April 2021

Fairness in the Eyes of the Beholder: AI; Fairness; and Alternative Credit Scoring

Janine S. Hiller
Pamplin College of Business, Virginia Tech

Follow this and additional works at: https://researchrepository.wvu.edu/wvlr

Recommended Citation
Available at: https://researchrepository.wvu.edu/wvlr/vol123/iss3/8

This Article is brought to you for free and open access by the WVU College of Law at The Research Repository @ WVU. It has been accepted for inclusion in West Virginia Law Review by an authorized editor of The Research Repository @ WVU. For more information, please contact ian.harmon@mail.wvu.edu.
FAIRNESS IN THE EYES OF THE BEHOLDER: AI; FAIRNESS; AND ALTERNATIVE CREDIT SCORING

Janine S. Hiller*

Abstract

Artificial intelligence is based, in part, on learning algorithms that can continually monitor and embed new data from large numbers of sources to create ever “improved” decisions, whether the decisions are applied to the physical world like the operation of smart cars, or whether the decision is about the extension of credit to a loan applicant. Yet, it is well known that algorithms and resulting decision making AI models are plagued by unintended bias and discriminatory results. Data science scholars have attempted to address bias and discrimination through the imposition of multiple mathematically represented options to measure fairness. However, these mathematical measures can conflict, and they can be incompatible with legal concepts of fairness, especially those found in non-discrimination laws. This article provides an introduction to the multiple meanings of fairness in both the data science and legal disciplines and uses a case study of AI and alternative data credit scoring to illustrate how the use of different disciplinary meanings of fairness will significantly affect societal outcomes. In conclusion, it is proposed that a socio-technical approach to AI fairness, which incorporates legal concepts, will increase the acceptance, legitimacy, and trust of those systems.

“[U]ltimately, science can only take us so far, and human judgments and norms will always play the essential role of choosing . . . notions of fairness we want to enforce.”
— The Ethical Algorithm

I. INTRODUCTION ........................................................................................................908
II. FAIRNESS IN THE EYES OF THE BEHOLDER ..................................................911

* Professor of Business Law, Sorensen Professor in Finance, Pamplin College of Business, Virginia Tech. The author extends thanks to Lindsay Sain Jones for helpful and insightful comments that contributed to this article.

A. Data Science Fairness .................................................................................. 912
   1. Fairness as a Mathematical Construct ................................................. 912
   2. Causality Fairness .............................................................................. 914
   3. Quasi-mathematical Fairness: Socio-technical Design
      Choice ......................................................................................... 915
B. Legal Fairness .......................................................................................... 916
   1. Fundamental Fairness ....................................................................... 916
   2. Administrative Agency Fairness ....................................................... 917
   3. Contractual Fairness ......................................................................... 918
III. Credit Scoring and Fairness Case Study ................................................. 919
   A. Fair Lending Laws .......................................................................... 921
   B. Alternative Credit Data and AI ........................................................ 923
   C. Multiple Lenses of Fair Artificial Intelligence for
      Credit Scoring .................................................................................. 925
      1. Credit Ratings and Mortgages: FHFA ........................................... 925
      2. Consumer Lending: CFPB and Upstart ........................................ 926
IV. An Approach to AI Fairness .................................................................... 932
V. Conclusion ............................................................................................... 935

I. INTRODUCTION

Artificial intelligence ("AI") is based, in part, on learning algorithms that can continually monitor and embed new data from large numbers of sources to create ever "improved" decisions, whether the decisions are applied to the physical world like the operation of smart cars, or whether the decision is about the extension of credit to a loan applicant. If data is the bloodstream of artificial intelligence, algorithms and models are its heart. It is well known that algorithms, and thus intelligent systems, are prone to bias and discrimination, at various stages. Scholars and practitioners in the disciplines of data and computer science also recognize the need to moderate and improve fairness in artificial intelligence systems. A fundamental problem is that the very definition of fairness is subject to widely different applications, and the scientific definition can be incompatible with legal concepts of fairness, such as the principles reflected in non-discrimination laws. Those who create intelligent systems, and those who regulate them on the basis of legal principles, must understand the language of fairness from differing viewpoints. Otherwise, intelligent systems will perpetuate inequality that is unacceptable to society and harmful to individuals, and laws may be less effective and have unintended consequences.

New intelligent algorithms, fed a continual diet of detailed and alternative data about lifestyle, buying habits, and personal information about individuals, in order to create a credit score, exemplify the complicated dynamic. Individual credit scores directly impact consumers in ways that are incredibly powerful. This individually attributed value impacts where a person can live,
who will employ them, and even what treatment their healthcare provider will recommend. Calculations of credit scores are opaque, and no one truly knows the exact methods and information utilized to arrive at a final score. In the wake of the COVID-19 pandemic, for example, it is unclear how deferred mortgage payments will ultimately affect credit scores, despite some credit agencies’ assurances. Why? Because new “artificial” credit calculation systems use broad new types of data to feed into moderated algorithms in order to provide “intelligent” scores that are always learning from more data. In 2017, the Consumer Financial Protection Bureau (“CFPB”) issued a Request for Information Regarding Use of Alternative Data and Modeling Techniques in the Credit Process. With no action from the CFPB, in March 2019, CEOs from the three main credit data aggregators were called before Congress to discuss how to open up the consumer credit industry to benefit individuals who have been locked out by the present credit scoring framework.

Separately, the Federal Trade Commission (“FTC”) worked to understand how to avoid bias and harms in AI. Clearly, artificial intelligence, fairness, and consumer protection are at the forefront of legal and policy concerns.

After an introduction to basic terms, Part II of this article provides a discussion of the multiple meanings of fairness from both the data science and legal disciplines. Part III presents a case study and comparison of AI used in credit scoring and fair lending to demonstrate how law must both respond to and exert influence on data discipline-based concepts of AI fairness. Part IV proposes an approach to bridging the data and legal disciplines that requires legal leadership and science integration in order to achieve the common goal of AI fairness. Algorithmic and artificial intelligence is becoming embedded in fine grained ways in everyday life, and if fairness is in the eyes of the beholder, it is essential for AI designers and regulators to share their visions and requirements for its fair application.

Before undertaking a discussion of fairness, a few definitional notes about terms are important. Some of these terms are disputed in their own disciplines, and this article does not presume to offer answers to these definitional challenges. However, the definitions are important because of the

---

5 See Mirka Snyder Caron, The Transformative Effect of AI on the Banking Industry, 34 BANKING & FIN. L. REV. 169, 176–77 (2019). Artificial intelligence is a term which itself is subject
comparative different meanings and perspectives and for identifying points where collaboration and interventions are most effective within the development and implementation of artificial intelligence. Several broad concepts are important to recognize and explain as the discussion begins.

The first essential definition immediately highlights the challenges, as AI disciplinary researchers state that “there is no widely accepted definition of Artificial Intelligence.”6 Artificial intelligence as used in this paper is a broad umbrella, incorporating a number of other concepts but ultimately referring to systems “that can sense, reason, and respond to their environment in real time.”7 Regardless of where an AI system falls on a scale of autonomous action, for purposes of this discussion, it means that the system is using data inputs and independently exercising some automated or prompted decision making.

