Tulane University
LAW SCHOOL

Written Testimony of

Kristin N. Johnson, McGlinchey Stafford Professor of Law and Associate Dean of Faculty Research, Tulane University Law School

Before the United States House Committee on Financial Services
Task Force on Financial Technology

Examining the Use of Alternative Data in Underwriting
and Credit Scoring to Expand Access to Credit

Thursday, July 25, 2019
2128 Rayburn House Office Building

The Future of Finance:
Alternative Data in Credit Underwriting

Chairwoman Waters, Ranking Member McHenry, Stephen Lynch, Chair of the Financial Technology Task Force, and Bill Foster, Chair of the Task Force on Artificial Intelligence, Members of the Committee and Members of the Task Force:

Thank you for inviting me to participate in this Hearing to discuss the use of alternative data in credit underwriting and credit scoring to expand access to credit.

I am a professor of law and associate dean of faculty research at Tulane University Law School where I teach courses on corporate and securities law, the integration of emerging technologies into financial markets, systemic risk across financial markets and financial markets regulation.[1] I am an affiliate of the Murphy Institute at Tulane and director of the Institute's Financial Market Stability Program.[2] I am here today solely in my academic capacity and am not testifying on behalf of any entity.

Over the last several years, a body of sophisticated algorithms commonly described as artificial intelligence ("AI") and distributed ledger technologies have altered the financial market ecosystem, creating a new class of financial institutions – fintech firms.[3]

[1] I have previously served as an analyst at Goldman Sachs, a Vice President and Associate General Counsel at JP Morgan and an associate at a New York law firm supporting many of the largest financial institutions in the country. For the last decade, as an academic, I have published research examining the risk management implications of nascent financial products and services in credit and capital markets.

[2] The Murphy Institute is a privately-funded, interdisciplinary center at Tulane University that aims to support and advance applied research in public policy, public affairs and civic engagement and to inspire and educate students and the interested general public in the understanding and analysis of challenging economic, moral, and political questions.

[3] The discussion presented here appears in forthcoming academic journal manuscripts.

Supplementing traditional credit underwriting data inputs and processes, fintech firms employ newer modeling techniques and consider a broader range of source data referred to descriptively (rather than normatively) as alternative data. These new inputs include information regarding consumers' financial transactions, recurring payments history and a behavioral score based on social networking and digital-interface. Fintech firms include both the non-depository digital platforms that operate independently and platforms that partner with legacy banks to originate loans.[4] Fintech firms servicing credit scoring and underwriting markets offer great promise but also present unique concerns.[5]

The introduction of alternative data may improve access to credit for many consumers with nonexistent or insufficient credit histories. According to estimates, twenty-six million Americans do not have traditional credit histories and are considered "credit invisible." Another nineteen million Americans have thin (limited), impaired or stale (outdated) credit histories and, as a result, cannot obtain credit scores using traditional scoring methodologies ("credit unscorable").

Unsavory lending practices, detestable marketing tactics and usurious interest rates have too often plagued marginalized consumers who face persistently fragile financial circumstances.[6] Unlike legacy credit scoring businesses such as Equifax, Experian and Transunion that rely on commercially available credit scoring models like the Fair Isaac Corporation Lenders ("FICO") methodology fintech firms increasingly rely on alternative credit scoring models and nontraditional source data. According to proponents, the development of nascent methodologies and alternative data enables fintech firms to expand access to credit to consumers historically deemed invisible or unscorable.

Legislative and regulatory authorities must, however, balance fintech firms' laudable promises of greater inclusion with the significant risks posed by integrating alternative data and new methodologies. Careful examination of the rise of alternative data and the evolution in consumer credit underwriting methods casts a spotlight on fintech firms' promises of inclusion and reveals the perils of relying on source data that may not be demonstrably predictive of creditworthiness as well as the potential for predatory or discriminatory practices to undermine the anticipated benefits of alternative source data and credit evaluation processes.

Fintech firms integrating alternative data and modeling techniques must satisfy long-standing fairness and accountability standards, engage in responsible innovation and commit to provide sufficient transparency, meaningful disclosure, auditing and necessary

[4] Christopher K. Odinet, *Consumer Bitcredit and Fintech Lending*, 69 ALA. L. REV. 781 (2018)
[5] FED. TRADE COMM'N, BIG DATA: A TOOL FOR INCLUSION OR EXCLUSION?, at i, (2016).
[6] Kathleen C. Engel & Patricia A. McCoy, *A Tale of Three Markets: The Law and Economics of Predatory Lending,* 80 TEX. L. REV. 1255, 1261 (2002).

internal controls to meet statutory obligations regarding their methodologies and minimize the potential for discriminatory effects on legally protected classes.

