european corporate governance institute

Consumer Credit in The Age of AI – Beyond Anti-Discrimination Law

Law Working Paper N° 663/2022
February 2023

Katja Langenbucher
Goethe University Frankfurt, SciencesPo, Fordham Law School and ECGI

© Katja Langenbucher 2023. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

This paper can be downloaded without charge from:
http://ssrn.com/abstract_id=4275723

https://ecgi.global/content/working-papers

european corporate governance institute

ECGI Working Paper Series in Law

Consumer Credit in The Age of AI – Beyond Anti-Discrimination Law

Working Paper N° 663/2022
February 2023

Katja Langenbucher

Katja is grateful to the DFG Project „Fair Scoring"/Law and Finance for LA1358/6-1. She also thanks the DFG Center for Advance their support (project FOR 2774). I am immensely grateful for having had the opportunity to present versions of this paper and receive invaluable feedback at: Law and Society/Lisbon; PennLaw Adhoc faculty workshop; AI and democracy conference/SciencesPo; European Central Bank legal conference/ Frankfurt; Fintech conference University of Hamburg; Fintech Symposion University of Mannheim; ECFR conference University of Helsinki; NYU Law School PRG group; RegHorizon conference ETH Zürich; Edinburgh University FinTech Lecture; FinCoNet conference OECD; FinTech conference Luxembourg; Max-Planck-Institute Bonn. I am especially grateful to Marion Fourcade, Talia Gillis, Sandy Maysoon, Katherine Strandburg and Olivier Sylvain for critical comments and discussion. All remaining errors are mine.

© Katja Langenbucher 2023. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Abstract

Search costs for lenders when evaluating potential borrowers are driven by the quality of the underwriting model and access to data. Both have undergone radical change over the last years, due to the advent of big data and machine learning. For some, this holds the promise of inclusion. Invisible prime applicants can perform better under AI than under traditional metrics. Broader data and more refined models help to detect them without triggering prohibitive costs. However, not all applicants profit to the same extent. Historic training data shape algorithms, biases distort results, and data as well as model quality are not always assured. Against this background, a debate over algorithmic discrimination has developed. So far, it has centered on the US with its legal framework dating back to the Civil Rights Act of the 1970s. With the AI Act and the reform of the Consumer Credit Directive, EU lawmakers have been catching on. This paper explores the EU and the US legal framework on anti-discrimination law. It submits that both face fundamental difficulties when fitting algorithmic discrimination into the traditional regime. Against this background, the paper suggests for algorithmic underwriting to reorient the discussion towards a better design of financial regulation.

Keywords: credit scoring methodology, AI enabled credit scoring, AI borrower classification, responsible lending, credit scoring regulation, financial privacy, statistical discrimination

JEL Classifications: C18, C32, K12, K23, K33, K40, J14, O31, O33

Katja Langenbucher
Professor
Goethe University Frankfurt
Theodor-W.-Adorno-Platz 3
60323 Frankfurt am Main, Germany
e-mail: langenbucher@jur.uni-frankfurt.de

LawFin Working Paper No. 42

Consumer Credit in the Age of AI – Beyond Anti-discrimination Law

Katja Langenbucher

DFG Center for Advanced Studies on the Foundations of Law and Finance | House of Finance | Goethe University
Theodor-W.-Adorno-Platz 3 | 60629 Frankfurt am Main | Germany
www.lawfin.uni-frankfurt.de

CONSUMER CREDIT IN THE AGE OF AI – BEYOND ANTI-DISCRIMINATION LAW

Katja Langenbucher[*]

ABSTRACT

Search costs for lenders when evaluating potential borrowers are driven by the quality of the underwriting model and access to data. Both have undergone radical change over the last years, due to the advent of big data and machine learning. For some, this holds the promise of inclusion. Invisible prime applicants can perform better under AI than under traditional metrics. Broader data and more refined models help to detect them without triggering prohibitive costs. However, not all applicants profit to the same extent. Historic training data shape algorithms, biases distort results, and data as well as model quality are not always assured. Against this background, a debate over algorithmic discrimination has developed. So far, it has centered on the US with its legal framework dating back to the Civil Rights Act of the 1970s. With the AI Act and the reform of the Consumer Credit Directive, EU lawmakers have been catching on.

This paper explores the EU and the US legal framework on anti-discrimination law. It submits that both face fundamental difficulties when fitting algorithmic discrimination into the traditional regime. Against this background, the paper suggests for algorithmic underwriting to reorient the discussion towards a better design of financial regulation.

TABLE OF CONTENTS

[*] Katja is a law professor at Goethe University, Frankfurt, Germany, She is affiliated at SciencesPo, Paris, France, and regular visiting faculty at Fordham Law School, NYC, US. Katja is grateful to the DFG Project „Fair Scoring"/ LA1358/6-1. She also thanks the DFG Center for Advanced Studies on the Foundations of Law and Finance for their support (project FOR 2774). I am immensely grateful for having had the opportunity to present versions of this paper and receive invaluable feedback at: Law and Society/Lisbon; PennLaw Adhoc faculty workshop; AI and democracy conference/SciencesPo; European Central Bank legal conference/Frankfurt; Fintech conference University of Hamburg; Fintech Symposion University of Mannheim; ECFR conference University of Helsinki; NYU Law School PRG group; RegHorizon conference ETH Zürich; Edinburgh University FinTech Lecture; FinCoNet conference OECD; FinTech conference Luxembourg; Max-Planck-Institute Bonn. I am especially grateful to Marion Fourcade, Talia Gillis, Sandy Mayson, Katherine Strandburg, and Olivier Sylvain for critical comments and discussion. All remaining errors are mine.

A. Introduction

The decision to hand out a loan and price interest rates includes an assessment of the borrower's credit risk. "Good" borrowers are separated from "bad" ones, credit default risk is predicted, and interest rates are calculated. Naturally, this involves distinguishing among applicants to make an informed choice. In the process, different groups of applicants emerge, some with excellent chances of obtaining attractively priced credit, others with reasonable chances, and some with low or no chances of affording interest rate payments or qualifying for a loan.

Lenders face transaction costs, uncertainty about an applicant's credit default risk and they operate in an environment of imperfect competition. To reconstruct hidden fundamental information about borrowers they rely on observable variables.[1] Historically, signals such as "capital, capacity, and character" were important clues towards fundamental information. Establishing them was a core part of the daily work of credit managers. Until the 1960s, a variable as qualitative and vague as "character" was "considered the foundation of consumer creditworthiness".[2]

Starting in the 1930s, lenders profited from advances in statistics. Reasonably good forecasts could be established based on a limited list of input variables. When deciding which input variables to use, lenders faced a choice:[3] Comprehensive checks were slow in information gathering. They could identify most attractive candidates but were costly. Alternatively, a focus on limited and standardized input variables allowed for quick decisions which captured many, if not all potentially attractive borrowers. For most lenders, privileging speed and volume over comprehensive searches seemed, on balance, more attractive.[4] This led to enormous market expansion based on standardized decision-making criteria.[5] Politically, statistical approaches of this type were looked upon favorably. They seemed to replace "vague intuitions, personal prejudices, and arbitrary opinions" by something more mathematical and objective.[6] They also held the promise of inclusion for groups which had found it hard to gain access to the financial system.

Today, a similar development is emerging with the advent of big data and machine learning algorithms.[7] Digital technology has achieved more efficient and lower-cost delivery of financial services than ever before.[8] Lenders can access data far beyond traditional financial variables, without compromising on speed and volume. One example is the use of borrower's cash flow and data with the lender. However, as the popular remark "All data is credit data. We just don't know how to use it

---

[1] Bartlett et al. (2022); Brito/Hartley (1995); Parlour/Rajan (2001); Stiglitz/Weiss (1981); see Guseva/Rona-Tas (2001) on uncertainty and institutions which allow for reducing uncertainty to measurable risk.
[2] Lauer (2017), pp. 199 et seq. on the five variables used by the mail-order firm Spiegel in the 1930s; tracing the historical development: Citron/Pasquale (2014), pp. 8 et seq.
[3] Lauer (2017), p. 210.
[4] Lauer (2017), p. 210.
[5] Burrell/Fourcade (2021), p. 222 ("national trust infrastructure").
[6] Burrell/Fourcade (2021), p. 222.
[7] Burrell/Fourcade (2021), p. 222; Citron/Pasquale (2014), p. 4.
[8] FSB (2022), p. 11, worrying at the same time that some incumbent financial institutions prioritize market share through sales rather than operating profit.

yet"[9] suggests, there is much more to explore. Online payment history, performance on lending platforms, age or sex, job or college education, ZIP code, income or ethnic background can all be relevant to predict credit default risk. Depending on a jurisdiction's privacy laws more variables can be scrutinized. This includes, for instance, preferred shopping places, social media friends, political party affiliation, taste in music, number of typos in text messages, brand of smartphone, speed in clicking through a captcha exercise, daily work-out time, or performance in a psychometric assessment.

