the No-Action Letter issued by the Consumer Financial Protection Bureau to Upstart.**

In 2007, then-New York Attorney General Andrew Cuomo explained to Congress that the use of education data in consumer lending posed significant risks to borrowers of color, warning that the specter of “educational redlining” warranted immediate attention from lawmakers.<sup>65</sup>

The findings of this report demonstrate the prescience of Cuomo’s warning. Big banks and fintech “innovators” are embracing education data when making new consumer loans. In doing so, these companies may be unlawfully discriminating against people of color and exacerbating economic inequality. Given the economic consequences potentially posed by a market-wide embrace of education data in consumer lending, Congress should deploy its full suite of investigatory, oversight, and legislative tools to protect consumers.

As part of this coordinated, market-wide oversight, Congress should investigate the CFPB’s handling of the 2017 No Action Letter awarded to Upstart. As described above, in 2017 the CFPB issued its first No-Action Letter (NAL) to fintech lender Upstart, pledging not to enforce federal fair lending laws so long as the company provides regular data about the company’s business practices to the Bureau. The preceding

case study, constructed using Upstart's own marketing materials, plainly illustrates the potential for racial disparities in credit pricing as a result of Upstart's lending practices. As Upstart expands the licensing of its underwriting algorithm to other financial services companies, scrutiny of these practices is even more important.

Congress should immediately demand the following historical data from Upstart to assess whether CFPB's 2017 NAL is consistent with the law and meets the needs of consumers, industry, and the marketplace:<sup>III</sup>

- Upstart's overall loan approval (expressed in dollars lent as well as consumers served) and denial rates for loans made using non-individualized education data (*e.g.*, school, school sector, major) in the underwriting process.
- Upstart's loan approval and denial rates where a consumer indicates that he or she attended an institution of higher education enrolling populations with significant percentages of undergraduate minority students.<sup>56</sup>
- Upstart's loan approval and denial rates where a consumer indicates that he or she attended an institution of higher education other than one enrolling populations with significant percentages of undergraduate minority students.<sup>57</sup>
- Upstart's loan approval and denial rates where a consumer indicates that he or she attended a community college.
- Upstart's loan approval and denial rates where a consumer indicates that he or she attended an institution of higher education other than a community college.
- Upstart's interest rate spread (25th percentile, median, 75th percentile) for loans made using non-individualized education data (*e.g.*, school, school sector, major) in the underwriting process.
- Upstart's interest rate spread (25th percentile, median, 75th percentile) where a consumer indicates that he or she attended an institution of higher education enrolling populations with significant percentages of undergraduate minority students.<sup>58</sup>

III To date, little public information has been produced by the CFPB about Upstart's disclosures to the Bureau under its NAL agreement. The limited disclosures made by the CFPB appear to have been based on a simulation, comparing Upstart's approach to underwriting and pricing against a hypothetical model that relies on FICO score. This approach is seriously flawed. It fails to isolate the effects of educational data on protected classes of borrowers when similarly-situated Upstart customers are compared to one another. The flaws in this design suggest a path forward for Congressional investigators—by demanding the production of data that allows for an apples-to-apples comparison across Upstart's existing portfolio of customers, including data on approvals and denials specific to each college or university attended by an Upstart customer, Congress can more accurately assess whether Upstart's approach to underwriting or pricing loans has a disparate impact. See Consumer Financial Protection Bureau, *An update on credit access and the Bureau's first No-Action Letter* (August 2019), <https://www.consumerfinance.gov/about-us/blog/update-credit-access-and-no-action-letter>.



- Upstart's interest rate spread (25th percentile, median, 75th percentile) where a consumer indicates that he or she attended an institution of higher education other than one enrolling populations with significant percentages of undergraduate minority students.<sup>69</sup>
- Upstart's interest rate spread (25th percentile, median, 75th percentile) where a consumer indicates that he or she attended a community college.
- Upstart's interest rate spread (25th percentile, median, 75th percentile) where a consumer indicates that he or she attended an institution of higher education other than a community college.