Fintech Firms and Learning Algorithms

"Fintech" is a catch-all term used to refer to the digital platform or internet-based financial services firms that engage in digital transfers, storage, payments systems, digital asset origination (such as cryptocurrency) and secondary market trading, investment advising and digital credit scoring and origination. To capitalize on economic efficiencies, reduce transaction costs and mitigate commonly-identified enterprise risks, fintech firms integrate artificial intelligence technologies such as supervised or unsupervised machine learning, deep learning or neural networks ("AI") or distributed ledger technologies into their business models. While there is no universally adopted definition of AI, the term refers to a diverse, but related, set of technologies that train through a reinforcement learning process, simulate human decision-making and cognitive behavior and engage in predictive analysis.[7]

Advancements in the collection, storage and analysis of vast volumes of data ("big data") fuel AI platforms designed to automate decision-making in several key sectors including healthcare, education, employment, criminal law, security, surveillance, communications and finance. While the inclusion of data crunching algorithms is nothing new – investment banking firms, for example, have long relied on sophisticated algorithms to predict timing, pricing, risk and other factors that influence investment and trading decisions - the rapid adoption of learning algorithms that interpret alternative data in consumer credit markets presents significant risks.

Automating Credit Decisions

Learning algorithms at the center of fintech platforms' credit evaluation processes analyze vast quantities of data in fractions of a second. Fintech platforms replace face-to-face meetings with loan officers and cumbersome and time-consuming paper-based credit application processes with applications accessible on internet-enabled smartphones, tablets and other mobile or personal devices. Removing human underwriting agents and their biases arguably reduces the likelihood of intentional discrimination. AI-based credit scoring methodologies may enhance consumer default predictions and lead to better credit classification and possibly lower-priced credit than traditional credit scoring methodologies. Together these process-oriented improvements enhance efficiency and accuracy, improve pricing, reduce operating and loan origination costs and enable fintech

[7] Examples of AI modeling techniques include but are not limited to decision trees, random forests, artificial neural networks, k-nearest neighbors, genetic programming, and "boosting" algorithms. Given the limited time available and suggested scope, only an abbreviated description of artificial intelligence and other referenced technologies appears in the submitted written testimony.

firms to offer credit to a greater diversity of consumers, in particular those who have struggled to obtain credit.

Traditional credit evaluation processes like FICO consider tradeline information, including but not limited to existing and previous loan obligations, repayment history, credit limits, account status for revolving accounts, credit inquiries, public records such as civil judgments, tax liens and bankruptcies. Incumbent credit scoring methodologies predominantly use multivariate regression analysis to correlate past credit history to consumer credit outcomes and evaluate the likelihood of default or delinquency. Increasingly, incumbent credit scoring firms and traditional methodologies are shifting their evaluation criteria. As fintech firms tout the benefits of AI driven decision-making, both incumbent credit scoring firms and insurgent fintech platforms rely on alternative sources of data and scoring methodologies.

Alternative Data

According to industry and federal and state agency reports, alternative data refers to information not traditionally used by the national consumer reporting agencies ("CRA") in calculating a consumer's credit score. In some instances, alternative data simply expands the categories of payment history beyond those considered by CRAs. For example, some fintech platforms integrate telecommunications (mobile phone and cable bills), utilities or residential rental payment history. In other instances, fintech firms expand the types of information considered in credit scoring processes and include financial transaction data (checking account cashflows).

Alternative data may assist historically marginalized (credit invisible and unscorable consumers) to gain access to conventional credit markets. There is good reason to believe that capturing nontraditional data may enable consumers with thin, impaired or nonexistent credit files to demonstrate a history of timely bill payment. The frequency of telecommunications, utility and rental payments may enable consumers to generate a different but valuable track record or consistent, timely bill payment history.

Limitations and conflicts arising from the use of alternative data to expand access to credit. Consumer advocates have, however, expressed some concerns regarding the impact of integrating certain data points, such as utility bill payments. Relying on utility or cable bill payment histories may disadvantage low-income consumers for various reasons. First, dispute resolution processes for public utilities and cable services may differ from other types of recurring obligations. Second, utility bill balances may fluctuate seasonally, prompting some consumers to delay payments or fall behind on pay utilities bills. Low-income or fixed income families are particularly susceptible to these circumstances.

Consider, for example, the families living in areas of the country that face severe seasonal weather patterns. For families living in the northeastern part of the country, for example, home heating bills may present significant monthly demands during the winter and families may not be able to pay utility bills on-time or in full at the close of each billing cycle. Similar challenges may arise for families living in southern states during the summer months. Finally, the significance assigned to recurring residential bills may disadvantage families that migrate seasonally based on employment opportunities or periodically relocate based on service in the armed forces.

Consumer advocates also expressed concerns that H.R. 435 proposed during the 115th Congress would preempt consumer privacy protections by amending Section 623 of the Fair Credit Reporting Act ("FCRA") to permit utilities and landlord to furnish payment information to a CRA "notwithstanding any other provision of law." Without modification, the earlier version of proposed H.R. 435 would override federal requirements that a subsidized housing provider obtain a consumer's consent before sharing rental payment information. The recently proposed Credit Access and Inclusion Act addresses the preemption concerns and solicits a two-year study and report from the Government Accountability Office on the impact of furnishing additional information.