At first glance, availability of data along those lines promises to enhance access to credit markets, due to lower search costs for lenders. Lenders are aware that an applicant with a low credit score might still be a good credit risk and present an attractive business case for the lender. However, it is often not cost-efficient for the lender to invest in locating such "invisible prime"[10] applicants. The lender will compare his search costs to the expected return on the loans he could hand out to applicants whose credit default risk is lower than what their score suggests. In the past, most lenders have found search costs for invisible primes to far outweigh the expected return.[11] As described in more detail below,[12] access to big data and ease of modelling via machine learning have significantly lowered search costs. For some invisible prime borrowers, this has raised hopes for inclusion through AI.[13] At the same time, data on the extent of inclusion along those lines is still sparse and newer studies caution that minorities profit only disproportionately.[14]

Framed as a question of economic efficiency from the micro-perspective of the lender, there is nothing wrong with using as many variables as cost-efficient. Market forces are assumed to single out the meaningful signals, lowering costs of information for the lender. Signaling of this type allows for risk-adjusted pricing according to the lender's business model.

By contrast, understood as a question of anti-discrimination law, some signals must not be used. Under US law, the Fair Housing Act (FHA), Title VIII of the Civil Rights Act of 1968, proscribes rejection of a mortgage which is motivated by race, color, religion, sex, disability, familial status or national origin.[15] The US Equal Credit Opportunities Act (ECOA) prohibits to deny a loan because of race, color, religion, sex, disability, marital status, age, national origin, receipt of income from a public assistance program or an applicant's good-faith exercise of any right under the Consumer Protection Act. Under EU law, various Directives prohibit specific instances of discrimination

---

[9] See https://archive.nytimes.com/query.nytimes.com/gst/fullpage-9A0CE7DD153CF936A15750C0A9649D8B63.html; discussion at Rosamond (2016).
[10] Term proposed by Di Maggio et al. (2021), p. 2.
[11] Lauer (2017), p. 210.
[12] See below B.
[13] See recital (49) of the Proposal for a Directive on Consumer Credits of 30 June 2021, COM(2021)347 final. For better readability I refer to this text. At the time of writing, the proposal is pending and the European Parliament has issued its final report, COM(2021)0347, issued on 25.8.2022. Where relevant, I include proposed amendments to the text proposed by the Commission: European Parliament, Report – A9-0212/2022, available at https://www.europarl.europa.eu/doceo/document/A-9-2022-0212_EN.html.
[14] On Fintech more generally: FSB (2022), p. 12.
[15] Below D.II.1.a.

between private citizens.[16] Most focus on employment discrimination, some rules cover access to goods and services which are publicly available.[17] Credit underwriting is not usually understood as a publicly available service. Against this background, discriminatory lending practices have for a long time remained a matter of private law of the Member States.[18] The 2021 proposal for a reform of the EU Consumer Credit Directive[19] includes a novel rule that explicitly targets discrimination in the consumer loan context.[20]

Anti-discrimination laws are applicable irrespective of whether lenders feel they are rejecting the applicant *because of* attributes which these laws protect. Lenders might wish to point out that they deny a loan because of high credit default risk, for which the protected attribute is merely a useful signal. However, as I will explore in more detail below, this is not an accepted defense in an anti-discrimination case. It is prohibited for the lender to use an observable protected attribute, such as sex, to form beliefs about unobservable attributes, such as credit default risk.

Instead of relying explicitly on protected characteristics lenders might use other variables which correlate with a protected attribute. One example is part-time employment which often correlates with sex.[21] More current examples concern big data, where first name, taste in music or preferred shopping places can correlate with race, sex, or religion.[22] As discussed in more detail below, the fact alone that a variable correlates (even narrowly) with a protected attribute is not *per se* grounds for an anti-discrimination case. Instead, plaintiffs must show that using the relevant variable triggers clusters of protected communities which are disproportionately faced with credit rejections or disadvantageous pricing. If this is the case, using the variable can violate anti-discrimination laws, even though the variable is facially neutral and does not fall under the list of protected attributes. US law addresses the concept as "disparate impact", EU law as "indirect discrimination".[23] Both legal principles combine individual protection against discrimination with combatting structural exclusion.[24] There is no US Supreme Court guidance so far on whether disparate impact doctrine applies to retail credit decisions.[25] By contrast, secondary EU law, when prohibiting discrimination, has so far understood this to apply to both, direct and indirect discrimination. The anti-discrimination Directives mentioned

---

16 Directive 2000/78/EC establishes a general framework for equal treatment and occupation; Directive 2006/54/EC concerns the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation; Directive 2000/43/EC implementing the principle of equal treatment between persons irrespective of racial or ethnic origin. The Directive addresses employment, vocational training, social protection, education and access to publicly available goods or services, including housing; Directive 2004/113/EC implements the principle of equal treatment between men and women in the access to and supply of public goods and services.

17 Below D.II.1.b.

18 For Germany see sec. 19, 20 *Allgemeines Gleichbehandlungsgesetz* (General Act on Equal Treatment).

19 Consumer Credit Directive/2021.

20 Below D.II.1.b.

21 ECLI:EU:C:1999:60 (part-time employment is a paradigm case in EU anti-discrimination law).

22 For econometricians trying to measure bias, this can lead to omitted-variable-bias if the relevant variable is not observed by the econometrician. The model will then attribute the effect of the missing variable to the observable variables which were included, Dobbie et al. (2019), p. 1.

23 On the reception of US disparate impact principles into EU Law: Mangold/Payandeh in Mangold/Payandeh (2022), note 21, 105.

24 Mangold/Payandeh in Mangold/Payandeh (2022), note 106.

25 Langenbucher (2020) on comparing this to the EU approach on indirect discrimination; see Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc., 135 S.Ct. 2507 (2015) on applying disparate impact doctrine to the Fair Housing Act which prohibits discrimination in home lending.

above refer to an "apparently neutral provision, criterion or practice" which puts protected groups "at a particular disadvantage compared with other persons, unless that provision, criterion or practice is objectively justified by a legitimate aim and the means of achieving that aim are appropriate and necessary".[26] However, the Consumer Credit Directive/2021 does not include a rule to that effect. Even assuming that credit underwriting qualifies under US disparate impact doctrine or EU indirect discrimination principles, I describe below why it is rarely straightforward to establish a case.[27]

The rest of the paper is structured as follows: Section B briefly presents various empirical studies on AI scoring and inclusion which have emerged over the last years. Section C summarizes several computer scientist's concerns with historically biased training data and algorithmic fairness. In Section D the paper moves on to an investigation of AI underwriting under the lens of US and EU anti-discrimination law. Section E suggests to reorient the discussion to designing better financial regulation law and offers a rough and preliminary outline of questions a future design will face.

The paper makes two main contributions to the debate. First, it brings out the obstacles received anti-discrimination laws face when trying to capture discriminatory credit underwriting. A core reason for this is the traditional understanding of individual attributes or practices triggering a decision. Its conceptual underpinnings provide an ill fit with AI-based decision-making based on bundles of correlating variables which are often opaque even for the decider. Second, the paper highlights areas of financial regulation law where further policy work is required to adequately capture novel challenges of AI credit underwriting.

B. Good Intentions: Inclusionary AI

Business models underlying credit pricing vary, depending on the information asymmetry between borrower and lender, the availability, and the cost of information. Scoring agencies deliver first signals, traditionally based on a short list of input factors. Most lenders refine these assessments by adding information of their own, using, for instance, cash flow and transaction data or proprietary algorithms to categorize potential borrowers. Empirically, US retail credit markets seem to move along highly standardized metrics.[28] The Consumer Financial Protection Bureau (CFPB) has found that input variables and modelling techniques are very similar across traditional lenders.[29] Multivariate linear regression analysis has been used to correlate a short list of (historically grown) variables (indicating past credit history and current credit usage) to consumer credit outcomes.[30] Compared with a sample of previous consumers with similar attributes, this is used to predict credit default risk.[31]

---

[26] Art. 2 para. 2 lit. b EU Directive 2000/43.
[27] See below D.
[28] On a troubling symbiosis between traditional and payday lenders see Di Maggio et al. (2021).
[29] CFPB (2017), p. 6; on imperfect competition see Parlour/Rajan (2001).
[30] Aggarwal (2021), p. 46.
[31] CFPB (2017), p. 6.