Secondly, it has also been noted that with regards to algorithms, “[t]here is little agreement in the relevant literature on the definition of an algorithm.”8 However, it is most often used to describe a mathematical application, or set of instructions, to address a particular problem or question. Algorithms are described as a “precisely specified series of instructions for performing some concrete task.”9 IBM explains an algorithm, in data science, to mean “a sequence of statistical processing steps.”10

Algorithms require a large amount of data in order to create models that are trained and verified. In different literatures, learning algorithms, machine learning, and artificial intelligence terms are sometimes used interchangeably when describing the fairness of data analytics outcomes or a data decision system. Finally, the term machine learning is a subset of artificial intelligence that can sense new patterns in data and adapt to those changes. IBM states that this type of AI can “learn from data and improve [its] accuracy over time without being programmed to do so.”11

In sum, artificial intelligence is broader than algorithms or machine learning; however, these are components of an autonomous (or semi-

to different definitions, as described in the 100 Year Study on Artificial Intelligence at Stanford University. Peter Stone, Rodney Brooks, Erik Brynjolfsson, Ryan Carlo et al., Artificial Intelligence and Life in 2030 12 (2016), https://ai100.stanford.edu/2016-report/section-i-what-artificial-intelligence/defining-ai.

6 Pei Wang, On Defining Artificial Intelligence, 10 J. ARTIFICIAL GEN. INTEL. 1, 1 (2019).
9 Kearns & Roth, supra note 1, at 12.
11 Id.
autonomous) reasoning system that one would call artificial intelligence. The terms are important for the analysis because artificial intelligence is not a monolith, and fairness considerations occur and are relevant at different levels. While a comparison of the fundamental meanings of “fairness” across the data science and legal disciplines is a large undertaking, the following comparison begins the journey.

II. FAIRNESS IN THE EYES OF THE BEHOLDER

The disconnect between legal and policy understandings of fairness and scientific implementations of fairness within artificial intelligence products is exemplified by the challenge by ProPublica to the fairness of Northpointe’s COMPAS algorithm that was used to predict criminal recidivism. ProPublica stated that the COMPAS algorithmic predictions were discriminatory based on race because “[b]lack defendants who do not recidivate were nearly twice as likely to be classified by COMPAS as higher risk compared to their white counterparts (45 percent vs. 23 percent).” The conclusion was based on the mathematical calculation of equalized odds. In other words, one would expect the odds of errors for high risk recidivism to be equal as compared across race. As described more fully below, this is a type of group fairness metric by making demographic comparisons. Northpointe did not argue that ProPublica was wrong per se, but it disputed ProPublica’s accusation of discrimination because the actual recidivism rate as compared across race was similarly accurate (or inaccurate), a measure known as predictive parity. The entire controversy caused a stir and debate across the data community, as well as the legal field. When the dust settled, both sides were found to have made at least partially valid statements, and it seemed clear that the dispute was, in large part, about what each side meant by fairness and the best methods to achieve that goal. The next sections discuss foundational concepts for understanding these different meanings.


13 Id. at 231 (quoting Jeff Larson, Surya Mattu, Lauren Kirchner & Julia Angwin, How We Analyzed the COMPAS Recidivism Algorithm, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm).

14 Id. at 231–32.

15 Id. This is also known as accuracy equity.

16 Id.; see also Alice Xiang, Reconciling Legal and Technical Approaches to Algorithmic Bias, 88 TENN. L. REV. (forthcoming 2021).
A. Data Science Fairness

There is no one encompassing framework for achieving fairness, as understood by computer scientists, statisticians, and data scientists ("data science"). In analytic science circles, there are at least 25 different definitions of fairness in artificial intelligence, and generally fairness methods are focused on alleviating one impact at a time. Further, researchers have argued that it is mathematically impossible to apply multiple concepts of fairness to the same problem at the same time; for example, some measures of both individual fairness and group fairness are simply incompatible. In sum, in data communities, fairness is described as a "nascent" field of study with some disagreement about terms and goals. Within these parameters and limitations, there are several important data science approaches to fairness described below.

1. Fairness as a Mathematical Construct

It has been said that "fairness refers to a concrete mathematical embodiment of some rule provided by an external party such as a government and which must be imposed on a learning algorithm." This definition contains two significant points: the mathematical expression and the externally imposed rule-based nature of fairness. There is an important connection here, as "machine

---

17 These disciplines are not the same, nor are they exclusively the domains of artificial intelligence, but for the general purposes of this article, the term data science is used to denote those academic and practitioner areas involved with a scientific approach to algorithms and artificial intelligence.


21 As one can easily infer from the discussion, there is much work being done in the realm of fairness for artificial intelligence and machine learning, and there are different ways to classify the categories of fairness. For a good overview of the different research avenues and another way to categorize these, see Irini Ntoutsi, Pavlos Fafalios, Ujwal Gadiraju, Vasileios Iosifidis et al., Bias in Data-driven Artificial Intelligence Systems—An Introductory Survey, 10 WILEY INTERDISC. REV.: DATA MINING & KNOWLEDGE DISCOVERY 1 (Feb. 3, 2020), https://doi.org/10.1002/widm.1356.

learning won’t give you things like gender neutrality ‘for free’ that you didn’t explicitly ask for.”\textsuperscript{23} Therefore, the external requirements, or rules, are essential to the process of machine learning meeting legal standards and principles. Expecting artificial intelligence to discover fairness on its own, so to speak, does not happen. In order to build in fairness to a predictive model, a number of mathematical approaches are proposed in data science. There are many measures, but they primarily fall into two categories: group fairness and individual fairness.

Group fairness is sometimes called statistical fairness and is a measure that seeks equality across groups based on a particular mathematical attribute, such as statistically similar demographics, false positive/negative error rates, or discovery/omission rates.\textsuperscript{24} A reflection of equality in these metrics is thought of as a way to avoid discrimination. However, it should be noted that specific measures of group fairness will conflict with each other, as they did in the COMPAS debate, and, further, that the goal of statistical fairness is to “give guarantees to ‘average’ members of the protected groups,”\textsuperscript{25} rather than fairness outcomes to each individual.

In contrast, individual fairness is based on the goal, measures to treat individuals the same as opposed to the average member of a group, and is tested by various outcomes, such as “similar individuals should be treated similarly”\textsuperscript{26} and “less qualified individuals should not be favored over more qualified individuals.”\textsuperscript{27} Chouldechova and Roth note that individual fairness is a contested concept because it presupposes that there is an external way that individuals are compared and found to be similar, a difficult presumption to meet, although there are proposals to achieve such a state.\textsuperscript{28} One side of the argument is that individual fairness “costs to accuracy are likely to be unpalatable; we’re simply asking for too much.”\textsuperscript{29} On the other side of the argument is that “[f]inding reasonable ways to give meaningful alternative

\textsuperscript{23} Kearns & Roth, supra note 1, at 61.


\textsuperscript{25} Chouldechova & Roth, supra note 24, at 3.

\textsuperscript{26} Id.

\textsuperscript{27} Id.

\textsuperscript{28} Id. at 4–5 (“Ultimately, however, these approaches all assume that fairness is perfectly defined with respect to some metric, and that there is some sort of direct access to it.”).