Financial transaction and social networking data. Expanding credit evaluation criteria beyond additional types of recurring payments, alternative data may also include personal consumer financial transaction data – bank account and credit/debit card transactions, including deposits, transfers or withdrawals. Methodologies integrating alternative data may also incorporate educational (major and university attended) or professional accomplishments.

Proponents of alternative data also advocate for the inclusion of nonfinancial, behavioral data. These data points may include digital interface information such as clickstream data, audio and text data, internet browsing and search habits, geo-spatial data and survey or questionnaire data.[8] Beyond simply browsing preferences, fintech firms are also integrating highly-personalized reputational data. For example, fintech firms are assessing consumers' social network status, web-scraping data from consumers' financial transactions and social media activities and ranking consumers based on relational social connections (consumers' status as "social influencers") through analysis of exchanged messages and friends tagged in social media posted photos.

It is not yet clear how these new sources of data will impact those without credit reports or with thin or stale credit files. It is also unclear how credit invisibles and unscorables who do not have conventional checking and savings accounts or credit cards will generate financial transaction data. Similarly, ranking consumers based on higher educational or

[8] Mikella Hurley and Julius Adebayo, *Credit Scoring in the Era of Big Data*, 18 YALE J. L. & TECH. 148, 159 (2016).

professional accomplishments seems likely to replicate the current credit scoring patterns. Finally, credit invisibles and unscorables that lack a presence on social media are unlikely to engender the relational benefits or rewards associated with social networking. In fact, familial and neighborhood associations may make it more difficult for consumers who have not traditionally qualified for credit on fair and reasonable terms to gain access to better, higher quality credit products.

Indisputably, however, the rising significance of alternative data has ignited interest across various markets for greater access to consumer financial data. Consistent with its dominance in the general technology market, Facebook has directly approached banks requesting access to consumers' financial transaction data[9] and registered for a patent for a technology that assesses users based on social network connections. Technology firms often seek to gather sensitive data from consumers but resist transparency regarding the uses of consumer data.[10]

Regulating Alternative Data

The harvesting, distribution and integration of financial transaction and behavioral scoring data raises significant questions regarding consumer protections, privacy and discriminatory practices.

Alternative data such as financial transaction data - credit and debit card and checking account transaction history- may offer valuable insights. Information regarding financial transaction activities and behavior may better inform evaluations of factors that are correlated to consumer credit risk assessment. A consumer's financial history is, however, sensitive information. Unmonitored use and distribution of this information challenges consumer protections and privacy norms.

Privacy Concerns – Existing and Proposed Federal Oversight

A host of state and federal regulators and this Committee are actively seeking to clarify the types of alternative data and the method for including these new class of information in emerging and evolving credit scoring processes. This Committee has held multiple hearings to explore these questions.

More specifically, in February of 2017, the Consumer Financial Protection Bureau ("CFPB") announced a comprehensive Request for Information Regarding Use of Alternative Data and Modeling Techniques in the Credit Process. The Government Accountability Office ("GAO") issued a report in March of 2018 - *Additional Steps by*

[9] Emily Glazer, Deepa Seetharaman and AnnaMaria Andriotis, *Facebook to Banks: Give Us Your Data, We'll Give You Our Users*, WALL ST. J., Aug. 6, 2018.

[10] FRANK PASQUALE, THE BLACK BOX SOCIETY (Harvard Univ. Press, 2015).

Regulators Could Better Protect Consumers and Aid Regulatory Oversight - and a second report in December of 2018 - *Agencies Should Provide Clarification on Lenders' Use of Alternative Data* – recommending a series of policies including proposals to coordinate agencies' regulatory efforts, clarify standards governing alternative data and minimize uncertainty regarding the use of alternative data in the underwriting process. In the absence of effective state or federal regulatory intervention, many warn that fintech firms will take advantage of gaps in oversight and engage in regulatory arbitrage.

Advocates argue that existing regulations sufficiently address consumer protection, privacy and antidiscrimination concerns. Under the Gramm Leach-Bliley Act, financial institutions may not distribute "raw" consumer data to third parties; instead, prior to distributing consumers' personal financial data, financial institutions must aggregate, anonymize and de-identify personalized transaction details. Financial institutions must also send consumers initial and annual privacy notices and allow them to opt-out of sharing their personal transaction information with unaffiliated third parties.