I. Broadening access to credit for some…

Applicants who are more costly to evaluate than the average often face obstacles in applying for a loan, even if they are low risk applicants.[32] Sometimes this is due to the scarcity of available information. In other cases the available information transmits a wrong or less meaningful picture of credit default risk.[33] For yet others, their future potential is not adequately reflected.[34] Such "thin-file applicants" or "credit invisible" have more recently been targeted by Fintech lenders who exploit digital big data footprints to offer lower cost scoring.[35] Those who perform well under the novel combination of AI and big data stand to gain when compared with standard metrics. This development has raised hopes of financial inclusion.[36]

One important source of data are mobile phones. A study by Agarwal et al. (2021) on Indian markets lists data collected from an individual's mobile phone.[37] When categorizing data, the authors work, for instance, with mobile phone data such as apps installed or operating systems, with data referring to the presence of social apps, preferred social networks or the number of contacts and calls, and, lastly, what the authors call "deep social footprint", understood as information obtained from call log patterns.[38] They provide many examples to illustrate their findings. Consumers who have no financial app installed increase their probability of default by 25 %. Having a dating app installed raises credit default risk by 17 %. Results hold after controlling for salary, age and education.[39] The authors find that a model that relies exclusively on mobile and social footprints outperforms traditional credit score models by 7 %.[40] They interpret this as a way to access "aspects of individuals' behavior which has implications for the likelihood of default"[41] – a thought not too far from last century's focus on "character".[42] Additionally, the authors use variables which measure frequency and duration of incoming, outgoing and missed calls as a way to capture an individual's "social capital". Again, they find that these measures are strongly correlated with the likelihood of default.[43]

Berg et al. (2018) analyze simple and easily accessible digital footprint variables using a data set of a German E-Commerce company. The company was interested in evaluating credit risk because goods are shipped first and paid for later.[44] Their set contains only ten digital footprint variables. Among those are the device type, the operating system, the time of day of the purchase and a dummy for a typing error. The authors find, for example, that the difference in default rates between customers

---

[32] On "error costs" in a different context but making a similar point: Hellman (2020), pp. 829, 836 et seq. on understanding error costs in their normative context.
[33] The CFPB estimates that 26 million Americans have no file with the major credit bureaus while another 19 million are unsociable because their credit file is either too thin or too stale to generate a reliable score, CFPB (2017), pp. 6–7.
[34] Di Maggio et al. (2021), p. 3.
[35] CFPB (2017), p. 15.
[36] Balyuk (2021); Bartlett et al. (2022), p. 55.
[37] Agarwal et al. (2021), p. 4.
[38] Agarwal et al. (2021), p. 4.
[39] Agarwal et al. (2021), p. 5.
[40] Agarwal et al. (2021), p. 5.
[41] Agarwal et al. (2021), p. 5.
[42] Lauer (2017), p. 199.
[43] Agarwal et al. (2021), p. 6.
[44] Berg et al. (2018) p. 2.

using an Apple and customers using an Android device is equivalent to the difference in default rates between a median credit score and the 80th percentile of the credit score.[45] The authors suggest that the variables they investigated provide a "proxy for income, character and reputation".[46]

The US online lending company Upstart provides another especially well-documented example for the promise of financial inclusion.[47] Upstart claims to outperform traditional scoring outfits not only as to all borrowers, but specifically as to those with traditionally low credit scores.[48] This covers approval decisions as well as interest rates.[49] Di Maggio et al. (2021)[50] find that "more than 30 % of borrowers with credit scores of less than 680 funded by Upstart over our sample period would have been rejected by the traditional model. We further find that this fraction declines as credit score increases, that is the mismatch between the traditional and the Upstart model is magnified among low-credit score borrower". The CFPB, which has investigated Upstart's business model, added that applicants with FICO scores from 620 to 660 were approved twice as frequently by Upstart if compared with a hypothetical lender using FICO. Applicants under the age of 25 were 32 % more likely to be approved and consumers with incomes under $50,000 were 13 % more likely to be approved.[51]

It is difficult to predict which variables will enhance an applicant's chances of being considered for a loan.[52] Education is one of the variables relied upon by Upstart which found that the better the educational background the higher the chances of profiting from Upstart's offer. By contrast, Agarwal et al. (2021), based on data from mobile phones and deep social footprints in India, found that the marginal benefits were likely to be higher for borrowers with low levels of education, indicating that relevant proxies and outcomes vary significantly across Fintech lenders and countries.[53]

There are more studies pointing towards financial inclusion through credit scoring based on alternative data. Agarwal et al. (2021) use proprietary data from a Fintech lender and find that mobile footprint, social footprint, and deep social footprints can expand credit access without adversely impacting the default outcomes. They suggest that 13 % of borrowers who were denied credit would be approved under the authors' alternate credit scoring model.[54] Bartlett et al. (2022) investigate the price of mortgages. They find that, today, Latinx and African American borrowers pay 7.9 and 3.6 basis points more in interest for home purchase and refinance mortgages because of discrimination. By contrast, Fintech algorithms discriminate 40 % less, with Latinx and African American borrowers paying 5.3 more in interest for purchase mortgages and 2.0 basis points for refinance mortgages originated on FinTech platforms.[55]

---

[45] Berg et al. (2018), p. 3.
[46] Berg et al. (2018), p. 3.
[47] Langenbucher/Corcoran (2022), p. 141; Di Maggio et al. (2021), p. 2.
[48] Di Maggio et al. (2021), p 3.
[49] Di Maggio et al. (2021), p. 4.
[50] Di Maggio et al. (2021), p. 4.
[51] CFPB (2019).
[52] Wachter (2022), p. 1 "algorithmic groups".
[53] Agarwal et al. (2021) using data from one of the largest Fintech lending firms in India.
[54] Agarwal et al. (2021), p. 8.
[55] Bartlett et al. (2022), pp. 31–32.

II. …. bringing out inequality for others

Empirical data does not always prove that algorithmic credit scoring will provide better access for all applicants. Fuster et al. (2022), using US data, show that a machine learning model provides only slightly better access and only marginally reduce disparity in acceptance rates. They find more pronounced cross-group disparity, especially very different interest rates across protected and not-protected communities, a concern I explore further below.[56]

Under the theoretical assumptions that the retail credit market is competitive, that lenders are risk-neutral and that they set interest rates contingent on borrowers' observable characteristics, we would expect to be able to explain differences in access to credit and in pricing of loans with differences in credit risk.[57] Inequalities in access to loans and differences in pricing for minority groups could then be understood as statistical discrimination: a population of loan applicants separated into groups according to their risk profile.[58] Statistical discrimination is the natural result of an efficient evaluation of credit risk. Under the same assumptions, we would not expect a situation commonly described as taste-based discrimination.[59] Taste-based discrimination occurs where individual preferences of lenders, such as a dislike of certain minority groups, provide the best explanation for observed inequality.[60] If lenders engage in taste-based discrimination, the population of loan applicants will fall into groups which do not correspond to their risk profile. Instead, membership in a group of loan applicants is determined by the lender's subjective preferences which do not (necessarily) correlate with credit default risk. For this reason, taste-based discrimination is not a rational reaction to the lender's situation of uncertainty. Rather, it can lead to inefficient rejections, if the lender's subjective preference for a specific set of loan applicants does not reflect their lower credit default risk.[61] In theory, taste-based discrimination is not expected to survive in a competitive market, given that it would indicate lenders are not objectively profit-maximizing.[62]

These theoretical forecasts are not always reflected in empirical data. Bartlett et al. (2022) explore data on the US Government Sponsored Enterprises (GSE), Fannie Mae and Freddie Mac. These GSE charge each loan a guaranty fee that depends only on the borrower's credit score and loan-to-value (LTV) ratio. In return, lenders are guaranteed against credit risk. The authors assume that interest rate differences between loans with identical credit score and LTV cannot reflect differential credit risk but must go back to some form of discrimination. Using this strategy, they find a mark-up of 7.9 basis

[56] Fuster et al. (2022), at p. 9 find the increase to be double the magnitude for Black and White Hispanic borrowers than for white non-Hispanic borrowers, see below D.III.1. on this point.
[57] Assumptions set by Fuster et al. (2022), p. 38.
[58] Arrow (1971); Phelps (1972); in the context of AI underwriting models: Hurlin et al. (2021), p. 6.
[59] Magen in Mangold/Payandeh (2022), note 58; von Ungern-Sternberg in Mangold/Payandeh (2022), note 11 et seq.
[60] Becker (1957); Becker (1993); newer models refer to stereotypes, assuming that loan examiners systematically underestimate the long-run profits of lending to minority applicants, Bordalo et al. (2016). Both models cannot explain why bias would persist in competitive markets, Dobbie et al. (2019), p. 7.
[61] See Bartlett et al. (2022), p. 32 for an example of unprofitable discrimination; further Hurlin et al. (2021), p. 6.
[62] Arrow (1971); Phelps (1972); Hurlin et al. (2021), p. 6.

points for purchase mortgages and 3.6 basis points for refinance mortgages for Latinx and African American borrowers. This results in borrowers paying $765 million in extra interest per year.[63]

Fuster et al. (2022) find that "a large fraction of borrowers who belong to the majority group (…) experience lower estimated default propensities under the machine learning technology" but "these benefits do not accrue to some minority race and ethnic groups (…) to the same degree."[64] The authors show that, for minorities, better technology produces "winners" and "losers". Winners are disproportionately white Hispanic and Asian. In Black and non-white Hispanic populations there are roughly equal fractions of winners and losers.[65]

Working with the services offered by Upstart, a US-based NGO ran a form of mystery shopping exercise. Applicants were construed as identical except for the college they had attended. Holding all other inputs constant, the authors of the study found that a hypothetical applicant who attended Howard University, a HBCU, would pay higher origination fees and higher interest rates over the life of their loans than an applicant who attended NYU.[66] Similar results were observed for applicants who attended NMSU, a HSI.[67]

There is a variety of hypotheses to explain these empirical results. We might be looking at taste-based discrimination, persistent despite its economic inefficiency. Alternatively, lenders might engage in strategic pricing, charging higher interest rates for protected communities because of their higher willingness to sign above-market. Along those lines, Bartlett et al. (2022) speculate that their findings might have to do with minority borrowers being prone to less comparison shopping on average, having less experience or acting in a more urgent time frame.[68] The observed unequal output across groups would then be a result of lenders targeting groups which - statistically - will be more vulnerable, hence open to predatory pricing.