\textsuperscript{29} Kerns & Roth, supra note 1, at 90.
fairness guarantees to individuals is one of the most exciting areas of ongoing research.\textsuperscript{30} Dwork compares group fairness, under the term statistical parity, with individual fairness:

Statistical parity is the property that the demographics of those receiving positive (or negative) classifications are identical to the demographics of the population as a whole. Statistical parity speaks to group fairness rather than individual fairness, and appears desirable, as it equalizes outcomes across protected and non-protected groups. However, we demonstrate its inadequacy as a notion of fairness through several examples in which statistical parity is maintained, but from the point of view of an individual, the outcome is blatantly unfair.\textsuperscript{31}

Whatever the approach, and while there is some disagreement within the scientific community, it is presently largely believed that mathematical fairness and predictive accuracy are opposites; if fairness is increased, then predictive accuracy will be decreased.\textsuperscript{32} Research to the contrary, however, suggests that mathematically increasing fairness will, for the most part, also increase accuracy.\textsuperscript{33}

2. Causality Fairness

Most recently, a third approach, causality fairness, has been the subject of work to create mathematical fairness metrics.\textsuperscript{34} In sum, the causal fairness approach asks a different question: instead of asking if the algorithm is producing a fair, non-discriminatory result, the causality fairness approach asks if using the protected classification, or its proxy, causes a discriminatory decision.\textsuperscript{35} If it does, then it would not be fair. There are several different ways to show causality or, for purposes of fairness, to show that a protected classification did not cause the negative outcome. Specific causality approaches include counterfactuals, no

\textsuperscript{30} Id.

\textsuperscript{31} Dwork et al., supra note 24, at 2 (emphasis omitted).

\textsuperscript{32} Wick et al., supra note 22, at 1.

\textsuperscript{33} Id.

\textsuperscript{34} Renzhe Xu, Peng Cui, Kun Kuang, Bo Li et al., \textit{Algorithmic Decision Making with Conditional Fairness}, in \textsc{Proceedings of the 26th SIGKDD International Conference on Knowledge Discovery \& Data Mining} 2125 (2020).

\textsuperscript{35} See Aria Khademi, Sanghack Lee, David Foley \& Vasant Honavar, \textit{Fairness in Algorithmic Decision Making: An Excursion Through the Lens of Causality}, in \textsc{Proceedings of the 2019 World Wide Web Conference} 2907 (2019); see also Xiang, supra note 16.
unresolved discrimination, no proxy discrimination, and fair inference.\textsuperscript{36} The causality fairness approach is fundamentally different than a group or individual fairness metric, and while causality is a common requirement for the law, it is not a traditionally mainstream approach for mathematical fairness.\textsuperscript{37}

3. Quasi-mathematical Fairness: Socio-technical Design Choice

Despite the intense work to mathematically represent fairness for algorithmic utilization, critics claim that methods are “poor measures for detecting discriminatory algorithms and, even more importantly, designing algorithms to satisfy these definitions can, perversely, negatively impact the well-being of minority and majority communities alike.”\textsuperscript{38} Instead, viewing fairness as a socio-technical design choice includes mathematical representations of fairness as only one aspect and social systems as at least a co-equal component. A socio-technical approach recognizes that “reality is messy, but strong frameworks can help enable process and order, even if they cannot provide definitive solutions.”\textsuperscript{39}

The socio-technical method is included in the discussion of mathematical fairness because it includes a technical, mathematical approach to fairness, although a technical approach alone is insufficient. The socio-technical approach is implicated, in part, when scientists reach out to include other inputs to the mathematical model. For example, researchers have proposed systems that would include public feedback on the fairness of a model\textsuperscript{40} or that would create a system where data would be submitted to a regulator to be corrected for bias and then returned for a decision maker to choose from a sliding scale of mathematical fairness.\textsuperscript{41} These proposals still use mathematical solutions as the


\textsuperscript{37} See Xiang, supra note 16.


core attribute of a decision-making algorithm and model. The socio-technical system lens requires, however, that in order to achieve a meaningful degree of fairness, "the social must be considered alongside the technical in any design" and is required in order to seek fairness.

B. Legal Fairness

While fairness metrics in data science for AI focus primarily on avoiding predictive discrimination, fairness is a broader recurring theme in the law in various ways. Notions of fairness are foundational to different categories of law, which are illustrated in three non-exclusive groups for purposes of discussion: constitutional, administrative, and contract law. To be clear, no attempt is made to catalogue every type of law that is relevant for the design of fair AI systems. Similar to the discussion of mathematical fairness metrics, these three categories are types of law that provide input to the meanings of legal fairness.

1. Fundamental Fairness

The Due Process Clauses of the Fifth Amendment and the Fourteenth Amendment prohibit federal and state governments, respectively, from depriving anyone of "life, liberty, or property, without due process of law." At a basic level, courts' application of this principle is based on "fundamental fairness." To meet the standards of due process fairness, the government's action must comply with both procedural due process and substantive due process. Procedural due process asks whether the government has followed proper procedures, while substantive due process asks the question of whether the

---

42 Critics of this narrow focus believe that "by abstracting away the social context in which these systems will be deployed, fair-ML researchers miss the broader context, including information necessary to create fairer outcomes, or even to understand fairness as a concept." See Selbst et al., supra note 39, at 59.

43 See, e.g., Federal Trade Commission Act § 5, 15 U.S.C.A. § 45 (West 2021) (authorizing the FTC to take enforcement action for "unfair" commercial practices); Red Lion Broad. Co. v. FCC, 395 U.S. 367 (1969) (upholding the FCC's fairness doctrine, which required broadcast license holders to present issues of public importance in a manner that was honest, equitable and balanced); Milliken v. Meyer, 311 U.S. 457, 463 (1940) (holding that in order for a court to assert jurisdiction over a defendant it must not offend principles of "fair play and substantial justice").

44 There is no intent to define all the types of law that involve fairness considerations, but to use these three areas to examine and discuss the differences so that the data science concept of fairness may be compared. Nor is this brief discussion meant to be exhaustive.

45 U.S. Const. amend. V; XIV.

46 LandMark Publications, Due Process: Historic Supreme Court Decisions 1 (2011) ("At a basic level, procedural due process is essentially based on the concept of 'fundamental fairness.'").
government’s action is justified by sufficient purpose. Due process rationale for government action is based to a great degree on ensuring that the power of the government over individuals is executed in a fair manner. AI has similar power over people in ever-increasing personal ways, such as credit scoring that is discussed in the case study in Part III.

Scholars have argued that due process should also be applied to data driven decisions more broadly because they affect the fundamental qualities of life, such as privacy. In order to ensure the fairness of algorithms that have the potential to deprive people of their life, liberty or property, they should also be subjected to this due process review. For example, Danielle Keats Citron and Frank Pasquale have argued that, “the underlying values of due process—transparency, accuracy, accountability, participation, and fairness—should animate the oversight of scoring systems given their profound impact on people’s lives.”

2. Administrative Agency Fairness

Administrative law is steeped in a history around the fair implementation of laws passed by legislatures, and “a focus on fairness has been a constant in administrative procedures.” Further, in a very relevant parallel to AI decision making, administrative action, which includes transparent procedures and process, lends legitimacy to rules and governing bodies if perceived as fair by the public. AI systems and decision-making also need to achieve legitimacy in order to be accepted broadly in society.

Administrative law fairness can also be substantive, as illustrated by the example of the Federal Trade Commission’s (“FTC”) broad enforcement power to prevent “unfair” business practices in commerce. According to the FTC, a business practice is unfair if (i) it causes or is likely to cause substantial consumer injury which a consumer could not reasonably avoid and (ii) it is not outweighed

51 Id.
by the benefit to consumers.\textsuperscript{53} Thus, the FTC is required to undertake a balancing analysis before undertaking regulatory action.