These protections are, however, weak and evidence suggests that they do not effectively protect consumers' confidential personal financial information. Using statistical methods, data scientists can decode or de-anonymize aggregated consumer social media and financial transaction data. In other words, data scientists can reverse the steps taken by financial institutions to de-identify consumer data and match consumer data with individual consumers' profiles. A recent study by Stanford and Princeton researchers details a theoretical methodology for de-identified web browsing histories and linking individual search histories to social media profiles using only publicly available data to facilitate the matching process.[11]

Behavioral scoring presents even more pernicious concerns. According to proponents of behavioral scoring, the likelihood that a consumer will default on payment obligations may be determined by evaluating the consumer's network of friends, neighbors, folks with similar interests, income levels, and backgrounds. Unlike consumer financial transaction data and payment history evaluations, however, behavioral scoring may not be demonstrably predictive of financial responsibility.

Credit is, indisputably, a critical resource. Individuals and families increasingly rely on credit to finance household purchases or overcome significant, unanticipated expenses.[12] Without access to credit on fair and reasonable terms, it can be extraordinarily expensive to be poor. For families with fragile financial circumstances, credit may serve as a lifeline,

[11] Jessica Su, Ansh Shukla, Sharad Goel, Arvind Narayanan, *De-anonymizing Web Browsing Data with Social Networks*, Apr. 3, 2017, https://5harad.com/papers/twivacy.pdf.

[12] REPORT ON THE ECONOMIC WELL-BEING OF U.S. HOUSEHOLDS IN 2018, BRD. GOVERNORS FED. RESERVE SYS. (2019), https://www.federalreserve.gov/publications/files/2018-report-economic-well-being-us-households-201905.pdf.

enabling consumers to meet short term debt obligations, and to pay for education, housing, and even food.[13]

Consumers navigate an ever-widening web of debt. According to the Federal Reserve Bank of New York's Center for Microeconomics – at the close of the first quarter 2019, families and individuals face over $13 trillion in debt obligations.[14] Rising college and university tuition rates have fueled an increase in educational debt obligations. Students and their families currently owe approximately $1.5 trillion in student loan debt.[15] A parallel narrative in the home mortgage loan market has led American households to borrow over $9 trillion in mortgage debt.[16]

Credit reporting agencies have a special role in financial markets and fintech firms operating at the intersection of startup innovation and consumer credit origination raise a number of the normative questions.[17] As AI increasingly influences the terms and availability of credit, this nascent technology will also inevitably perform a gatekeeping function, determining who receives access to credit, and for those with access, learning algorithms will likely decide the most fundamental terms of any credit arrangement.

Privacy Concerns - Adopted and Proposed State Laws and Regulation

In the absence of definitive federal regulation addressing the use of alternative data, several state laws require disclosure regarding the use of alternative data by credit scoring platforms or limit the use of alternative data.

California Consumer Privacy Act. Signed by Governor Jerry Brown on June 28, 2018, the California Consumer Privacy Act ("CCPA") grants a consumer the right to request that a business "disclose the categories and specific pieces of personal information that it collects about …consumer[s], the categories of sources from which that information is collected, the business purposes for collecting or selling the information, and the categories of [third] parties with which the information is shared."[18] The CCPA also enables consumers to request the deletion of personal information, opt out of the sale of personal information, and access the personal information in a "readily useable format."[19]

---

[13] S*ee* Abbye Atkinson, *Rethinking Credit as Social Provision*, 71 STAN. L. REV. 1093 (2019) (describing the dangers of making credit a key determinant of whether and how basic needs are met).

[14] QUARTERLY REPORT ON HOUSEHOLD DEBT AND CREDIT (Q1 2019), CTR. MICROECON. DATA: FED. RESERVE BANK N.Y. (2019), https://www.newyorkfed.org/medialibrary/interactives/householdcredit/data/pdf/HHDC_2019Q1.pdf.

[15] *Id*.

[16] *Id*.

[17] E. Gerald Corrigan, *Are Banks Special?:* ANNUAL REPORT 1982, FEDERAL RESERVE BANK OF MINNEAPOLIS (raising fundamental questions regarding the role of banks and prudential regulation).

[18] Assembly Bill No. 375, https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375.

[19] Dipayan Gosh, *What You Need to Know About California's New Data Privacy Law*, Harvard Business Review (July 11, 2018), https://hbr.org/2018/07/what-you-need-to-know-about-californias-new-data-privacy-law.

The CCPA construes "personal information" broadly. Under the CCPA, "personal information" means "information that identifies, relates to, describes, is capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household."[20] Similarly, the law also offers a broad definition of the term "sell;" consequently, any of the following activities constitutes a sale of consumer data: "disclosing, disseminating, making available, transferring or otherwise communicating orally or in writing or by electronic or other means" a consumer's personal data. Examples of personal information include consumer's personal identifiers, education information, geolocation, biometric data, internet browsing history, psychometric data, and "inferences" drawn from information used to create a profile about a consumer, reflecting the consumer's preferences, predispositions or behavior, among other attributes.