C. Biases and Algorithmic Unfairness

AI underwriting models have raised high hopes when compared with the limited list of input variables of traditional scoring bureaus as well as the biases and cognitive limitations of human credit officers.[69] At the same time, computer scientists caution against mistaking predictions based on big data and AI as objective forecasts.[70] Rather, the predictive power of an algorithm depends on factors such as data and model quality and context of its use. Using a spam filter as an illustration, Kim lists conditions

[63] Bartlett et al. (2022), p. 31.
[64] Fuster et al. (2022), p. 8.
[65] Fuster et al. (2022), pp. 31–32.
[66] Student Borrower Protection Center (2020), methodology described at p. 16.
[67] Student Borrower Protection Center (2020), findings described at pp. 18–19.
[68] Bartlett et al. (2022), p. 32: "the fact that the relation between the rate differential and either credit score or realized default is minor suggests the income and LTV results may instead reflect something else, such as the correlation between income, financial sophistication, and a propensity to shop for rates"; similarly Gillis (2022), p. 39 ("personalized pricing"); Hurlin et al. (2021) "lack of fairness".
[69] Sunstein (2019).
[70] For a critical discussion see Blattner/Nelson (2021); Kaminski (2019), p. 1538; O'Neil (2016); from the perspective of sociology: Burrell/Fourcade (2021), p. 224; Burk (2021), p. 1163: AI as a "prosthetic extension of human judgment"; Kiviat (2019a), p. 283; Kiviat (2019b), p. 1134; Kim (2022), p. 1 on the promise of an evidence-based approach.

for AI algorithms to be efficient: the target variable they look to is clear, the answer is a straightforward yes/no binary choice, there is a correct outcome, an error the algorithm makes can be used to refine it and unbiased data is readily available.[71] For credit scoring and underwriting, not all conditions are met. To name just a few: The target variable "credit-default risk" refers to non-performance on a loan but it is only one possible target which the algorithm could optimize. It might look merely to a binary answer along the lines of: did the borrower perform or not? Instead, the lender might have a variety of target variables in mind when maximizing his return on a payday loan and looking for maximum payback over a short period of time. There will often be shades of grey which are necessary to understand the risk a potential borrower poses. A predatory loan will be tougher to pay back than a standard, market-priced loan.[72] An unforeseen event, macro-economic, hence affecting everyone, or micro-economic, affecting only the borrower, might explain the non-performance, without being adequately reflected in the data.

Additionally, for credit underwriting flawed predictions are often impossible to fix because there is no counterfactual data. The AI learns which loans it should not have offered because borrowers did worse than predicted (false positives). But it does not learn which loans would have been attractive, for the simple reason that they were not offered to the applicant.[73] Another explanation has to do with training data. Current AI underwriting algorithms use machine-learning and existing datasets on borrowers. Such datasets include their performance in repaying loans in the past as well as alternative data of the type I have described above.[74] Based on the training data set, the AI learns to connect alternative data to performance. This allows the AI to judge and evaluate future applicants for a loan. The more similar the applicant's characteristics are to the characteristics of borrowers which were successful in the past, the better the score the AI attributes to this applicant: Yesterday's world shapes today's predictions.[75]

I. The credit-default risk target variable

The quality of an AI's prediction is only as good as the match between the world according to the training data and the world as it is today.[76] If the training data reflect past inequality, then an applicant who shares features with a historically underserved group will be flagged as a higher credit risk than a comparable applicant who does not share the relevant feature (historic bias).[77] The fact that training data are, in this way, shaped by history has direct implications for how the AI builds its model.[78] Variables it finds for most candidates which were successful in the past will be accorded most weight,

---

[71] Kim (2022), p. 3, for a similar list see Kaminski (2019), p. 1539 (clear and mathematical objective, detailed and direct data, transparent inputs and code, easily verifiable outcome, fair and accurate output).
[72] See below C.I.; D.III.1.b.
[73] See below C.III; Kim (2022), p. 5.
[74] See above B.
[75] Mayson (2019), p. 2251: "look to the past as a guide to the future".
[76] Hellman (2020), p. 841 ("compounding injustice"), p. 842 ("carrying forward injustice"); Mayson (2019), p. 2251: "The premise of prediction is that, absent intervention, history will repeat itself".
[77] See below C.III. on market forces contributing to ameliorating models.
[78] Blattner/Nelson (2021), p. 12 ("model bias").

for instance a specific sex or race. Candidates whose profile does not include the relevant positive variable will face a risk premium (majority bias).[79]

The same logic applies to variables which send a negative signal. The AI learns from historical data and singles out variables which have in the past been a predictor for high credit-default risk. Applicants whose profile includes the risk-variable see their credit score sink. This happens even if a particular risk-variable does not reflect relevant details of the default situation across all applicants. The same is true if the observed risk-variable is less informative for some applicants if compared with others.[80] Take, for instance, an AI model which optimizes credit-default risk. It predicts the probability of non-performance on loan and interest rate payments across borrowers. To illustrate the problem, assume that one group of the population is especially vulnerable to signing predatory loans. For some of these borrowers the reason for default might be the inability to pay the predatory interest rate. Hence, there is a good chance that the same person would have been able to perform, had she received the same loan under the same circumstances, but with a market-standard interest rate. However, if the model attributes most weight to non-performance, a borrower who fails to pay back a loan with a predatory interest rate will be treated the same as a borrower who fails to pay back a loan with a market-standard interest rate.[81] If the group of borrowers which are (or were in the past) likely to accept predatory loans share a protected characteristic such as race, any applicant with that characteristic will be accorded a higher risk factor. A "colorblind" algorithm would be restricted from noticing the difference and – in that way – hurt, rather than help applicants of that race.[82]

II. Redundant Encoding

The flaws due to algorithmic biases are not per se novel concerns or worries which are specific to AI underwriting. Traditional scoring models, for instance the US FICO score, with their limited number of input variables, provide a much cruder picture than today's AI underwriting models.[83] The FICO score methodology is shaped by path-dependent historical choices of relevant variables and of the balance struck between accuracy, search costs and market expansion.[84] It is common knowledge that applicants who do not fit the FICO profile find it hard to qualify for a loan.

---

[79] Barocas/Selbst (2016), p. 689; Gillis (2022); Graham (2021), p. 211; Langenbucher (2020); Burk (2021), p. 1163; von Ungern-Sternberg in Mangold/Payandeh (2022), note 15 et seq.

[80] EDPB/EDPS (2021); Gillis (2022): "biased measurements"; Burk (2021), p. 1164 on the lack of context for late payments with the conclusion that credit scoring entail a "reproduction of social context", although it was "originally intended to help neutralize bias in lending".

[81] See for a similar problem in the context of predicting crime Mayson (2019), p. 2263 (three prior arrests in New Orleans were for a black man not unusual since "black men were arrested all the time for trivial things", by contrast, the same variable (three prior arrests) for a white man "was really bad news").

[82] See Hellman (2020), pp. 818, 848 et seq. on why the assumption that it is under US law prohibited for the algorithm to look into protected variables; Mayson (2019), p. 2259, making the point that "differential crime rates do not signify a difference across racial groups in "an individuals' innate" propensity to commit crime" and making an argument for allowing an algorithm to assess risk factors contingent on a protected characteristic; Langenbucher (2022), p. 364 on Art. 10 of the EU AI Act Proposal which allows the processing of data on protected variables if it is "strictly necessary for the purposes of ensuring bias monitoring, detection and correction in relation to high-risk AI systems".

[83] In the EU, not all Member States have credit reporting and credit scoring agencies similar to the US. While Germany and the UK do, France does not and has lenders score applicants in-house.

[84] Bourrell/Fourcade (2021), p. 5.

However, it is important to understand that, from the perspective of the borrower who is subject to scoring, FICO-scoring will often be more transparent than AI and big data scoring. Traditionally, because of the limited number of variables, many reasons for a denial of credit will be obvious. A recent immigrant might, for example, face a risk premium for lack of a repayment history with US credit card companies. Filing a claim under US law will help him to learn about this concern by getting access to some of the factors underlying his score.[85] Understanding his score can guide him towards a solution if he is in a position to influence the relevant factors.[86] With AI models and big data input, this is much less straightforward. Not only will it be difficult for consumers to make sense of computer code as an answer to what influenced his score.[87] Even if he succeeds in understanding distinct variables which drove his score, the model redundantly encodes the same or very similar information in various variables.[88] Put differently: it might not help to bring in additional information, if the AI will extract the identical score by relying on other variables.