The 2015 case, \textit{FTC v. Wyndham Worldwide Corp.},\textsuperscript{54} is illustrative. The FTC charged the Wyndham Hotel ("Wyndham") with unfair practices because it failed, after multiple vulnerabilities and notices, to secure its networks, which contributed to yet another data breach of customer information.\textsuperscript{55} Wyndham argued that to prove unfairness to consumers due to the failure to secure its networks required a showing that it had acted in a way that was unethical, inequitable, unjust, or not impartial.\textsuperscript{56} The history and definition of unfairness was discussed in some detail by the court, which disagreed with Wyndham’s argument,\textsuperscript{57} and the FTC was allowed to continue its case based on consumer injury, whether the consumer was able to avoid that injury, and whether the harm incurred was not outweighed by a benefit to the consumer.\textsuperscript{58} If this same analysis is applied to AI unfair practices that harm consumers, it is highly unlikely that a consumer would have the ability or means to protect themselves to avoid such a black box impact. If an AI unfairness harm exists, then it would need to be balanced against any benefits of the same.

In sum, administrative actions must in general be fair, which requires transparency and reasonable processes. Trust and legitimacy are reasons for these standards. Further, the FTC is an agency that decides substantive issues of unfairness based on consumer injury, regardless of the intent of the defendant, balanced against the benefit to the consumer. This is not unlike AI systems that can provide insights into data at levels that a human could not, therefore benefitting consumers in multiple ways. But AI systems can also harm consumers because they may make decisions based on group characteristics rather than individual ones, while consumers have little or no power over the data used or input into the decisions made. Transparency and processes that support legitimacy are similarly important for AI systems to gain consumer acceptance and trust. These concepts are discussed further in Part IV.

3. Contractual Fairness

The law compels a minimal level of fairness in private agreements, contracts, as well. A court will not enforce a contract if it is deemed to be unconscionable. Although dependent on state laws, an unconscionable term has


\textsuperscript{54} 799 F.3d 236 (3d Cir. 2015).

\textsuperscript{55} \textit{Id.} at 240–42.

\textsuperscript{56} \textit{Id.} at 244–46. Wyndham also argued that the meaning of unfairness was so unclear that it did not have notice under the Due Process Clause, but the court disagreed. \textit{Id.} at 249–59.

\textsuperscript{57} \textit{Id.}

\textsuperscript{58} \textit{Id.} at 248–49.
been defined as a term that "shock[s] the conscience" of the court or is "unreasonably and unexpectedly harsh" under the circumstances. In essence, the terms are so unfair to the weaker party, such as a consumer who lacks any real choice, that a court will not enforce the agreement. For commercial transactions, rather than consumer transactions, unconscionability is judged by commercial reasonableness.60

Similar to the due process analysis described above, a court will review the contract for unconscionability in terms of substance and procedure. First, a court will determine whether the contract terms are so unfair and unreasonable as to constitute substantive unconscionability, and then the court analyzes the relative bargaining power of the contracting parties to determine whether procedural unconscionability exists.61 Thus, if an unfair contract is executed under unfair circumstances, it will not be enforceable.62 These unfair circumstances, when one party has little choice over the terms and there are highly unfavorable terms for the weaker side, resonate with certain applications of automated AI decision making being applied to individuals, such as in the next case study of credit scoring and AI fairness.

Part III applies both the mathematical and legal concepts of fairness to the case study of individual credit scoring and artificial intelligence.

III. CREDIT SCORING AND FAIRNESS CASE STUDY

Credit scores are formulaic numerical expressions used to evaluate an individual's creditworthiness at a given point in time.63 The credit scoring process involves the application of a statistical model to historical data to predict a borrower's ability to pay.64 To develop the model, a scorer selects a random sample of customers and analyzes the sample to identify characteristics or variables that relate to risk.65 Each of the variables is then assigned a weight

62 For more specifics about procedural and substantive unconscionability, see Melissa T. Lonegrass, Finding Room for Fairness in Formalism—The Sliding Scale Approach to Unconscionability, 44 Loy. U. Chi. L.J. 1 (2012).
65 Id.
based on how strong it is as a predictor of creditworthiness.\textsuperscript{66} Once an individual’s credit data is run through the algorithmic model, a numerical score is calculated.\textsuperscript{67} Regardless of the nature of the inputs, creditors rely on credit scores in lending decisions to improve efficiency and reduce costs. Without credit scoring, loan underwriting would be a much more time-consuming process. The Fair Credit Reporting Act ("FCRA")\textsuperscript{68} and the Equal Credit Opportunity Act ("ECOA")\textsuperscript{69} are the two primary laws for fair lending. The FCRA’s purpose is to "require that consumer reporting agencies adopt reasonable procedures for meeting the needs of commerce for consumer credit . . . in a manner which is fair and equitable to the consumer"\textsuperscript{70} while the ECOA’s worthy goal is to "make that credit equally available to all creditworthy customers."\textsuperscript{71}

It is important to note that fair lending laws were enacted to correct credit discrimination based on attributes such as race and gender. Historically, married women were not able to obtain credit in their own name,\textsuperscript{72} and race was once an explicit factor that excluded persons of color from credit to obtain a mortgage.\textsuperscript{73} Enactment of the fair lending laws did not ameliorate the effect of past disparate treatment, which continues to impact present credit scores today; facially objective longitudinal data, for example payment history, is skewed by historical discriminatory effects. Neither is discrimination in the far away past; in 2015, for

\textsuperscript{66} Id. FICO claims that its scores are calculated by weighting payment history 35%, amounts owed 30%, length of credit history 15%, new credit 10%, and credit mix 10% but also states that "for some people, the importance of these categories can be different." What’s In My FICO Scores?, FICO, https://www.myfico.com/credit-education/whats-in-your-credit-score (last visited Feb. 28, 2021).

\textsuperscript{67} Traditional FICO scores, for example, range from 300 to 850. Citron & Pasquale, supra note 49, at 9.


\textsuperscript{69} Id. §§ 1691–1691f (West 2021).

\textsuperscript{70} Id. § 1681(b).


\textsuperscript{72} See Katherine S. Clarke, Is It a Violation of the Equal Credit Opportunity Act To Require a Spouse To Guarantee a Loan? If Not, It Should Be, 22 N.C. BANKING INST. 135, 137–38 (2018) (explaining the history of this discrimination and the ECOA purpose to address it); Leslie A. Kulick, Guaranteeing Credit for Others; The Federal Reserve Board’s “Regulation B” Requires Amendment, 67 J. Mo. BAR 224, 227 (2011) (explaining “a main purpose of the ECOA was to eradicate credit discrimination against married women who had traditionally been denied individual credit”).