Should information produced by Upstart demonstrate that the company's practices have a disparate impact on protected classes with respect to the cost of credit, or offer evidence that Upstart's approach to consumer lending perpetuates economic inequality, Congress should immediately clarify to the CFPB that these outcomes are inconsistent with the intent behind the No-Action Letter Program. Further, Congress may wish to consider new legislation to prohibit the CFPB from waiving the Equal Credit Opportunity Act (ECOA) for any companies seeking a No-Action Letter in the future, narrowing the scope of CFPB's authority to issue these types of letters.

**Recommendation 2: Federal and state financial regulators should prioritize oversight of the use of education data in underwriting to ensure lenders comply with fair lending laws.**

Federal and state financial regulators supervise compliance with and enforce fair lending laws. Regulated financial institutions include both large banks like Wells Fargo and nonbank specialty consumer lenders like Upstart. Based on the findings of this report, federal and state financial regulators should prioritize the oversight of consumer lending where regulated entities use education data in underwriting or pricing credit.

**Federal financial regulators, including prudential regulators and the CFPB, should examine the use of education criteria in lending decisions by big banks and nonbank consumer lenders.** Federal regulators, including the Office of the Comptroller of the Currency (OCC), the Federal Reserve Board, the Federal Deposit Insurance Corporation (FDIC), the Federal Trade Commission (FTC), and the CFPB, oversee or enforce laws that may apply to the use of education data in consumer lending. In particular, these regulators may enforce ECOA, which prohibits certain types of discrimination in the extension of credit.<sup>60</sup> As the first case study in this report demonstrates, large regulated financial institutions may use education data when determining access to credit or pricing financial products, despite the fair lending compliance

risks it creates for these financial institutions.<sup>61</sup> This report offers ample evidence to suggest Wells Fargo's consumer lending practices, in particular, create risks for protected classes of consumers.

There is recent precedent for the CFPB and other regulators to consider the use of non-individualized education data as a fair lending compliance risk for financial institutions. In 2012, the CFPB studied the use of schools' Cohort Default Rate (CDR) in private student lending, finding that, "[g]enerally . . . lenders' consideration of CDR in either school eligibility or underwriting and pricing criteria may reduce credit access and increase costs disproportionately for minority borrowers."<sup>62</sup>

Following publication of the 2012 report, the CFPB incorporated this finding into its examination procedures by instructing examiners to consider the use of CDR when evaluating both bank and nonbank private student lenders for compliance with ECOA. Shortly thereafter, the FDIC took an enforcement action against Sallie Mae Bank for violating ECOA by using this particular piece of education data in underwriting and pricing private student loans.<sup>63</sup>

Based on the evidence presented in this report, other regulators should adopt the same approach as the FDIC—prioritizing scrutiny of these practices across the financial services sector and taking enforcement actions where appropriate.

**States should prioritize action to stamp out educational redlining when overseeing consumer lending by banks and nonbanks.** Since 2017, the CFPB has ceased to bring new enforcement actions policing discrimination in the financial sector, drawing criticism from state law enforcement officials, civil rights groups, and Members of Congress for failing to appropriately administer the nation's fair lending laws.<sup>64</sup> Fortunately for consumers, the Dodd-Frank Act empowers state attorneys general and state banking regulators to enforce these laws with respect to the companies they regulate. This authority presents an opportunity for state officials to scrutinize the use of education data in consumer lending within their states, stepping in where the CFPB has recently failed to act.

In addition, states may enforce and administer a wide range of state civil rights and anti-discrimination statutes. Evidence suggests that some states are already beginning to scrutinize these entities for violations of state law. As part of any expanded state oversight effort, state regulators and law enforcement should scrutinize Upstart's practices for compliance with these state fair lending laws in the context of the CFPB's Upstart No-Action Letter.

**Recommendation 3: Consumer lenders, including banks and fintech specialty lenders, should regularly publish information on underwriting decisions and pricing that relies on education data.**

Banks and specialty lenders such as Wells Fargo and Upstart that use education data in their underwriting decisions should make available data on the impact of these criteria on access to credit (including both approvals and denials) and on pricing of loans for consumers. This information should track access and pricing both for borrowers who attend minority-serving institutions and for borrowers who attend non-minority serving institutions. This additional information about credit decisioning and pricing should be made available to the public at large, including stakeholders inside and outside of government, through publication on the lender's website and disclosure at the time of application. For this public disclosure to be effective, it should include data that allows for comparison across a company's existing portfolio of customers, including data on approvals and denials specific to each college or university attended by an applicant for credit.