The CCPA requires companies to obtain consent from customers before selling their personal data to third parties, but it does not apply to consumer information that is de-identified. "De-identified" information is personal information that cannot reasonably identify, relate to, describe, or be linked to a particular consumer.[21] In addition, the CCPA does not apply to "aggregate consumer information," which is information that relates to a group or category of consumers, from which individual consumer identities have been removed, that is not linked or reasonably linkable to any consumer or household or device.[22]

Critics have challenged the breadth of the CCPA and the likely impact that the law would have on established business models in the technology sector including several of the largest technology companies such as Facebook, Twitter, and Google.[23] This restriction may extend to internet service providers such as AT&T and Verizon, which collect broadband activity data (web browsing data) and may generate behavioral profiles to enable digital advertising.[24]

New York State Senate 2302 and Department of Financial Services Regulatory Guidance

As of July 2019, the New York State Assembly is considering the adoption of Senate bill 2302 - a bill that would prohibit consumer reporting agencies from using information about the members of a consumer's social network to evaluate the consumer's creditworthiness.[25] The bill defines the term "members of a consumer's social network" as "a group of individuals authorized by a consumer to be part of his or her social media communications and network."[26] The bill prohibits consumer reporting agencies from

[20] Assembly Bill No. 375, https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375.
[21] *Id.*
[22] *Id.*
[23] *Id.*
[24] *Id.*
[25] https://legislation.nysenate.gov/pdf/bills/2019/S2302.
[26] *Id.*

"collect[ing], evaluat[ing], report[ing], or maintain[ing] in the file on a consumer the credit worthiness, credit standing or credit capacity of members of the consumer's social network for purposes of determining the credit worthiness of the consumer; the average credit worthiness, credit standing or credit capacity of the consumer's social network; or any group score that is not the consumer's own credit worthiness, credit standing or credit capacity."[27] In addition to pending legislation limiting the use of social networking behavioral information in consumer credit evaluation processes, New York state financial services regulators have expressly limited the use of alternative data in the context of life insurance underwriting methodologies.

New York State Department of Financial Services Insurance Circular: Use of External Consumer Data and Information Sources in Underwriting for Life Insurance

On January 18, 2019, the New York State Department of Financial Services ("NYSDFS") issued an insurance circular with two guiding principles on the use of alternative data in life insurance underwriting. First, insurers must independently determine that external data sources do not collect or use prohibited criteria. Insurers may not rely on a vendor's claim of that alternative data does not reflect bias or result in discrimination against protected classes. Insurers may not evade their obligations to comply with antidiscrimination laws by pointing to the proprietary nature of a third-party process.[28] Notwithstanding the fact that alternative data may be provided by third-party vendors, the NYSDFS emphasized that "the burden" to ensure compliance with antidiscrimination laws "remains with the insurer at all times."[29]

Second, insurers should not use external data unless they can establish that it is not "unfairly discriminatory."[30] Insurers must be confident that the use of alternative data is demonstrably predictive of mortality risk. The Circular also notes that "transparency is an important consideration in the use of external data sources to underwrite life insurance." Insurers using external data should be confident that the use of the data is demonstrably predictive of mortality risk and that they can explain how and why this is the case.[31]

Fair Credit Reporting – Alternative Data as a "Consumer Report"

The Fair Credit Reporting Act ("FCRA") imposes obligations on CRAs - entities that provide consumer reports - as well as anyone who uses or furnishes information included in consumer reports. The FCRA defines consumer reports as "communication[s] of any information by a consumer reporting agency bearing on a consumer's creditworthiness, credit standing, credit capacity, character, general reputation, personal characteristics, or mode of living which is used or expected to be used or collected in whole or in part for

[27] *Id.*
[28] *Id.*
[29] https://www.dfs.ny.gov/industry_guidance/circular_letters/cl2019_01
[30] *Id.*
[31] *Id.*

determining a consumer's eligibility for credit, employment purposes, or any other purposes enumerated in the statute."

A number of questions arise as fintech firms begin to gather alternative data and generate credit assessments. If fintech firms' consumer credit assessments based on alternative data constitute "consumer reports," consumers and consumer advocates may assert that fintech firms are subject to the obligations imposed on CRAs under the FCRA. In addition, CRAs may only distribute consumer reports for limited purposes identified in the statute. Consumer reports may be furnished (i) in connection with a credit transaction involving the consumer, (ii) for employment purposes, (iii) in connection with insurance underwriting, or (iv) in accordance with the consumer's written instructions. Consequently, entities gathering data and fintech firms and other firms that obtain and resell data may violate the FCRA by impermissibly using and transferring assessments based on alternative data if such assessments constitute consumer reports. As described in the CFPB request for information and the GAO reports, federal regulators should clarify the contexts in which nontraditional data or alternative data will be deemed "consumer reports" and the instances in which fintech firms may be deemed CRAs.

Adverse Action Notices – Explainability

The FCRA and Equal Credit Opportunity Act ("ECOA") also impose an adverse action notice requirement for entities that take action with respect to any consumer that is based, in whole or in part, on any information contained in a consumer report. State law parallels federal obligations for adverse action notices.