III. Inaccurate Data

Algorithmic biases matter even more when combined with concerns about data quality,[89] incorrect labelling,[90] or omitted variables. Data can vary in its reliability across a population, for example if there is less data available for specific groups such as recent immigrants.[91] Additionally, the US CFPB has stressed that the use of alternative data, for instance stemming from social networks, increases the risk of inaccuracies. One of the reasons for that are quality standards. Social networks consist mostly of data uploaded by its users. Platforms do not engage in double-checking the accuracy of that data, much less scrutinize it along the lines of credit reporting agencies. This concerns the applicant for a loan if the data used to evaluate him is inaccurate. It also impacts the AI model which learns from (partially) inaccurate training data. The more inaccuracies are hidden in big datasets, the more the AI's model is shaped by a world which does not even adequately reflect yesterday's world, much less today's.

Market forces should eventually solve some of these issues. This is especially true for models with too many false positive results. The lender will at some point realize that his model does not adequately predict credit default risk and switch to a more powerful one. Of course, the road can be long and there is the risk of credit bubbles and concerns for financial stability to consider.

---

[85] See below E.III. Not all variables are covered because scoring agencies and lenders are allowed to treat the scoring model, including the weight of each variable, as a trade secret. See Langenbucher (2020) comparing this to the (identical) German situation. A ECJ case is pending on this matter.
[86] This is not to deny the many flaws of traditional scoring systems nor to claim that changing the input variables which inform traditional scoring agencies is an avenue open to all communities. This paper does not contribute to the extensive debate on the discriminatory potential of current scoring metrics, see: Bourrell/Fourcade (2021); Burk (2021), pp. 1163 et seq.; Citron/Pasquale (2014), pp. 11 et seq.
[87] Information asymmetries between highly trained coders and consumers facing a denial could theoretically be overcome if a market for intermediaries develops, see Citron/Pasquale (2014), p. 29 for "creative customer relations" demystifying credit scoring via feedback and control mechanisms.
[88] "Flexibility" in the terminology of economists.
[89] See below E.I.
[90] Von Ungern-Sternberg in Mangold/Payandeh (2022), note 16 (sampling bias), note 17 (labelling bias), note 18 (feature selection bias).
[91] See Mayson (2019) for the same problem leading to racial distortion.

For models with too many false negative results, market forces will be less efficient in weeding them out. Some of this has to do with a specific feature of credit decisions, namely that there is no counterfactual data.[92] If the lender denies a loan to an applicant who is considered high risk, he will never know whether the better decision would have been to grant that loan. Accordingly, the AI will not learn which borrowers it denied a loan although it should have offered them one.[93] This is different in other use cases of AI. Imagine a doctor, using AI to scan melanoma. The AI might wrongfully overlook a critical result in year one. If the same patient returns in year two and three, the AI might eventually put his melanoma in the correct category. In this way, it expands its data and learns that it should have categorized the melanoma differently in year one. For credit decisions, this counterfactual data is not available, hence, the AI cannot learn from the false negative decision.

D. Anti-Discrimination Law and AI Credit Underwriting

Understanding algorithmic discrimination in credit scoring and underwriting does not start with a clean slate. Relevant court cases on anti-discrimination concern areas as diverse as employment, housing, bail, mortgages, scoring and credit underwriting.[94] Legal rules and principles address the employer's decision on whom to hire, the landlord's criteria for selecting tenants, the judge making bail decisions, and the lender denying credit. While there are common core principles of anti-discrimination law, each of these scenarios has in the past called for a different balance between competing interests and values. This is perhaps most pronounced for credit scoring and underwriting, given that distinguishing different sets of loan applicants based on sophisticated statistical risk models has been an institutionalized part of making a credit decision. Additionally, macro concerns of financial stability and protecting borrowers from over-indebtedness have for decades fueled the search for powerful predictors of credit default risk.

The US has since the late 1960s regulated fair lending. The EU has only recently started to harmonize relevant laws, partly motivated by the fear that current law does not adequately capture risks of discriminatory practices in AI-based lending.[95] Before, there had been EU rules prohibiting discrimination based on race, ethnic origin, and sex, but these require the loan to be a product or service available to the general public irrespective of the borrower's personal situation. Beyond that, fair lending had remained a question of EU Member State law.[96] The EU Consumer Credit

---

[92] See above C. before I.
[93] Kim (2022), p. 5. A problem which raises similar concerns has been described for predicting future crime, see Mayson (2019), p. 2252. "criminal justice risk-assessment tools purport to predict future crime. But that is not actually what they predict. They generally predict future arrest".
[94] For an overview see Kissinger et al. (2021).
[95] See above A.
[96] See Langenbucher (2020). In Germany, §§ 19, 20 *Allgemeines Gleichbehandlungsgesetz* (General Act on Equal Treatment) have transposed EU Directives.

Directive/2021 introduces an explicit anti-discriminatory rule.[97] It protects consumers legally resident in the EU from discrimination on grounds of nationality, place of residence, sex, race, color, ethnic or social origin, genetic features, language, religion, belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation.[98] The European Parliament's final report proposes an amendment to include the right to be forgotten as from 10 years after the end of their treatment (five years if the diagnosis was made before the age of 18). Further, it specifically requests that Member States should ensure equal access to all people which are cured of relevant communicable and non-communicable diseases.[99]

I. Building Blocks of Underwriting Decisions

Anti-discrimination laws do not distinguish between taste-based and statistical discrimination, nor do they allow for statistical discrimination on grounds of a protected attribute.[100] Some of the strategies an economist views as efficient when faced with uncertainty will be impermissible under the law. Put differently: An efficient underwriting decision can be unlawful even if it reflects an unequal distribution in the world.

1. The Concept of Building Blocks

Anti-discrimination law is triggered by input to a decision-making process. It starts from the assumption that such input comes in the form of different attributes or practices.[101] I refer to them as "building blocks" of a decision. There are outright prohibited building blocks, and facially neutral ones. Impermissible building blocks must not be used, even if they are of direct empirical relevance.[102] At first glance, facially neutral ones can be used. However, in some situations even facially neutral attributes might still be "suspicious", as it were. This is the case if they are "fair in form, but discriminatory in operation".[103] A paradigm case is the US Supreme Court decision in *Griggs,* where the score in an intelligence test was decisive for the position as a manual laborer. While the employer did not explicitly look to age, sex or race, the test triggered a discriminatory result. The same was true for the facts of *Smith,*[104] where years of experience in a job lead to a proportionately lower pay raise. Both, the score in an IQ test and job experience are facially neutral variables. However, they can operate in a way which masks the true reasons underlying the decision. Such is

[97] Art. 6. The EP amendment to the Commission proposal includes: "refusal to provide services in a Member State where the creditor or, where applicable, the credit intermediary or the provider of crowdfunding credit services does not conduct business shall not be considered discrimination", see European Parliament, Report – A9-0212/2022, available at https://www.europarl.europa.eu/doceo/document/A-9-2022-0212_EN.html, amendment 22.
[98] Art. 6 of the proposal, see recital (45) of the current Consumer Credit Directive 2008/48/EC which mentions fundamental rights, among those "the right to non-discrimination".
[99] European Parliament, Report – A9-0212/2022, available at https://www.europarl.europa.eu/doceo/document/A-9-2022-0212_EN.html, amendment 22, 142.
[100] See above C.
[101] On input see Berman/Krishnamurthi (2021), p. 99; Koppelman (2022), p. 10 (the latter criticizing the former, but in agreement about this basic point).
[102] See Gillis (2022), pp. 49, 51 for the claim that this blurs the distinction between anti-discrimination law and affirmative action.
[103] Griggs v. Duke Power Co., 401 US 424 (1971) at p. 431.
[104] Smith v. City of Jackson, 544 U.S. 228 (2005).

the case if the seemingly neutral attribute is picked *because* it correlates, for instance, with age or with race.[105]

This paper puts a spotlight on the role of such building blocks. Metaphorically speaking, traditional doctrine requires these building blocks to form a chain which eventually leads to a decision. The law's role is to carefully examine each building block and to determine whether the decision would have looked different, if one unlawful building block was removed from the chain: Would the person have been hired if the employer had not known her sex? Would the landlord have signed the lease if he had been ignorant as to the tenant's race? But also: Would the same employees have been hired if the IQ test was not run? Would the same bonus payments have been made, if years of job experience were not considered?

With improving technology this core notion of anti-discrimination law faces a novel challenge. Economists such as Fuster et al. (2022) predict that we will soon reach a situation which makes prohibiting the use of specific variables ineffective.[106] Big data furnishes a universe of different variables. Machine learning algorithms unearth innumerable correlations between those variables. Depending on the type of machine learning model employed, its human user might often not be able to identify salient variables or core correlations. Instead, we are looking at an opaque bundle of building blocks which drive a decision. As I will argue below, it is difficult to square this with the conceptual underpinning of anti-discrimination law, based on distinct building blocks, which form a neat chain of causation leading up to a final decision.