By embracing new transparency with respect to the effects of education data on lending, market participants can empower borrowers to shop for financial products with an accurate understanding of the costs and risks associated with each product. Further, such transparency efforts will empower federal and state regulators to perform more effective oversight over the industry.

## Conclusion

Communities of color have historically been locked out of mainstream credit markets. But while companies tout the use of education-based criteria in underwriting as a means to broaden credit access for marginalized consumers, the use of such factors may actually undermine equitable access to credit. Indeed, by creating situations where protected classes of consumers are offered less favorable credit terms, the use of education data in credit underwriting decisions can reinforce systemic barriers to economic opportunity.

Discrimination in consumer credit markets is not new. But as this analysis shows, the use of education data in underwriting could charge borrowers more for a loan simply for choosing the most accessible path for pursuing the American Dream. Is this what is meant by a mission of 'innovation'? Access to credit should not simply mean 'more people getting more loans.' It is imperative to examine the variance in the cost of those loans. Otherwise, expanded access to credit will not expand equity.

With mortgage redlining, borrowers are given worse loans simply because of who their neighbor is. Now, with educational redlining, borrowers are given worse loans simply because of who is sitting next to them in the classroom. Just as law enforcement took action against mortgage redlining, they must do the same with education redlining. Innovation should not re-package age-old discrimination. Rather, true innovation should provide a means to equitably broaden credit access for historically marginalized communities.



## Endnotes

- <sup>1</sup> Charles Davidson, *Lack of Access in Financial Services Impedes Economic Mobility*, Fed. Res. Bank of Atlanta Economy Matters (Oct. 18, 2018), <https://www.frbatlanta.org/economy-matters/community-and-economic-development/2018/10/16/lack-of-access-to-financial-services-impedes-economic-mobility>; see also Christopher K. Odinet, *The New Data of Student Debt*, 92 S. Cal. L. Rev. 1617, 1673 (Dec. 2019).
- <sup>2</sup> *Report on the Economic Well-Being of U.S. Households in 2016-May 2017*, Fed. Res. Bank (June 14, 2017), <https://www.federalreserve.gov/publications/2017-economic-well-being-of-us-households-in-2016-banking-credit.htm>.
- <sup>3</sup> Lori Teresa Yearwood, *Many minorities avoid seeking credit due to generations of discrimination. Why that keeps them back*, CNBC (Sept. 6, 2019), <https://www.cnbc.com/2019/09/01/many-minorities-avoid-seeking-credit-due-to-decades-of-discrimination.html>; Lisa Rice, *Missing Credit: How the U.S. Credit System Restricts Access to Consumers of Color*, Nat'l Fair Hous. All. (Feb. 26, 2019), <https://financialservices.house.gov/uploadedfiles/hhrg-116-ba00-wstate-ricel-20190226.pdf>; Bradley L. Hardy et al., *The Historical Role of Race and Policy for Regional Inequality*, The Hamilton Project, at 8 (Sept. 2018), [https://www.hamiltonproject.org/assets/files/PRP\\_HardyLoganParman\\_1009.pdf](https://www.hamiltonproject.org/assets/files/PRP_HardyLoganParman_1009.pdf); Tracy Jan, *Redlining was banned 50 years ago. It's still hurting minorities today*, Wash. Post (Mar. 28, 2018), <https://www.washingtonpost.com/news/work/wp/2018/03/28/redlining-was-banned-50-years-ago-its-still-hurting-minorities-today>; Danyelle Solomon et al., *Systemic Inequality: Displacement, Exclusion, and Segregation: How America's Housing System Undermines Wealth Building in Communities of Color*, Ctr. for Am. Progress, 4-10 (Aug. 2019), <https://cdn.americanprogress.org/content/uploads/2019/08/06135943/StructuralRacismHousing.pdf>.
- <sup>4</sup> Paul Hancock et al., *Supreme Court vs. HUD: The Race to Decide "Impact or Intent"*, K&L Gates (Nov. 17, 2011), <http://www.klgates.com/emsupreme-court-vs-hudem--the-race-to-decide-impact-or-intent-11-17-2011>; see also Lisa Rice & Deidre Swesnik, *Discriminatory Effects of Credit Scoring on Communities of Color*, Suffolk L. Rev. (Dec. 19, 2013), [http://suffolkreview.org/wp-content/uploads/2014/01/Rice-Swesnik\\_Lead.pdf](http://suffolkreview.org/wp-content/uploads/2014/01/Rice-Swesnik_Lead.pdf).
- <sup>5</sup> See, e.g., Upstart, *Upstart Receives First No-Action Letter Issued by Consumer Financial Protection Bureau* (Sept. 14, 2017), <https://www.upstart.com/blog/upstart-receives-first-no-action-letter-issued-consumer-financial-protection-bureau> [hereinafter *Upstart Release*].
- <sup>6</sup> See Aryn Bussey, *Educational Redlining? The use of education data in underwriting could leave HBCU and MSI graduates in the dark*, Student Borrower Prot. Ctr (July 24, 2019), <https://protectborrowers.org/educational-redlining>.
- <sup>7</sup> See, e.g., *Upstart Release*, *supra* note 5.
- <sup>8</sup> See Bussey, *supra* note 6.
- <sup>9</sup> See *id.*
- <sup>10</sup> See *id.*
- <sup>11</sup> A series of lenders were examined for their use of education-based data in underwriting decisions. In particular, Wells Fargo Bank and Upstart were selected for the availability of their products' rate calculation without a credit check (Wells Fargo) and only a soft credit inquiry (Upstart).
- <sup>12</sup> See *Who are the credit invisibles?*, Consumer Fin. Prot. Bureau (CFPB) (Dec. 2016), [https://files.consumerfinance.gov/f/documents/201612\\_cfpb\\_credit\\_invisible\\_policy\\_report.pdf](https://files.consumerfinance.gov/f/documents/201612_cfpb_credit_invisible_policy_report.pdf) [hereinafter *Credit Invisibles*].
- <sup>13</sup> See, e.g., *Upstart Release*, *supra* note 5.