\textsuperscript{73} See Lisa Rice & Deidre Swesnik, Discriminatory Effects of Credit Scoring on Communities of Color, 46 SUFFOLK U. L. REV. 935, 940–43 (2013) (describing the policies of the Home Owners Loan Corporation, Federal Housing Authority, and the Veterans’ Administration that were explicitly discriminatory). The historically disadvantaged continue to feel the impact in current data that are used in credit scoring. See Terri FRIEDLINE, BANKING ON A REVOLUTION 24 (2021).
example, a New Jersey bank paid a $33 million fine for discriminatory mortgage lending. 74

A. Fair Lending Laws

The FCRA promotes fairness in lending by requiring credit bureaus to “follow reasonable procedures to assure maximum possible accuracy of the information concerning the individual about whom the [credit] report relates.” 75 The implications here are twofold: (1) accuracy of the information is required in order to achieve fair lending decisions and (2) accuracy can be achieved through the implementation of proper procedures. 76 To that end, the FCRA includes mandatory disclosures that are intended to promote accuracy in credit reporting. For example, the credit bureaus are required to provide a consumer with (1) a credit report once a year, (2) a current credit score, and (3) a disclosure of key factors that adversely affect such credit score. 77 These disclosures give consumers at least a small degree of transparency with respect to the credit reporting and scoring system. However, a study by the FTC found that 1 in 5 consumers have errors in their credit reports and 1 in 20 have errors that would result in them being denied credit or offered more stringent credit terms. 78 Credit bureaus derive revenue from selling reports to lenders so they have reason to prioritize low predicted default rate over fairness, and consumer disputes may be seen as an expense to be minimized. 79 The FCRA only imposes civil liability on a credit bureau when it is “either negligent or willful in failing to comply with any requirement imposed under the FCRA.” 80 This standard places an

---

76 Id. § 1681i. One procedure required by the FCRA, for example, is that the credit bureaus are obligated to conduct a “reasonable reinvestigation” within 30 days of being notified of an inaccuracy in a credit report. Id. § 1681i(a)(1)(A).
77 15 U.S.C.A. § 1681g (West 2021). Note that the credit bureaus are required to provide the report for free but may charge a reasonable fee for the credit score disclosure. Id. § 1681g(f)(8).
79 Chi Chi Wu, supra note 76.
extraordinary burden on plaintiffs and, consequently, is a continuing barrier to achieving lending fairness through data accuracy.  

On the other hand, the ECOA focuses on fairness in lending by making “credit equally available to all creditworthy customers.”82 Thus, a component of fairness from the ECOA’s perspective is equal treatment of similarly situated applicants. To address discrimination in lending decisions under the ECOA and its implementing regulation, lenders are prohibited from using credit scoring systems that take into account race, color, religion, national origin, or sex to evaluate an applicant’s creditworthiness.83 The ECOA seeks to achieve this worthy goal, at least in part, through transparent procedures. For example, the ECOA’s implementing regulations require a creditor to notify an applicant of its decision within thirty days of an application and to provide the reason for an adverse decision upon request.84 Implied in this requirement is that if a creditor must disclose the reason for an adverse action, then it is less likely to discriminate.

Despite ECOA prohibitions, fair and equal lending continues to be an unsolved problem. A University of California, Berkley study found that Latino and African American borrowers were charged 7.9 and 3.6 basis points more for original and refinance mortgages, respectively, than white applicants who had the same FICO score and loan-to-value ratio—regardless of whether a loan officer or a software-based underwriter set the rates.85 The ECOA is enforced through government action and private litigation by proving discrimination through “disparate treatment” or “disparate impact.”86 Disparate treatment claims are difficult to prove because “smoking gun” evidence of lending discrimination is rare since lending discrimination is likely to be subtle,

---

sophisticated[,] and difficult to prove, especially given the use of computerized credit scoring systems to evaluate applicants.\textsuperscript{87} Disparate impact claims are even more difficult because when a plaintiff is able to establish a prima facie case, the defendant will be liable only if the plaintiff can then prove that an alternative policy could serve the business purpose with a less discriminatory effect.\textsuperscript{88} Legal scholarship has concluded that the use of data-driven decision making and artificial intelligence likely makes both the potential harms and proof of discrimination by disparate treatment harder yet.\textsuperscript{89}

B. Alternative Credit Data and AI

As described above, traditional financial data is collected and then used to create an individual credit score, which is a large part of whether a person will be granted a loan and under what terms. Due to historical and continuing discrimination,\textsuperscript{90} laws prohibit the use of protected classifications as inputs to the scores. Definitions of data outside of these categories are still evolving. However, in this article, two other kinds of data are defined as follows. Non-traditional financial data includes payment history related to rent, mobile phone, cable, and bank account cashflow.\textsuperscript{91} For purposes of clarity, in this discussion, alternative data is used to mean data that is not directly related to the financial ability to pay. Newer AI predictive scoring models, fueled by large amounts of alternative data, also consider transaction histories, physical address histories, education level, work histories, social network information, and other Internet activity.\textsuperscript{92} Worldwide, the breadth of data input into AI to derive credit risk include the length of email subject lines, the prevalence of selfies,\textsuperscript{93} type of computer used, email host, time of day an application was made, the device used


(mobile, PC, etc.), and the list goes on. AI systems find predictive correlations within alternative data, so much so that for any individual data point that "there is probably a way to integrate it into a credit model." Questions about proxy discrimination and the fairness of using non-financial alternative data in AI models are unresolved.

Of late, the Consumer Financial Protection Bureau ("CFPB"), the agency responsible for enforcement of the ECOA, has taken an "innovation friendly" approach to regulating the use of alternative data in credit decisionmaking, despite its potential for discriminatory impact. For example, despite the potential for discriminatory impact, the CFPB issued a no-action letter to UpStart Network, Inc., a company that uses alternative data to make credit underwriting and pricing decisions. The 2017 letter indicated that the agency had no "present intention to recommend initiation of an enforcement or supervisory action against Upstart with regard to application of the Equal Credit Opportunity Act." Yet, a recent study found that UpStart is charging higher interest rates on student loans to graduates of historically Black or predominately Hispanic colleges. In contrast, the Economic Growth, Regulatory Relief, and Consumer Protection Act of 2018 ("Economic Growth Act") required the Federal Housing Financing Agency ("FHFA") to create a process to validate and approve credit scoring methods, fed by consumer data and algorithms to create a credit scoring model. The statute requires that the model be validated on accuracy, reliability, and integrity. These two approaches are drastically different views of the relationship between artificial intelligence driven products and legal, social, and fairness priorities. A case study comparing these intersections of regulation and algorithmic scoring is instructive.

94 Aaron Klein, Credit Denial in the Age of AI, BROOKINGS (Apr. 11, 2019), https://www.brookings.edu/research/credit-denial-in-the-age-of-ai/.
95 Id.
99 Id. § 310.
100 Id. § 310(a)(7)(C).
C. Multiple Lenses of Fair Artificial Intelligence for Credit Scoring

It has been argued that the use of alternative data and AI automated credit decision systems could expand access to credit. On the other hand, it has been argued that alternative credit scoring is opaque and unfair. The FHFA and the CFPB approaches to addressing the potential for both good and harm are in stark contrast.

1. Credit Ratings and Mortgages: FHFA

The FHFA is the regulatory body for the secondary mortgage market that is run by organizations known as Fannie Mae and Freddie Mac. The FHFA sets overall rules for assessing credit scoring,\(^{102}\) pursuant to which Fannie Mae and Freddie Mac adopt more detailed implementation of those rules.\(^{103}\) The FHFA established a process for validating and approving credit scoring models, adopting the three guiding principles for credit scoring: accuracy, reliability, and integrity.\(^{104}\)

The accuracy standard requires that the scoring system “appropriately reflects a borrower’s propensity to repay a mortgage loan” under the relevant terms.\(^{105}\) Statistical tests are applied using past data to compare a proposed new method to a benchmark of current scoring methods.\(^{106}\) The reliability standard evaluates whether the scoring system is accurate in different economic cycles and compares data across two different time periods with significantly different economic conditions.\(^{107}\) The two different time periods are, at present, the 2009–2010 time period and the 2016–2017 time period, which would show different economic stress periods. The integrity standard is described as, “[a] Model has integrity if, when producing Credit Scores, the Model uses relevant data that reasonably encompasses the borrower’s credit history and financial performance.”\(^{108}\) The meaning of integrity is worth examining in more detail, as it relates to what information may fairly be used in a scoring algorithm/model.