Under relevant provisions of New York Insurance Law referenced above, for example, insurers must notify consumers of their right to receive the "specific reason or reasons for a declination, rate differential, or other adverse underwriting decision." According to the NYSDFS Circular issued earlier this year, if an insurer uses alternative data to underwrite insurance, the reason(s) provided to the consumer for any adverse action "must include details about all information" underlying the decision, including the specific source of the information.

Satisfying adverse action notice requirements may present a significant challenge for platforms using learning algorithms to review large volumes of alternative data. The inscrutable and non-intuitive nature of learning algorithms suggests that even developers may be unable to explain the specific rationale underlying an algorithm's credit or insurance underwriting decision.[32] As a result, it may be difficult for CRAs to explain adverse actions as contemplated under the existing regulatory framework.

[32] Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085 (2019)

Bias, Fairness and Inclusion

Under ECOA and federal fair lending regulations, intentional discrimination based on a protected trait is prohibited under antidiscrimination statutes. Facially-neutral algorithms mitigate the risk that consumers will face intentional discriminatory treatment based on legally protected traits such as race, gender or religion; this suggests that fintech firms employing automated decision-making platforms are not likely to engage in intentional discrimination and therefore are less likely to violate antidiscrimination statutes. The operational mechanics of learning algorithms may, however, mask an algorithm's reliance on a trait that functions as a proxy for a legally protected trait.

Evidence demonstrates that incomplete or inaccurate data sets may influence the objectivity of learning algorithms. Perhaps even more alarming, learning algorithms are designed to identify the most expedient path or optimal variable for solving a problem or making a decision. Learning algorithms seek to identify variables that simplify and expedite the sorting, classifying and ranking of identified subjects. To that end, learning algorithms may rely on proxies or traits that are highly-correlated with protected traits.[33]

This approach may result in the learning algorithm relying on facially-neutral variables in a manner that masks prohibited decision-making behavior.[34] In other words, the algorithm may make decisions using facially neutral variables that function as proxies in the decision-making process for prohibited criteria, violating antidiscrimination protections.

Even if developers expressly program algorithm's not to discriminate on the basis of a protected trait, the developers' biases may creep in and influence the algorithm's operation. Three examples illustrate concerns regarding biases in the data sets.

First, inaccurate, incomplete and otherwise flawed data sets may potentially amplify discrimination.[35] To illustrate this concerns, consider Amazon's attempt to use an automated decision-making platform to evaluate, score and rank job applicants for a software developer position.

Amazon created a resume review platform designed to identify and sort candidates with desirable attributes for a software developer position. The platform received facially-neutral instructions regarding educational or skill prerequisites and analyzed the resumes of employees recently hired for similar computer programmer positions. Beyond this

[33] Daniel Schwarcz and Anya Prince, *Proxy Discrimination In The Age Of Artificial Intelligence And Big Data*, IOWA L. REV. (*Forthcoming* 2020).

[34] Solon Barocas & Andrew Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV. 671, 678 (2016).

[35] SAFIYA NOBLE, ALGORITHMS OF OPPRESSION (2018); VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE AND PUNISH THE POOR (2018); Margaret Hu, *Algorithmic Jim Crow*, 86 FORDHAM L. REV. 633 (2017).

initial data set and series of instructions, the platform taught itself to mimic human-like decision-making behavior. As the platform began to review real candidates' resumes, it operated independently, using cognitive analysis to decide which candidates to interview without specific instructions regarding the submitted resumes.[36]

Amazon's goal was to identify best athletes in a competitive pools of applicants.[37] Notwithstanding programmers' intentions, the platform began to "penalize resumes that included the word 'women's,' as in 'women's chess club captain,'" and "downgraded graduates of. . . women's colleges."[38] Amazon's experiment illustrates the risk that an automated platform will inherit the biases that data sets and developers unknowingly introduce, leading to unanticipated and potentially prohibited discrimination against individuals who are members of a legally protected class.

Second, selecting and cleaning data sets involves human judgment. Data sets are often compiled by third party vendors and distributed to developers who utilize the data to create a training data set. A learning algorithm's successful analysis depends significantly on the data used to train the algorithm.

In order to achieve the predictive benefits of learning algorithms, data sets require a large number of observations. Even if a data set has a sufficient number of observations, the data must be subjected to several pre-processing steps including, among others, cleaning, partitioning, sampling, scaling and feature selection. These steps are necessary because datasets are rarely free from missing or inaccurate values. Data scientists must decide how to resolve missing values. The options for addressing these concerns may include removing the subjects with missing values from the data set and excluding them from the analysis. At each step from data collection decisions to the development of the algorithm, human judgment will influence how the algorithm operates. Finally, some commentators have demonstrated that underrepresentation, particularly of members of legally protected classes, may lead to digital discrimination.[39]

One study suggests that fintech firms using AI based methodologies are replicating historic biases. According to the results of the study both fintech and traditional mortgage

[36] See Jeffrey Dastin, Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women, REUTERS (Oct. 9, 2018), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

[37] *Id*.