US and EU anti-discrimination laws both single out specific sets of building blocks, for instance sex, race, or religion, which they consider directly discriminatory. Both sets of anti-discrimination laws react by prohibiting decisions "because of" [107] or "based on"[108] a protected attribute.

However, reasons, intentions and motives are not always a straightforward phenomenon. Attributes which lead to the denial of a loan may, for instance, reflect a discriminatory taste-based preference but at the same time provide a useful signal for statistical discrimination.[109] This is especially likely if many members of a population have shared the same taste-based preference about, for instance, sex or race in the past. Shared preferences of this type might have triggered credit past rejections. This can today result in clusters of groups with high credit default risk.[110] This narrative might explain why lenders could feel they do not discriminate *because of* any bias on their side but *because of* the efficiency arguments underlying statistical discrimination. As we will see in more detail below, an argument along those lines will not usually exclude liability of the lender. Anti-discrimination laws do not require a protected attribute to be the sole building block. Rather, decisions where only one

[105] This is not to understand the law as blind to economic explanations of discrimination as a statistical phenomenon. Some of the reasons leading to statistical discrimination can show at a later stage of the reasoning process, namely where disparate impact doctrine allows for a legitimate business defense. See below D.III. and see Gillis (2022), p. 48.
[106] Fuster et al. (2022), p. 8; along similar lines: Gillis (2022).
[107] See the ECOA and the FHA for the lending context.
[108] Art. 2 para. 1 EU Directive 2000/43/EC.
[109] CFPB (2017), p. 19; Dobbie et al. (2019) p. 1; Arrow (1971); Phelps (1972).
[110] Sunstein (2019), p. 509: algorithms using factors which are "an outgrowth of discrimination"; Gillis (2022), p. 18: "biased world".

out of a variety of variables is impermissible are considered unlawful, if the variable is one among various building blocks. [111]

Building blocks which are not outright prohibited might still be "suspicious", even if they are facially neutral. Dealing with these is where US disparate impact doctrine and EU indirect discrimination come into play. Disparate impact doctrine deals with decisions which are motivated by a facially neutral attribute, but which still trigger a disproportionately disadvantageous outcome for protected communities. Using the attribute is impermissible unless the decision-maker can establish justificatory reasons such as a business defense. Similarly, indirect discrimination may be legitimate if the relevant variable is a necessary tool, in that no equally appropriate and less discriminatory measures can be identified and the disadvantages caused are proportionate to the aims pursued by the discriminator.[112]

2. Introducing three hypothetical lenders

To illustrate a variety of potential building blocks underlying a credit underwriting decisions, I introduce three hypothetical lenders. For the sake of the example, they all claim that their AI underwriting model does not violate anti-discrimination laws. I further assume that the lender's optimization goal[113] is to assess credit default risk.[114] I assume further that the lender is the one developing the model. There are additional scenarios in practice, for instance an AI credit scoring agency furnishing a report to the lender or a Fintech platform screening lenders but partnering with a bank to originate the loan. Such scenarios raise questions of liability of each involved party which are beyond the scope of this paper.

The first hypothetical lender reasons as follows: "Yes, I have trained the AI model to include protected attributes in the observable variables I use to calculate credit risk. However, this is just one of the many observable variables I use. I include it because, statistically, protected traits are a good indicator for credit default risk."

The second lender claims: "I understand denying credit because of a protected characteristic is impermissible. Therefore, I use an input-control procedure which makes sure that no protected characteristic enters my AI model. Instead, I choose other variables which correlate narrowly with protected traits." [115]

The third lender submits: "I also use an input-control procedure. Furthermore, I don't select any input attributes, my model does. It is a black-box model based on a huge number of variables. I know that the model can find narrow correlations between neutral variables and protected attributes, but as I understand it, this is not prohibited."

---

[111] See below D.II.1.
[112] ECJ ECLI:EU:C:2015:480; ECLI:EU:C:2012:657.
[113] On these see below E.III.2.
[114] Optimization goals such as targeting vulnerable groups for strategic pricing are discussed further below at E.III.2.
[115] See for this strategy in Fintech lending: Di Maggio et al. (2021), p. 4; for the case of Upstart: Langenbucher/Corcoran (2022), p. 143.

Arguably, these hypothetical lenders raise different but related concerns. The first lender's claim has to do with the received understanding of causation and intent if he claims: "the protected attribute was not the sole cause of my denial of credit, it was just one of the variables I use".

The second lender's case is more complicated. Throughout the paper, it has become clear that the combination of big data and AI will lead to proxy variables standing in for protected attributes. Can a lender escape liability if he deliberately chooses a variable which, to his full knowledge, correlates narrowly with a protected attribute?

The third lender is the one we should worry about most. He is aware of the model finding narrow correlations but still uses it.[116] He is what Fuster et al. (2022) might have in mind when they claim that with improving technology it will become ineffective to prohibit the use of certain variables.[117] Arguably, decisions produced by a vast array of variables, not necessarily known to the human who employs the AI model, are hard to square with received anti-discrimination law. The reason for this is that one of the cornerstones of anti-discrimination doctrine has been the concept of distinct building blocks forming a chain of causation towards the underwriting decision. If these become hard to pin down and easily interchangeable without altering the result, this cornerstone loses significance.

II. US Disparate Treatment and EU Direct Discrimination Law on algorithmic credit underwriting

1. Establishing a case

a) Disparate Treatment under US Law

In the US, discrimination in a lending context is addressed by the ECOA and the FHA. Both, the ECOA and the FHA prohibit decisions which are motivated by a protected characteristic, the FHA in the context of mortgages, the ECOA for more general access to credit. While the conceptual framework is straightforward, proving a discriminatory building block is often difficult.

There are generally two regimes available to make a disparate treatment case, both developed in Title VII which covers employment discrimination. The first regime, following *McDonnell Douglas*, is focused on strategies which allow to prove the existence of a discriminatory motive.[118] The plaintiff is sometimes able to establish overt or other direct evidence.[119] However, given the awareness of many employers (or lenders) of anti-discrimination laws, such evidence is often hard to come by. This is even more likely if subconscious motives played a role in decision-making. Against that background, two further strategies allow for inferential proof. Individual inferential proof is common

[116] See Hellman (2020), p 852 on the difference between "because of" and "in spite of".
[117] Fuster et al. (2022), p. 8.
[118] McDonnell Douglas Corp. v. Green, 411 U.S. 792 (1973).
[119] Equivalent in Germany: Sacksofsky (2017), p. 73.

in Title VII cases. A plaintiff will need to establish a prima facie case by showing that she is a member of a protected group, was qualified for a position, was rejected and the position remained open. If she succeeds, the defendant must establish a legitimate non-discriminatory explanation. To do so, he must show that there was no discriminatory motive at play.[120] He does not have to prove that the reason he advances is true. Instead, it is only a burden of production. The plaintiff may react by attempting to prove that these reasons are pretextual. Group or systemic inferential proof is another way to make a disparate treatment case. Plaintiffs use statistics to prove a pattern and practice which reveals that their group is underrepresented.[121] Defendants may bring in different statistics or put forward a legitimate nondiscriminatory explanation to rebut.[122]

The second regime to make a disparate treatment case concerns a mix of factors. It has been applied in situations where it is not in doubt that a discriminatory element contributed to the decision, but a defendant still disputes causation. Courts will remove each building block and check whether the decision would have come out differently. An example is provided by the US Supreme Court decision in *Price Waterhouse v. Hopkins.* The defendant's decision to let a woman go was at stake.[123] She claimed she was fired because of her sex. The defendant had to establish that he had legitimate reasons which, standing alone, would have led to the same decision.

In *Manhart*, the Supreme Court ruled on another defense for using protected attributes. It held that an employer's policy of requiring women to make larger pension fund contributions than men violated Title VII. There was no doubt about an unlawful factor since the policy specifically targeted women. Still, the employer argued that he had no discriminatory intent and did not treat women differently *because of* their sex. Rather, actuarial logic dictated a "life-expectancy adjustment".[124] It is a claim any economist would have embraced, pointing to the logic of statistical discrimination. The US Supreme Court did not follow this reasoning. Instead, it was sufficient to establish that one impermissible attribute was a building block towards the decision. Removing the attribute "female" from the set of variables, so the Court held, would have led to a different, non-discriminatory outcome.[125] In a comparable case, the European Court of Justice rejected the claim of insurance companies which had argued statistics and actuarial logic required an adjustment of fees for women.[126]

b) Direct Discrimination under EU Law

---

[120] Similar in Germany: Sacksofsky (2017), p. 73.
[121] Similar in Germany: Sacksofsky (2017), p. 84 on § 22 *Allgemeines Gleichbehandlungsgesetz* (General Act on Equal Treatment).
[122] See below D.III. for the role of statistics in making a disparate impact case.
[123] Price Waterhouse v. Hopkins, 490 U.S. 228 (1989).
[124] City of Los Angeles v. Manhart, 435 U.S. 702 (1978); the ECJ followed the same logic: ECJ ECLI:EU:C:2008:397; Sacksofsky (2017), p. 73.
[125] The argument differs from the Pricewaterhouse case. In that decision, the sex of the woman was one factor. However, the employer introduced a hypothetical line of causation. He argued that he could have reached the same decision with a different, non-discriminatory motive in mind, see Koppelman (2022), p. 14. In Manhart, the employer argued that his "real" motive was non-discriminatory but simply a short observable variable (sex) to predict an unobservable variable (life-expectancy).
[126] ECJ ECLI:EU:C:2011:100, below at D.II.2.b.