- 14 See Brian Kreiswirth et al., *Using alternative data to evaluate creditworthiness* (Feb. 16, 2017), <https://www.consumerfinance.gov/about-us/blog/using-alternative-data-evaluate-creditworthiness/>.
- 15 See *Credit Invisibles*, *supra* note 12, at 5.
- 16 See Dowse B. Rustin IV et al., *Pricing without Discrimination: Alternative Student Loan Pricing, Income-Share Agreements, and the Equal Credit Opportunity Act*, Am. Enter. Inst. (Feb. 2017), <https://www.aei.org/wp-content/uploads/2017/02/Pricing-Without-Discrimination.pdf>.
- 17 Odinet, *supra* note 1, at 1644-1648.
- 18 Odinet, *supra* note 1, at 1645.
- 19 See, e.g., *Upstart Release*, *supra* note 5.
- 20 *Examining the Use of Alternative Data in Underwriting and Credit Scoring to Expand Access to Credit: Hearing Before the H. Task Force on Financial Technology*, 116th Cong. 7 (2019) (statement of Chi Chi Wu, Staff Attorney, National Consumer Law Center).
- 21 See Jeremy Ashkenas et al., *Even with Affirmative Action, Blacks and Hispanics are More Underrepresented at Top Colleges Than 35 Years Ago*, N.Y. Times (Aug. 24, 2017), <https://www.nytimes.com/interactive/2017/08/24/us/affirmative-action.html>.
- 22 See, e.g., Daniel Aaronson et al., *The Effects of the 1930s HOLC "Redlining" Maps*, Fed. Res. Bank of Chicago (Feb. 2019); see also Hannah Fry, *Lori Loughlin's daughters remain at USC amid college admissions scandal*, L.A. Times (Mar. 26, 2019), <https://www.latimes.com/local/lanow/la-me-ln-college-cheating-giannulli-20190326-story.html>.
- 23 Bussey, *supra* note 6.
- 24 Despite a half-century ban, redlining based on geography persists in cities across the country. In the early 1930s, the federal government made efforts to steady the nation's housing market by making changes to valuation assessments. Aaronson et al., *supra* note 22, at 3. The government classified neighborhoods by their relative lending risks; in addition to housing-based characteristics such as price, housing age, and quality, homes were also classified by race and ethnicity. *Id.* Predominantly black neighborhoods were the lowest-rated, and thus those neighborhoods were denied access to credit. *Id.* Although the Fair Housing Act of 1968 bars housing discrimination, redlining and the residual effects of residential segregation persist. Sam Fullwood III, *The United States' History of Segregated Housing Continues to Limit Affordable Housing*, Ctr. for Am. Progress (Dec. 15, 2016) <https://www.americanprogress.org/issues/race/reports/2016/12/15/294374/the-united-states-history-of-segregated-housing-continues-to-limit-affordable-housing>. Residential segregation of African Americans remains high, particularly in large urban areas. Douglas S. Massey, *American Apartheid: Segregation and the Making of the Underclass*, Am. J. of Sociology, Vol. 96, No. 2, 329-357, 354 (Sept. 1990).
- 25 A 2018 analysis by the Center for Investigative Reporting found race-based lending disparities in 61 metropolitan areas across the country. Aaron Glantz & Emmanuel Martinez, *Modern-day redlining: How banks block people of color from homeownership*, Chicago Tribune (Feb. 17, 2018) <https://www.chicagotribune.com/business/ct-biz-modern-day-redlining-20180215-story.html>.
- 26 See Kevin Drawbaugh, *Lawmakers Quiz Student Lenders on 'redlining'*, Reuters (June 7, 2007) <https://www.reuters.com/article/us-studentloans-congress/lawmakers-quiz-student-lenders-on-redlining-idUSN0724029120070607>.
- 27 See Assoc. Press, *Cuomo charges 'redlining' in student loan probe*, NBC News (June 19, 2007), [http://www.nbcnews.com/id/19316230/ns/business-personal\\_finance/t/cuomo-charges-redlining-student-loan-probe/#.XJhm9GhKhPY](http://www.nbcnews.com/id/19316230/ns/business-personal_finance/t/cuomo-charges-redlining-student-loan-probe/#.XJhm9GhKhPY).
- 28 See Drawbaugh, *supra* note 26.
- 29 See, e.g., Kreiswirth, *supra* note 14; Rustin, *supra* note 16.
- 30 George Kuh et al., *What Matters to Student Success: A Review of the Literature at 77*, Nat. Postsecondary Educ. Coop. (July 2006), [https://nces.ed.gov/npec/pdf/kuh\\_team\\_report.pdf](https://nces.ed.gov/npec/pdf/kuh_team_report.pdf).
- 31 Rustin, *supra* note 16.
- 32 See also Drawbaugh, *supra* note 26.
- 33 See Jennifer Ma & Sandy Baum, *Trends in Community Colleges: Enrollment, Prices, Student Debt, and Completion*, College Bd. Research (Apr. 2016), <https://research.collegeboard.org/pdf/trends-community-colleges-research-brief.pdf>.