When adopting the final rule, the FHFA responded to comments pointing out the difference between the objective (i.e., mathematical) foundations of accuracy and reliability and the subjective determination of integrity by acknowledging that integrity is “more subjective” but that it is

\(^{102}\) Validation and Approval of Credit Score Models Final Rule, 12 C.F.R. § 1254 (2021).

\(^{103}\) For purposes of discussion, the process is presented in its entirety without reference to the implementing body.

\(^{104}\) 12 C.F.R. § 1254.7.

\(^{105}\) Id. § 1254.7(c)(1).

\(^{106}\) FANNIE MAE & FREDDIE MAC, JOINT ENTERPRISE CREDIT SCORE SOLICITATION 12 (2020).

\(^{107}\) The two time periods used are March 2009–February 2010 and July 2016–June 2017. Id.

\(^{108}\) Id.
"necessarily" so. The integrity assessment includes, at a minimum, a review of three aspects of the data. First, the model may only use legally permissible data. Second, the assessment reviews the impact of any smoothing, truncating, censoring, or aggregating of any data elements in the model and whether data has been "omitted, modified, or discounted in the Model." Third, a Fair Lending certification is required to be executed, confirming that the following is true with regards to the credit scoring model:

1. There is no use of any variable that is a protected class or is prohibited under any laws;
2. There is no use of a variable that is highly correlated with, a proxy for, or predictive based on a correlation to protected classes;
3. There is a "reasonable, causal, and understandable relationship" between the variables and credit risk; and
4. Evaluation and testing of the model complies with Fair Lending Laws, and processes and procedures for compliance.

In sum, the FHFA transparent processes treat fair lending as an essential goal. In that vein, its response confirming the necessity of integrity, i.e. fairness, as compared to accuracy is not trivial. The three-legged stool of accuracy, reliability, and integrity are interrelated, and quantifiable or mathematical constructs do not take precedence. In turn, integrity measures require compliance with fair lending laws that prohibit the use of data related to immutable and protected characteristics about people in calculations about creditworthiness. Furthermore, the rules incorporate the concepts of causality and reasonableness that support principles of fairness. Prohibiting use of particular variables, including any proxy or predictive variables for protected classifications, showing causality, and making reasonable uses of data support the overall legitimacy of both the mortgage market and public perceptions, similar to the administrative law procedures for legitimacy. The approach also preferences individual fairness over group fairness and process over outcomes.

2. Consumer Lending: CFPB and Upstart

The approach of the CFPB, in comparison, is drastically different. In early 2021, a CFPB Taskforce on Federal Consumer Financial Law Report

---

110 FANNIE MAE, supra note 106, at 9.
111 Id.
112 Id. at 24.
addressed a broad number of consumer finance issues, over two volumes; it mentioned the use of AI and machine learning, and discussed alternative data for credit scoring in somewhat more detail. The most significant conclusion seemed to be that lenders can save money by using machine learning for credit decisions and processing. The Taskforce warned that using alternative credit data could violate FCRA, be discriminatory, and use financially unrelated correlations for decision making, yet the Taskforce concluded the discussion by noting its no-action letter extension to Upstart, an alternative data and AI credit company. It also recommended that the use of alternative data reported in consumer reports be studied; it was silent about actions relevant to the fairness of machine learning that is fueled by non-financial data. After a review of the potential violations of the ECOA and the Federal Reserve’s warning that AI models can be difficult to validate, it nonetheless expressed optimism that market forces would deliver benefits to consumers.

Examination of the CFPB no-action letter to Upstart, and its ramifications, provides further insight into its notions of fairness. Upstart describes itself as “one of the first consumer lending platforms to leverage artificial intelligence and machine learning to price credit and automate the borrowing process” by using “non-conventional variables at scale.” Although initially very little was publicly disclosed about the data feeding the artificial intelligence model, it was recently made known that Upstart creates a credit scoring model and sets terms using 1,500 variables and that although no variable is paramount, “[e]ducation characteristics are among the highly predictive variables selected by the model.”

---

114 Id. at 497–98, 500–03.
115 Id. at 498.
116 Id. at 502.
117 Id. at 519–21.
Upstart first requested a no-action letter from the CFPB in 2017.\textsuperscript{120} It explained that its model is based on a number of factors, including traditional credit scoring but also based in large part on educational and employment information.\textsuperscript{121} It stated that its model was not necessarily equally predictive across all demographic groups and that "the expected evolution of Upstart's automated underwriting model and potential changes in the applicant pool over time" made it unclear what Upstart would be required to do to address potential discrimination.\textsuperscript{122} It recited the processes and procedures it followed to ensure fair lending principles, but stated that because of its proprietary methods the no-action letter was the only avenue open for obtaining approval for its product.\textsuperscript{123} Furthermore, Upstart agreed to share information about the performance of the model confidentially and to

\begin{quote}
compare applicant outcomes from its underwriting model against outcomes that would result under a model without non-traditional variables. This will include an analysis of any different outcomes for specific applicant groups, including groups defined by race/ethnicity, sex, age, income, credit history, educational background, and other non-credit based variables.\textsuperscript{124}
\end{quote}

In 2020, Upstart requested and received an extension of the no-action letter, for the reason that "[i]n particular, there is a lack of certainty regarding the sufficiency of the analysis required to confirm that the use of AI and facially neutral alternative variables do not have an unjustified disparate impact on applicants and borrowers."\textsuperscript{125} In the second application, Upstart gave overall statistics of improved access to credit at lower rates as compared to traditional consumer underwriting, but with regards to demographics, it only stated that "[t]he model provides higher approval rates and lower interest rates for historically underserved demographics as compared to traditional models."\textsuperscript{126} It provided general statements of consumer risk including the potential for disparate impact and that the model could include "bias in the training data, or

\begin{footnotes}
\footnotetext[121]{\textit{Id.} at 1.}
\footnotetext[122]{\textit{Id.} at 8--9.}
\footnotetext[123]{\textit{Id.}}
\footnotetext[124]{\textit{Id.} at 14.}
\footnotetext[126]{\textit{Id.} at 6.}
\end{footnotes}
other errors\textsuperscript{127} that are unaccounted for. Once again, Upstart agreed to a process and procedure for fair lending compliance, to a Model Risk Assessment Plan in conjunction with the CFPB, and to confidentially share data and results.\textsuperscript{128}

The confidential nature of the agreement with Upstart in the no-action letter makes it difficult to determine what type of fairness measures their AI models meet. The CFPB later reported an update, but also stated very generally that, “[w]ith regard to fair lending testing, which compared the tested model with the traditional model, the approval rate and APR analysis results provided for minority, female, and 62 and older applicants show no disparities that require further fair lending analysis under the compliance plan.”\textsuperscript{129} However, due to an investigation by the Student Borrower Protection Center (“SBPC”) in July of 2020, more information about Upstart’s credit scoring algorithm and testing was made available by a Senate subcommittee.\textsuperscript{130}