In the EU, several Directives address discrimination between private citizens. Many of these focus on employment, hence, have nothing to say on discrimination in the credit context.[127] Two Directives concern access to publicly available goods or services.[128] However, given the relevance of personal attributes in the credit context, loan contracts typically do not qualify.[129] Additionally, Directive 2004/113 stresses that it does not wish to "prejudice the individual's freedom to choose a contractual partner".[130]

A rule explicitly prohibiting discriminatory lending is envisaged in the EU Consumer Credit Directive/2021.[131] It protects consumers which are legally resident in the EU "on grounds of nationality, place of residence, sex, race, color, ethnic or social origin, genetic features, language, religion, belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation".[132]

Litigating a case is largely subject to each Member State's national law. This includes burden of proof issues. However, EU law requires Member States to adjust their rules on burden of proof to guarantee a minimum standard of enforceability.[133] The plaintiff must make a *prima facie* showing of facts, "from which it may be presumed that there has been direct or indirect discrimination". If he succeeds, burden of proof shifts to the defendant "to prove that there has been no breach of the principle of equal treatment".[134] As part of his defense, he might point out that the plaintiff is not similarly situated to the set of persons he claims is treated more favorably.[135]

2. The first Hypothetical Lender: Mixed Motives and Intent

The first hypothetical lender I described above explicitly used an impermissible attribute. He made two claims to justify its use. First, he suggested that it is but one of the many variables he feeds into his model. Second, he stressed that he had no discriminatory intent but just followed business logic.

---

[127] Directive 2000/78/EC establishing a general framework for equal treatment and occupation; Directive 2006/54/EC on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (recast).

[128] Directive 2000/43/EC implementing the principle of equal treatment between persons irrespective of racial or ethnic origin. The Directive addresses employment, vocational training, social protection, education and access to publicly available goods or services, including housing: Art. 3 para. 1 reads: "access to and supply of goods and services which are available to the public, including housing". Directive 2004/113/EC implementing the principle of equal treatment between men and women in the access to and supply of goods and services addresses access to public goods and services, too, however focuses exclusively on sex discrimination: Art. 3 para. 1 reads: "a service which is available to the public irrespective of the person concerned" and "offered outside the area of private and family life".

[129] For examples see Rödl/Leidinger in Mangold/Payandeh (2022), note 42 et seq.

[130] Art. 3 para. 2.

[131] See above A.

[132] Art. 6 of the proposal, see already recital (45) of the current Consumer Credit Directive 2008/48/EC.

[133] See ECJ ECLI:EU:C:1989:383 pointing to *effet utile* (in an employment context).

[134] Art. 8 EU Directive 2000/43; similar: Art. 9 EU Directive 2004/113. In the context of employment and occupation, Art. 10 EU Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation provides a similar rule. Following up (but only in this context), Art. 19 EU Directive 2006/54 of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation lays down more granular rules to be transposed by the Member States.

[135] ECJ ECLI:EU:C:2015:473; ECJ ECLI:EU:C:2010:703; ECJ ECLI:EU:C:2001:648: discussing "comparable situations", see below at D.III.1.a. for US law on disparate impact turning on two sets of people being similarly situated.

a) Causation and Mixed motives

The first claim has to do with causation. The lender would be right if the law required the protected attribute to be the sole building block towards the decision. Under US law, the Supreme Court in *Price Waterhouse v. Hopkins* rejected this argument, given that the text of the statute did not read "solely because of".[136] Instead, for Title VII cases, established practice requires plaintiffs to prove that the protected attribute was one out of various building blocks of the decision.[137] Events can have multiple but-for causes of this type.[138] Plaintiffs must show that removing the relevant building block changes the outcome.[139] If AI models are used, this entails proof that the lender would have reached a different outcome, had there been an input restriction on the variable sex. For most algorithms, this will be hard to show because many variables correlate with sex. Even if the lender restricts input, due to redundant encoding, the outcome will often remain the same.[140]

If the ECOA and the FHA adopt the standard to prove that a discriminatory variable caused the decision in the way Title VII does, a lender could not escape liability by citing a variety of variables which he included in his decision. So long as a protected attribute was one but-for cause, it is enough to trigger the prohibition. Arguably, the wording of the ECOA and the FHA support this line of reasoning. Section 2000e-2(a)(1) of Title VII stipulates that it is unlawful to discriminate "because of" a protected attribute. It is this term which the *Bostock* Court has invoked to apply what it understands as the but-for standard of causation. Similarly, the FHA speaks of discrimination "because of" protected characteristics and the ECOA makes it unlawful to discriminate "on the basis of" certain protected attributes. [141] None of these statutes require that the outcome was reached "solely" because of the protected attribute. [142] This suggests that both statutes can be read along the same lines as Title VII.

Today, Section 2000e-2(m) of Title VII explicitly allows for a complaining party to demonstrate that the protected attribute was "a motivating factor for any employment practice, even though other factors also motivated the practice". The statute's text was changed to its current wording after the decision in *Price Waterhouse v. Hopkins*. [143] By contrast, Congress did not change the wording of the ECOA and the FHA. Arguably, this cannot be construed as prohibiting a mixed-motive test along the lines of Title VII's mixed-motives test. However, Title VII'S Section 2000e-2(b) includes a follow-up rule. It addresses a situation where the defendant can establish that he would have reached the same decision in the absence of the motivating factor. The rule still allows for declaratory relief,

---

[136] Price Waterhouse v. Hopkins, 490 U.S. 228 (1989).
[137] Berman/Krishnamurthi (2021), pp. 100 et seq.; Eidelson (2022), pp. 797 et seq. on a more narrow reading of the term "because of" as "by reason of"; see Dembroff/Kohler-Hausmann (2022), pp. 74 et seq. for a critique of applying causation standards borrowed from torts to anti-discrimination law.
[138] The EU equivalent for but-for causation is a *conditio sine qua non* test, see Rebhahn/Kietaibl (2010), p. 378.
[139] Bostock v. Clayton County, 590 U.S. (2020), p. 6; Dembroff/Kohler-Hausmann (2022), p. 58.
[140] Von Ungern-Sternberg in Mangold/Payandeh (2022), note 27.
[141] EU law's close analogue reads: "on grounds of".
[142] 15 USC Chapter 41 § 1691; 42 USC § 3604.
[143] See Berman/Krishnamurthi (2021), pp. 99 et seq. on the decision and on Congress following up by amending Title VII to encompass a mix of motivating factors.

injunctive relief and attorney's fees and costs but it limits the relief available (disallowing damages, reinstatement and more). The ECOA and the FHA lack a provision along those lines, hence, it remains an open question whether the limited relief is allowed in ECOA and FHA cases.

Under EU law the text of the relevant Directives is similar to US law. Less favorable treatment is prohibited "because of" a protected attribute, not "solely because of".[144] This includes an intentional reference of a protected characteristic as well as stereotypes or biases.[145]

In litigation, EU law differs from the standard of US law. Once the plaintiff made a *prima facie* showing of facts, rather than establish full proof,[146] he profits from a second rule. He must show that from the facts he established *prima facie* "it may be presumed that there has been direct or indirect discrimination".[147] This is commonly read as the standard of proof for causation being a probability standard. Lowering the burden of proof in this way helps with a disadvantage of EU law: it does not require access to information along the lines of US discovery.[148] Under the probability standard, plaintiffs must establish that it is more likely than not that the building blocks underlying the defendant's decision caused the discriminatory output.[149] In accordance with US law, it is sufficient to show a situation of mixed motives.[150]

b) Discriminatory Motives

The first hypothetical lender does not escape liability by claiming that sex was not the sole cause of his decision. Under both, EU and US law, liability ensues if sex was one of the decision's building blocks. The hypothetical lender's second claim moves beyond causation.[151] He now has discriminatory motives or intent in mind. He points out that the protected attribute correlates with high credit default risk and that this is the only reason why he includes it. He claims that his decision should not be understood as taken *because of* an impermissible motive in the way a taste-based discriminator decides according to his discriminatory preferences.[152] Instead, he screens potential borrowers for sex, race, or similar attributes *because of* an entirely different reason, namely as a