- 34 See *id.*
- 35 See *id.*
- 36 See Wells Fargo Private Student Loans, Wells Fargo, <https://www.wellsfargo.com/student/> (last visited Feb. 3, 2020).
- 37 See Wells Fargo Undergraduate Private Student Loans, Wells Fargo, <https://www.wellsfargo.com/student/collegiate-loans/> (last visited Jan. 30, 2020).
- 38 See Wells Fargo Private Loans for Community College, Wells Fargo, <https://www.wellsfargo.com/student/community-college-loans/> (last visited Jan. 30, 2020).
- 39 See *Today's Rates*, Wells Fargo, <https://wffs.wellsfargo.com/terms/TodaysRates> (last visited Feb. 3, 2020).
- 40 See also Ellen Wexler, *Geography Matters*, Inside Higher Ed (Feb. 3, 2016), <https://www.insidehighered.com/news/2016/02/03/when-students-enroll-college-geography-matters-more-policy-makers-think> ("At public four-year colleges, the median distance students live from home is 18 miles. That number is 46 miles for private nonprofit four-year colleges, and only eight miles at public two-year colleges.").
- 41 See U.S. Dep't of Ed., *Characteristics of Minority-Serving Institutions and Minority Undergraduates Enrolled in These Institutions*, Nat'l Ctr. for Ed. Statistics, U.S. Dep't of Ed. (Nov. 2007), <https://nces.ed.gov/pubs2008/2008156.pdf>.
- 42 Walter R. Allen, *Foreword to Understanding Minority-Serving Institutions*, xv-xvi, xv-xix (Marybeth Gasman et al. eds., State University of New York Press) (2008).
- 43 See Alisa Cunningham et al., *Issue Brief: Minority-Serving Institutions Doing More with Less*, Inst. for Higher Ed. Policy (Feb. 2014), <https://vtechworks.lib.vt.edu/bitstream/handle/10919/83120/MinorityServingInstitutions.pdf?sequence=1&isAllowed=y>.
- 44 See Josh Mitchell & Andrea Fuller, *The Student-Debt Crisis Hits Hardest at Historically Black Colleges*, Wall St. J. (Apr. 17, 2018), <https://www.wsj.com/articles/the-student-debt-crisis-hits-hardest-at-historically-black-colleges-11555511327>.
- 45 See Upstart Network, <https://www.upstart.com/about#who-we-are> (last visited Jan. 15, 2020).
- 46 See *id.*
- 47 See *Upstart Release*, *supra* note 5.
- 48 See *id.*
- 49 See CFPB Announces First No-Action Letter to Upstart Network, Consumer Fin. Prot. Bureau (Sept. 14, 2017), <https://www.consumerfinance.gov/about-us/newsroom/cfpb-announces-first-no-action-letter-upstart-network/>.
- 50 *Id.*
- 51 *Id.*
- 52 See generally Kenneth P. Brevoort & Michele Kambara, *CFPB Data Point: Becoming Credit Visible*, Consumer Fin. Prot. Bureau (June 2017), [https://files.consumerfinance.gov/f/documents/BecomingCreditVisible\\_Data\\_Point\\_Final.pdf](https://files.consumerfinance.gov/f/documents/BecomingCreditVisible_Data_Point_Final.pdf) (finding that the majority of consumers are credit visible by age 25).
- 53 *Compare Rates With Our Loan Calculator*, Upstart, <https://www.upstart.com/blog/compare-rates-loan-calculator> (last visited Feb. 3, 2020).
- 54 See United States Department of Education *Lists of Postsecondary Institutions Enrolling Populations with Significant Percentages of Undergraduate Minority Students*, U.S. Dep't of Educ. (accessed Feb. 3, 2020), <https://www2.ed.gov/about/offices/list/ocr/edlite-minorityinst.html> [hereinafter *MSI List*].
- 55 See Drawbaugh, *supra* note 26.
- 56 See *MSI List*, *supra* note 54.
- 57 See *id.*
- 58 See *id.*
- 59 See *id.*
- 60 See 15 U.S.C. §§ 1691 et seq.; see also *Equal Credit Opportunity Act*, CFPB Consumer Laws and Regulations, Consumer Fin. Prot. Bureau, [https://files.consumerfinance.gov/f/201306\\_cfpb\\_laws-and-regulations\\_ecoa-combined-june-2013.pdf](https://files.consumerfinance.gov/f/201306_cfpb_laws-and-regulations_ecoa-combined-june-2013.pdf).