Among the 1,500 variables input into the model, educational factors are a subset that Upstart explains “are among the highly predictive variables selected by the model.”\textsuperscript{131} Individual data collected includes: “most recent school attended, highest degree, area of study and graduation year.”\textsuperscript{132} Schools are placed into categories based primarily on “average incoming standardized test scores.”\textsuperscript{133} Upstart states that the education data adds to the model’s accuracy and predictability, and it is therefore reasonable to assume that this specific data is a significant input. This type of educational data is widely known to be “methodologically flawed, biased, and causally related to systemic discrimination,” including in a report by the CFPB itself.\textsuperscript{134} Analysis by the SBPC showed that two historically Black colleges were in the top 50% of the average SAT groupings used by Upstart, while the remainder, approximating

---

\textsuperscript{127} Id.
\textsuperscript{128} Id. at 7–8.
\textsuperscript{132} Id. at 5.
\textsuperscript{133} Id.
96% of the schools, were in the bottom 50% of the rankings. In essence, under the artificial intelligence scoring model, the school that a person attended, major, and graduation year are among factors that impact whether a loan is approved and at what terms. Indirectly, even SAT scores matter. Upstart does report increased access to credit. For example, in 2018, “Upstarts [sic] model increased approval rates for Black applicants by 28% with 17% lower APRs, compared to a traditional model developed by Upstart in accordance with specifications from the CFPB.” The improvement in lending rates and terms is positive, but it is difficult to confirm and to square with the evidence of data use that is discriminatory in scope.

Upstart does put its AI scoring system through tests for lending fairness, creating a simulated traditional credit score model specifically for purposes of comparison to their alternative data model; though more detailed information is unavailable for public review. It states that to determine whether its AI credit model is fair, it uses a “ratio test” to identify if there are any disparities based on age, gender, and race between the actual model and the simulated model. There is no further information about what ratio is adopted, except that there is a “predefined threshold” for any protected group, which if violated would make the model fail and be subject to further review. Upstart reviews the model’s fairness by means of whether it “under- or over-predicts defaults” for different groups based on predictions and outcomes. Upstart does not answer the specific question posed by Congress as to whether its scoring system has a disparate impact on protected classes, which may imply that it does not undertake such analysis or use such a threshold. Upstart summarizes its procedures to test for fairness in the following way:

These two tests combined ensure that any potential bias in Upstart’s model would be quickly identified. They identify underwriting and pricing disparities in the Upstart model compared to the status quo; stated differently, they analyze whether our model treats historically disadvantaged groups better or worse in a material way than the Traditional Model. If they do, they also assess to what extent disparities exist, if our model treats that group in a manner commensurate with true credit risk, i.e. are these outcomes warranted/accurate. In the event the model fails both these tests, additional specific

135 Welbeck & Kaufman, supra note 130.
137 Id. at 1; see also Press Release, United States Senate Committee on Banking, Housing, & Urban Affairs, Brown, Warren, and Harris Call on CFPB to Protect Borrowers from Discrimination (July, 31, 2020), https://www.banking.senate.gov/newsroom/minority/brown-warren-and-harriscall-on-cfbp-to-protect-borrowers-from-discrimination.
variable and model redevelopment tests must also be performed.\footnote{138}

In a footnote, Upstart states that it has never performed additional tests because "its model has never failed the initial tests performed for it's [sic] lender's programs."\footnote{139} However, it does provide a simple graph showing its model's predicted default rate on the left hand axis and the actual default rate on the bottom line, with three lines denoting the default rates by borrowers who are Black, Hispanic, and White. At low default rates, the three lines are close to equal, but as default rates rise the lines diverge, the Black default rate being highest and the White default rate being the lowest. Upstart concludes, "[i]f the use of alternative data in our model was introducing bias to the credit decision, this test would show lower default rates for the disadvantaged groups."\footnote{140} This interpretation is suspect, but what the graph does show is that the Upstart model using alternative non-financial data is not as predictive across classes of people at high risk/high default rates. One could as easily interpret the graph as showing a failure of the model to treat the most vulnerable people fairly and without discrimination, thus adding to a negative credit history and contributing to the continuing cycle of the credit underserved.\footnote{141}

As a result of the information that Upstart and five other companies provided, several Senators requested that the CFPB increase its oversight of fair lending and the use of alternative data for credit decisions.\footnote{142} One of the reasons included Upstart's use of educational data as highly correlated with race. On December 1, 2020, the SBPC and the Legal Defense Fund ("LDF"), an arm of the NAACP, entered into a two-year agreement with Upstart whereby a third party auditor, a civil rights legal firm, will monitor, review, and report upon the credit scoring model and whether it meets fair lending standards, specifically but not exclusively with regards to educational data.\footnote{143} Shortly thereafter, on January 9, 2021, Upstart filed a prospectus for an initial public offering, indicating it

\footnotesize{\begin{itemize}
\item Upstart Feb. Letter, \textit{supra} note 131, at 2.
\item \textit{Id.} The report states that it performed the second test anyway and shows a simple graph that plots Black, Hispanic, and White default rates compared by actual annual default rates versus predicted annual default rates. While the three categories are almost equal at low default rates, there is clear (but undeterminable from the graph) discrepancy at high default rates, with Black at higher rates of default than Hispanic, which is higher than White. Upstart believes that proves that its AI scoring is unbiased, or else the opposite would be true. This logic is debatable.
\item \textit{Id.} at 4.
\item \textit{Id.} It is also unclear if the terms of the loans across default rates were accounted for.
\item \textit{See} Press Release, United States Senate Committee on Banking, Housing, & Urban Affairs, \textit{supra} note 137.
\end{itemize}}
would become a publicly held company.\textsuperscript{144} Initial pricing of its stock indicated a public valuation of over $1.5 billion.\textsuperscript{145}

Relative to the FHFA, the CFPB takes a hands-off approach to the issue of AI and credit scoring, and lets industry lead.\textsuperscript{146} As compared to the FHFA, the CFPB process of a no-action letter and confidential review of an individual company's AI system is opaque and procedurally vague. Evidence indicates that the CFPB has prioritized increased lender efficiency and market expansion as benefits in the consumer lending process. This approach favors numerical outcomes and mathematical predictive accuracy as opposed to substantive fairness. It does not investigate or adopt a prohibition against proxy variables, for which it was roundly criticized.

IV. AN APPROACH TO AI FAIRNESS

The broad overview of mathematical fairness exemplified several different approaches to AI systems. The choices include whether to prioritize group or individual fairness, whether to allow for correlations to feed predictions or whether causality is required, and whether the model development is situated within the socio-technical context. The broad overview of legal concepts identified different strands of procedural and substantive fairness and administrative and contractual fairness principles. The case study of AI systems using non-financial alternative data for credit scoring illustrated the vast difference between regulators choosing to let AI algorithms lead, as compared to using legal principles to set standards for the creation and application of AI models. The case also illustrates the difference between mathematically preferred concepts of group and outcome fairness as allowed under the opaque CFPB no action letter as compared to individual and process fairness as followed by the FHFA approval process.