[38] *Id*.

[39] Joy Buolamwini and Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROCEEDINGS OF MACHINE LEARNING RESEARCH, CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 1-15 (2018) (present an approach to evaluate bias present in automated facial analysis algorithms and datasets).

origination firms lending practices result in discrimination against Latinx and African-American borrowers.[40]

Cyber Security Concerns

In addition to privacy and discrimination concerns, permitting fintech firms and CRAs to collect alternative data heightens cybersecurity concerns. The rising cost, frequency, and severity of data breaches now dominate risk management discussions. Over the last ten years, more than 4,000 known data breaches have shocked, debilitated, and even (temporarily) paralyzed markets. Commentators estimate that vast numbers of records containing confidential or sensitive data have been compromised. Experts suggest that data breaches cost the global economy more than $ 400 billion dollars of losses annually.

As cyberattacks multiply, governments, corporations, and citizens scramble to mount a successful defense against cyber-intrusions. The size, sophistication, and diversity of styles of the cyberattacks renders these activities among the most perilous of emerging risk management concerns.

The cyberattacks against financial institutions threaten the stability of financial markets and create personal costs for consumers exposed during data breaches. As the New York State Department of Financial Services noted, "[c]yber hacking is a potentially existential threat to our financial markets." Federal regulators have warned that cybersecurity threats may "wreak serious havoc on the financial lives of consumers."

Financial transaction and social media data present particularly attractive targets for hackers. Pursuant to federal regulation and consistent with their business models, large financial institutions acquire, collect, and retain significant volumes of personal information. Collection, storage and transfer of this sensitive data renders financial institutions and retailers highly attractive targets for hackers.

Cyberattacks capture national and international attention because of their pervasive effects. For example, in December 2013, Target, a national retailer, announced that it was the target of a massive data breach. The hackers who orchestrated the data breach obtained the confidential credit and debit card information of more than 40 million customers. As investigations ensued, Target continued to adjust its estimate of the number of records accessed, ultimately reporting that hackers captured the personal data of as many as 110 million customers. In 2014, hackers invaded home improvement retailer Home Depot's records and acquired 56 million customers' credit and debit account information and 53 million customers e-mail addresses.

[40] See, e.g., Robert Bartlett et al., Consumer-Lending Discrimination in the Era of FinTech (October 2018) (unpublished manuscript), https://faculty.haas.berkeley.edu/morse/research/papers/discrim.pdf (an empirical study comparing discrimination in lending by traditional mortgage origination firms with face-to-face interaction with borrowers and decisions made by fintech platforms; the study finds that "lenders charge Latinx/African-American borrowers 7.9 and 3.6 basis points more for purchase and refinance mortgages respectively, costing them $765M in aggregate per year in extra interest").

Equifax's settlement this week illustrates the perils of cyberattacks against credit reporting agencies. Between mid-May 2017 and July 2017, Equifax, one of the country's largest CRAs suffered one of the largest known financial data breaches, exposing the personal information (names, addresses, dates of birth, Social Security numbers, and driver's license numbers) of more than 148 million Americans, 8,000 Canadians, and nearly 700,000 UK citizens.

Former Equifax CEO Richard Smith in testimony before Congress explained that the data breach resulted from hackers' exploitation of a flaw in "Apache Struts," an open source web application. While a patch was released during the first week of March 2017, Equifax failed to apply the security updates until two months later. Equifax should have addressed this vulnerability within forty-eight hours, but it did not.[41] Equifax's information security scans also failed not detect the Apache Struts vulnerability.[42]

On May 13, 2017, hackers exploited this vulnerability to access Equifax's systems and consumers' personally identifiable information.[43] Between May 13, 2017 and July 30, 2017, evidence suggests that the attackers continued to access sensitive information, exploiting the same Apache Struts vulnerability without being detected by Equifax's security tools.[44]

Mr. Smith notified the Equifax board about the breach on August 22, 2017.[45] On September 7, 2017, Equifax disclosed the breach to the American public.[46] In other words, Equifax waited six weeks from the time they discovered the breach until they disclosed said breach to the American public. The Equifax settlement marks one of the largest data breach settlements and will provide up to $425 million in consumer restitution. The settlement reflects a number of measures that Equifax will take to protect consumers' personal data and assist with fraud detection.

The Equifax data breach demonstrates the systemically important role of CRA in credit markets and US financial markets. As the universe of fintech firms expands, regulatory oversight of these entities must reflect the nature of the information that the firms will collect, store and transfer. Regulation must also reflect the significant role of the these firms in the stability of consumer credit markets and broader financial markets.