- <sup>61</sup> See e.g., *Equal Credit Opportunity Act*, CFPB Consumer Laws and Regulations, Consumer Financial Protection Bureau, [https://files.consumerfinance.gov/f/201306\\_cfpb\\_laws-and-regulations\\_ecoa-combined-june-2013.pdf](https://files.consumerfinance.gov/f/201306_cfpb_laws-and-regulations_ecoa-combined-june-2013.pdf).
- <sup>62</sup> *Fair Lending Report of the Consumer Financial Protection Bureau*, Consumer Fin. Prot. Bureau (Dec. 2012), [https://files.consumerfinance.gov/f/201212\\_cfpb\\_fair-lending-report.pdf](https://files.consumerfinance.gov/f/201212_cfpb_fair-lending-report.pdf).
- <sup>63</sup> See FDIC Announces Settlement with Sallie Mae for Unfair and Deceptive Practices and Violations of the Servicemembers Civil Relief Act, Fed. Deposit Ins. Corp. (May 13, 2014), <https://www.fdic.gov/news/news/press/2014/pr14033.html>.
- <sup>64</sup> See Kate Berry, *Where Have all the CFPB Fair-Lending Cases Gone?*, Am. Banker (Dec. 16, 2019) <https://www.americanbanker.com/news/where-have-all-the-cfpb-fair-lending-cases-gone>.