In large part, achieving AI fairness is much about optimization, a data science term, which indicates which goal will be preferred in the model design and outcomes. In data science, it has been said that, "[w]e can be confident of

\textsuperscript{144} Upstart's Initial Public Offering Prospectus, Upstart Network, Inc. (Dec. 2020), https://ir.upstart.com/node/6491/html/toct. The prospectus refers to the agreement Upstart entered into with the LDF and SBPC.

\textsuperscript{145} Tomi Kilgore, Upstart Holdings Sets IPO Terms, Could Be Valued at up to $1.6 Billion, MarketWatch (Dec. 4, 2020), https://www.msn.com/en-us/money/companies/upstart-holdings-sets-ipo-terms-could-be-valued-at-up-to-1-6-billion/ar-BB1bD71c. The stock was offered at $20 per share and, as of February 19, 2021, was trading at over $89 per share. Upstart Holdings, Inc. (UPST), Stock Analysis (Feb. 19, 2021, 1:58 PM), https://stockanalysis.com/stocks/upst/company/.

little about a trained model other than that it will perform well according to the objective function that [it] was optimized for, which is usually some narrow and myopic proxy for aggregate error or profit.\textsuperscript{147} Optimization goals in socially important areas such as consumer credit should not be limited to profit. Not every AI application has the same level of societal impact, but as the importance of the function increases, so does the imperative for an optimization that adopts “socially acceptable definitions of fairness and meaningful interventions to ensure the long-term well-being of all groups.”\textsuperscript{148} This means starting from broadly accepted legal principles, such as those found in the fair lending laws and illustrated by not using protected classifications or proxies for protected classifications as found in the FHFA certification. Regulators and legislators should be active participants in setting standards and reviewing algorithmic systems. This could mean adopting regulatory processes that approve specific tests to meet the standards of fairness, as the legislature and FHFA did for mortgage credit scores.\textsuperscript{149} It may also require the adoption of new and explicit laws that clarify or bring AI and data companies underneath existing laws that protect equality and elevate fairness.

In comparison, consumers have little choice about the 1500 data points that are used to create their Upstart credit score, and the known educational data points have been widely shown to reflect discrimination based on race. Yet, the Upstart credit model makes automated credit decisions 60% of the time through automated AI. In sum,

Algorithms generally, and especially machine learning algorithms, are good at optimizing what you ask them to optimize, but they cannot be counted on to do things you’d like them to do but didn’t ask for, nor to avoid doing things you don’t want but didn’t tell them not to do.\textsuperscript{150}

Legal leadership is needed in this regard, especially to set choice boundaries for consumer protection, so that if necessary the model development will “preference gains in fairness”\textsuperscript{151} over other goals, such as error rate or profit.

Specifying and preferencing legal and ethical fairness goals for optimization must be supported by fair processes and procedures that are

\textsuperscript{147} KEARNS & ROTH, supra note 1, at 32.

\textsuperscript{148} Ntoutsi, supra note 21, at 2.

\textsuperscript{149} See, e.g., Babak Salimi, Bill Howe & Dan Suciu, Database Repair Meets Algorithmic Fairness, 49 ACM SIGMOD Rec. 34 (2020) (stating that only omitting the classification from the data is not enough without considering underlying bias in other data).

\textsuperscript{150} KEARNS & ROTH, supra note 1, at 87.

\textsuperscript{151} Nat’l Cmty. Reinvestment Coal., Comment on Request for Information on the Equal Credit Opportunity Act (Dec. 1, 2020), https://www.regulations.gov/comment/CFPB-2020-0026-0123 (proposing that lenders should be required to choose fairness outcomes over accuracy outcomes in a three-step, iterative, balanced process that accommodates both attributes in a set ratio).
transparent and trustworthy. Mathematical design principles and legal procedural due process are perhaps the closest and most easily transferable concepts to ensure fairness. The law requires notice, an opportunity to be heard, access, and procedures that are applied equally to all individuals. Substantive due process requires regulatory processes of agencies, such as the FHFA review for credit scoring validation, to be transparent and follow accepted procedures for fairness. In addition to due process, the law also establishes the meaning of fairness as the avoidance of surprise, as in contract law. This principle can be applied to many choices in the AI community that do not rise to a due process consideration, such as using data and inferences that are surprising, and which have societal ramifications and a negative impact upon trust in AI.

Work towards AI design principles is receiving much attention in the science of fairness literature, and design principles are often discussed as ethical principles. Microsoft, for example, is one leader in the adoption of responsible AI principles that include the substantive principle of fairness, but which also include design guidelines that, like due process, operationalize fairness within an organization. In the public sector, in January 2021, the U.S. Department of Defense announced that it had established AI ethics principles for its systems engineering that will, among other things, “minimize unintended bias” and that they would work to ensure that it would provide training across the services to accomplish these goals. Human computer interaction and design disciplines are applying process principles to AI as well. For example, Yu et al., use the COMPAS case to show how a process works to reach the goals of the users when the different approaches to fairness and accuracy, each of which had different measures of effectiveness, were disputed. The first step is to identify the objectives, which can be summarized as; do not detain individuals who will not recidivate (reduce false-positives), do not parole individuals who will commit future offenses (reduce false-negatives), and apply a fair release rule across demographics, which is defined as no disparity in either false positives or false-negatives. Furthermore, the research team adopted a process that marshals the various algorithmic models that could be applied and, then, creates a series of


155 Id. at 1248.
visualizations that can be used to show the changes in impact with different decisions.\textsuperscript{156} Importantly, the visual was created to communicate the impact to non-technical decision makers.\textsuperscript{157} It is essential that law and regulatory fundamentals are included in the pursuit for AI fairness, as constituting part of essential and optimized goals for socially responsive and trusted systems.

Optimizing legal principles of fairness in an AI system is essential when it affects fundamental decisions about individuals and their access to important economic and societal benefits. Credit scoring is one of those situations. The two studies of regulatory input into credit scoring using alternative data for algorithmic models, the FHFA and the CFPB approaches, show how leading with regulatory fairness principles as compared with leading with mathematical accuracy principles produce strikingly different results. Regulatory agencies do not need to stand back to let algorithmic fairness concepts take precedence over legally established principles of fairness. Data scientists should embed the application of legal fairness principles into the process of creating algorithmic models at every stage, which could avoid conflicts such as the COMPAS dispute over the optimization goals, after the fact. Courts and regulators should not hesitate to require legal fairness principles to be met by algorithmic models and AI systems, and data scientists should appreciate that optimized goals of procedural and substantive due process are fundamental to both fairness and accuracy within socio-technical AI systems.

V. Conclusion

Mathematical and legal definitions of fairness are both important concepts related to AI systems and their design. However, rather than being considered a constraint for achieving another optimized goal for accuracy, situated concepts of legal fairness should inform and define AI design goals. The case study of AI alternative data credit scoring illustrated the distinct results between similar systems, depending on whether the design for fairness was proactively led by legal standards or whether the AI was developed outside of a legally led framework. In the context of fair lending, the impact of leading legal principles on AI fairness included transparency over confidential processes, causality rather than correlations used in AI design, non-discrimination over algorithmic predictions, and individual data protection over consumer surprise. A socio-technical approach to fairness that implements legal constitutions of fairness within mathematical AI systems would improve the legitimacy and benefits of those systems while creating consumer trust. Regulatory leadership is necessary to create the circumstances for that result.

\textsuperscript{156} Id. at 1248–49.
\textsuperscript{157} Id. at 1249.