As the Office of the Comptroller of the Currency ("OCC") and Federal Deposit Incorporation's ("FDIC") consider paths for granting fintech firms special purpose

---

[41] *Oversight of the Equifax Data Breach: 7 Answers for Consumers, Hearing Before the U.S. H. Comm. on Energy and Commerce Subcomm. on Digital Commerce and Consumer Protection*, 115th Cong. 3 (2017) (statement of Richard F. Smith, CEO, Equifax), https://democrats-energycommerce.house.gov/sites/democrats.energycommerce.house.gov/files/documents/Testimony-Smith-DCCP-Hrg-on-Oversight-of-the-Equifax-Data-Breach-Answers-for-Consumers-2017-10-03.pdf.

[42] *Id.*

[43] Stacy Cowley, *2.5 Million More People Potentially Exposed in Equifax Breach*, N.Y. TIMES (Oct. 2, 2017), https://www.nytimes.com/2017/10/02/business/equifax-breach.html.

[44] *Oversight of the Equifax Data Breach: 7 Answers for Consumers, Hearing Before the U.S. H. Comm. on Energy and Commerce Subcomm. on Digital Commerce and Consumer Protection*, 115th Cong. 3 (2017) (statement of Richard F. Smith, CEO, Equifax), https://democrats-energycommerce.house.gov/sites/democrats.energycommerce.house.gov/files/documents/Testimony-Smith-DCCP-Hrg-on-Oversight-of-the-Equifax-Data-Breach-Answers-for-Consumers-2017-10-03.pdf.

[45] *Id.*

[46] *Id.*

nonbank charters and Industrial Loan Corporation ("ILC") charters concerns mount regarding careful monitoring of fintech firms' privacy and cybersecurity measures and their ability to protect the collection, storage and transfer of alternative data.

Blockchain-Based Credit Scoring and Lending Models

For several years, fintech firms and conventional CRAs have integrated learning algorithms into credit scoring models. In more recent years, developers began to advocate for credit scoring models built on decentralized, distributed digital ledger protocols.

On January 27, 2018, Jesse Leimgruber, Alain Meier, John Backus published a whitepaper for Bloom Protocol, a "credit staking" decentralized credit scoring platform powered by Ethereum and the Interplanetary File System. According to the whitepaper, Bloom plans to offer three main services: Bloom ID (Identity Attestation), BloomIQ (Credit Registry) and BloomScore (Credit Scoring). According to Bloom, its model addresses the shortcomings of traditional credit scoring by transitioning the credit scoring process to the blockchain protocol. Touting its success as one of the first distributed ledger credit scoring and lending platforms in the world, Bloom promises to facilitate cross border credit scoring, accommodate users with no credit history, secure personal information, increase global access to credit development and provide greater competition in the credit risk evaluation market.

Bloom introduces three unique models: the BloomID, BloomIQ and Bloom Score. Using a peer assessment methodology, Bloom claims that consumers with thin, limited, impaired or no credit history may demonstrate creditworthiness and enjoy access greater access to credit. While the whitepaper clearly indicates that the model will evaluate conventional criteria such as loan and bill repayment history, Bloom relies heavily on social networking to assess a consumer's eligibility to receive credit. A number of important details regarding Bloom's methodology are not revealed in the whitepaper, but there is significant potential for a decentralized distributed ledger based credit scoring platform to assist invisible and unscorable consumers by offering greater transparency in the credit evaluation process, a more easily reviewable and correctible credit report and reduced incidents of fraud and data breaches.

For decades, consumer advocates, academics, regulators and state and federal legislators have recognized that low-income consumers pay remarkably more for basic financial services such as check cashing, money transfers and short-term loans. Nearly ten percent of American households continue to lack access to traditional savings and checking accounts.

Consumers with limited access to basic banking services, those living in financial services deserts (requiring them to commute significant distances to bank branches) have had too few options for obtaining access to credit on fair and reasonable terms. Check cashing storefronts, payday loan outlets and other predatory financial services providers exploited invisible or unscorable consumers' lack of access to conventional banking and credit services.[47]

Fintech firms operating at the intersection of startup innovation and consumer credit evaluations raise a number of the normative questions.[48] As artificial intelligence increasingly influences the terms and availability of credit, this nascent technology and the firms adopting it will come to perform an important gatekeeping function, determining whose receives access to credit, and for those with access, learning algorithms will likely decide the most fundamental terms of any credit arrangement.

To be sure, the advent of artificial intelligence technology disrupts legacy banking, inspires a new market infrastructure and spurs development that *may* benefit unbanked and underbanked consumers. The successful expansion of access to credit may depend largely on regulators' effective supervision of the integration of alternative data and reliance on opaque, inscrutable and non-intuitive algorithms.

[47] Twenty percent of families with traditional bank accounts still rely on alternative financial services outlets. Jason Furman, *Financial inclusion in the United States*, June 10, 2013 available at https://obamawhitehouse.archives.gov/blog/2016/06/10/financial-inclusion-united-states.

[48] E. Gerald Corrigan, *Are Banks Special?:* ANNUAL REPORT 1982, FEDERAL RESERVE BANK OF MINNEAPOLIS (raising fundamental questions regarding the role of banks and prudential regulation).