[PROTECTBORROWERS.ORG](https://PROTECTBORROWERS.ORG)



**To:** House Financial Services Committee Task Force on Artificial Intelligence

**From:** Upstart Network, Inc.

**RE:** Report on Upstart by Student Borrower Protection Center

---

Student Borrower Protection Center (SBPC) released a report on February 4, 2020 claiming that Upstart's model is biased. Upstart would like to respectfully share its viewpoints on this study.

**Traditional lending models based on credit score and income are significantly biased**

- Traditional credit models based on factors like FICO and income disadvantage millions of Americans, disproportionately affecting Hispanics, African-Americans and women. Improving access necessarily requires data beyond those traditional factors.
- According to the Federal Reserve<sup>1</sup>, traditional credit scores classify more than 3 times as many blacks (53%) and almost two times as many Hispanics (30%) as whites (16%) into the lowest two deciles of credit scores.

**Use of alternative data improves outcomes for all race, gender, and age segments**

- According to the CFPB<sup>2</sup>, for borrower applicants during 2018, Upstart's AI model increased access to credit across all tested race, ethnicity, and gender segments by 23-29% while also decreasing average rates by 15-17%.
- In 2018, Upstart increased approval rates for African-Americans by 28% with 17% lower APRs.
- More recent results are even more encouraging: 2019 results showed that Upstart's model increased approval rates for African-American applicants by more than 45% with 21% lower APRs.

---

<sup>1</sup> <https://www.federalreserve.gov/boarddocs/rptcongress/creditscore/performance.htm#toc9.2>

<sup>2</sup> <https://www.consumerfinance.gov/about-us/blog/update-credit-access-and-no-action-letter/>

- In 2019, near-prime borrowers on Upstart (620-660 FICO) saw more than twice the approval rates with 25% lower APRs compared to traditional models.

**All consumer lending should be subject to rigorous and regular testing for bias**

- All credit models, regardless of the variables used, should be tested for bias using a rigorous and complete methodology on a recurring basis.
- Upstart systematically conducts fair lending testing, in accordance with our CFPB No-Action Letter, on millions of real applicants on a quarterly basis.

**SBPC's study was significantly flawed and misrepresented its findings**

- SBPC's study was based on a single individual misrepresenting his education 26 times between November 14, 2019 and February 3, 2020 to request a rate on Upstart.com. In reality, the applicant is a 24-year old graduate of a top-ten school, with a 787 FICO score and no student debt.
- SBPC's claim that the loan applications were identical except for the applicant's college attended is inaccurate and misleading. The 26 applications had many differences among them, including changes to the applicant's credit report. These differences affected the interest rates offered.
- Even though the hypothetical Howard University graduate received a better rate than more than half the other hypothetical applicants, the study chose to selectively cite an example where the hypothetical Howard graduate's APR was higher than one from a different school.
- Upstart's model does not consider the specific institution that any applicant attended. Rather, it groups schools by academic and economic-outcome characteristics.
- In Upstart's model, Howard University is virtually identical to more than 200 other institutions. The characteristics of these institutions are stronger than those of the vast majority of more than 5000 universities in the US.
- Finally, if SBPC is interested in conducting an academically rigorous and comprehensive review of our model's outcomes for borrowers, Upstart is willing to collaborate with them. In any case, this flawed "report" should be retracted and replaced with a rigorous academic-style study.