[144] Rödl/Leidinger in Mangold/Payandeh (2022), note 52; Sacksofsky in Mangold/Payandeh (2022), note 43 et seq.
[145] See Sacksofsky (2017); further Mangold/Payandeh in Mangold/Payandeh (2022), note 103; Rödl/Leidinger in Mangold/Payandeh (2022), note 72 et seq.; Sacksofsky in Mangold/Payandeh (2022), note 43 et seq. (distinguishing reference to a protected category ("Anknüpfung") from causation ("Kausalität") understood as the sole factor driving the decision).
[146] Above D.II.1.b.
[147] Art. 10 para. 1 EU Directive 2000/43.
[148] Denying plaintiff rights along those lines: ECJ ECLI:EU:C:2012:217; see Muthorst in Mangold/Payandeh (2022), note 69, 79 et seq.
[149] On German law transposing EU law see: Grünberger (2021); Muthorst in Mangold/Payandeh (2022), note 28, 46, 48; Thüsing (2021), § 22 AGG note 6, 10; defendants can rebut by a *prima facie* showing of facts which suggest that discrimination is not more likely than not, Muthorst in Mangold/Payandeh (2022), note 72; Weigert (2018), p. 1168.
[150] Thüsing (2021), § 22 AGG note 10; Rödl/Leidinger in Mangold/Payandeh (2022), note 52.
[151] By contrast, Dembroff/Kohler-Hausmann (2022), pp. 74 et seq. link the question to causation when they claim that causation in anti-discrimination law requires a "Normative Showing" along the following lines: "If not for the defendant's *discriminatory* conduct, policy, motive, or intent, would the plaintiff have experienced this employment practice or loss?"
[152] See above C for taste-based and statistical discrimination.

predictor for credit default risk. Put differently, he submits that he had no discriminatory intent whatsoever, but just implemented his business strategy.

Neither EU nor US doctrine are open to this line of argument.[153] If the impermissible attribute was one building block of the lender's decision, the law will understand the lender as having discriminated *because of* the protected attribute. As illustrated by the US Supreme Court case *Manhart*,[154] it was sufficient to show that the employer used sex when designing differences in payment plans for men and women.[155] It did not help that he had a business reasons for doing so. The ECJ decided a case involving ethnicity along those same lines in *Feryn*.[156] An employer was looking to recruit fitters but was not ready to employ "immigrants", claiming that his customers were reluctant to give them access to their private homes.[157] The court did not ask the plaintiff to prove any form of discriminatory motivation or intent.[158] I will get back to this point when discussing proxies.[159]

Along those same lines, there is no such thing as a statistical-efficiency defense if a protected trait is used.[160] Actuarial statistics in *Manhart* suggested a risk premium for women as a group, but this did not help the defendant. In *Bostock* the US Supreme Court confirmed its earlier reasoning. It referenced *Manhart* and the inadmissibility of a "life-expectancy adjustment".[161] This suggests that it would be just as futile for a lender to call a policy which discriminates against women a "credit-default risk adjustment", even if looking to sex allowed for efficient statistical discrimination. Arguably, the Court would reach the same conclusion for a disparate treatment case under the ECOA or the FHA. Just like the employer in *Manhart*, a lender might wish to bring in statistics, showing why a risk premium should attach to, for instance, sex, race, or age. He might wish to add that this risk premium motivated his decision, which he understands as non-discriminatory. If the Court decided along the reasoning in *Manhart*, it would not be open to this argument. As soon as an unlawful attribute appears as one of the building blocks leading to the underwriting decision, a disparate treatment case is established, irrespective of any additional discriminatory intent.

---

[153] On EU law: Sacksofsky in Mangold/Payandeh (2022), note 41 et seq.; on German law Rödl/Leidinger in Mangold/Payandeh (2022), note 76, 83.
[154] See above at D.II.1.
[155] Authors seem to disagree about requiring intent. Arguably, some disagreements have to do with imprecisions on what is meant by "intent": (i) discriminatory goals along the lines of taste-based discrimination described above or (ii) deliberately looking to a protected attribute as one motivational building block, but for other reasons, for instance statistical discrimination (as was the case in Manhart). See Mayson (2019), p. 2240 for requiring type (ii) intent. Adams-Prassl et al. (2022), p. 6 understand US, but not EU and UK law as requiring intent type (i); see further Campbell/Smith (2022), p. 3.
[156] ECJ ECLI:EU:C:2008:397.
[157] ECJ ECLI:EU:C:2008:397 note 16.
[158] See explicitly in ECJ ECLI:EU:C:2017:204; a similar question was raised in ECJ ECLI:EU:C:2013:275; Adams-Prassl et al. (2022), p. 6 claim that unintentional discrimination can be direct under European law but not under US law. However, cases such as *Manhart* show that an intention to discriminate is not necessary if the protected attribute is one building block of the decision. For aligning US and EU law as proposed here: Hacker (2018), p. 55; Wachter et al. (2021), p. 41; Zuiderveen Borgesius (2020), p. 31.
[159] See below at D.II.3.b.
[160] See for a comparison to disparate impact's business defense: Barocas/Selbst (2016), p. 713 referencing 42 U.S.C. § 2000e-2(k)(2) (2012).
[161] Bostock v. Clayton County, 590 U.S. (2020), p. 2.

The European Court of Justice, arguably, would concur for discrimination based on sex or race. For these attributes, the Court invalidated EU Directive 2004/113/EC which had initially made room for a business defense, justifying disparate output between men and women for certain insurance premium payments.[162] So far, the Court has not extended this jurisprudence to discrimination on grounds of other protected characteristics.[163] For the consumer credit context, this leaves an open question as to a potential justification of direct discrimination if the interests of the lender or more general macro-economic concerns tied to the stability of credit markets outweigh those of loan applicants.

3. The Second and the Third Hypothetical Lenders: Discriminatory Output when using Neutral Variables – the Exception becomes the Standard Case

The defense put forward by the second hypothetical lender is more intricate. He claims to escape liability because he controls input, making sure the model does not use a protected characteristic for its computations. A court, so this second hypothetical lender submits, which examines the relevant building blocks will not find an unlawful trait among them. This should rule out liability for disparate treatment/indirect discrimination even if the second lender deliberately chose variables which correlate narrowly with protected traits. The third hypothetical lender did not deliberately choose the relevant variables. Still, he understood that the model would likely do that.

a) The first Hard Case: Blurring the Line between Protected and Neutral Characteristics

Courts and scholars in the EU and in the US have so far followed the ground rule that disparate treatment/indirect discrimination liability requires a protected trait to feature as one building block of the decision. This would rule out liability for the second and the third lender as they both use an input control mechanism, making sure no protected trait played a role in the underwriting decision.

However, in both jurisdictions there have been hard cases, blurring this bright-line distinction. They are intended to catch a decisionmaker who hides his true intent or motive behind a seemingly neutral attribute.[164] It follows from there, that discriminatory motives suddenly play a role to detect the attempt at circumventing the rule. The argument goes as follows: If the decisionmaker explicitly uses a protected attribute as a building block for his decision, liability ensues, his motives for doing so are irrelevant.[165] By contrast, if he uses a facially neutral attribute, there is usually no disparate treatment/indirect discrimination liability. The exception to the rule concerns a situation where the decisionmaker deliberately picked the neutral attribute, in full knowledge that it would lead to the same result as using the protected characteristic.

Paradigm examples mirror different historical and cultural paths taken in the US and in the EU. In the US, scholars have pointed to variables which are identical to the protected characteristic and differ

---

[162] ECJ ECLI:EU:C:2011:100.
[163] See von Ungern-Sternberg in Mangold/Payandeh (2022), note 57.
[164] "Covert discrimination", Sacksofsky in Mangold/Payandeh (2022), note 54.
[165] See above D.II.2.b.

in name only or to variables which serve as a pretext for the real motive.[166] Redlining is a classic illustration. It refers to the practice of denying a creditworthy applicant a loan for housing in a certain neighborhood, even though he may otherwise be eligible for the loan. Redlining on a racial basis has been held by courts and regulatory agencies to be an illegal practice[167] entailing disparate treatment liability.[168] The term "race" is not openly used; however, redlining is generally understood to be a good enough proxy to serve as a pretext for race. The lender is treated as if he had used the protected attribute itself.

In the EU, there is no history of redlining,[169] but Europe has its own hard cases. An example is provided by a UK court case concerning sex and age discrimination. The case concerned a town who lowered prices for a public swimming pool based on pensionable age. The term "pensionable age", so the court held, had become a shorthand expression referring to the age of 60 in a woman and to the age of 65 in a man. [170] Where ZIP codes stood in for race in US cases, "pensionable age" worked as a proxy for sex and age and allowed to consider it a pretext.[171] Some scholars have suggested to go further and understand any intentional use of a variable which correlates narrowly with the protected attribute as triggering disparate treatment/indirect discrimination liability.[172] By contrast, most courts and scholars seem open to considering a variable which correlates 100 % with a protected attribute, but refuse to step beyond that line.[173] Instead, scholars have pointed out that the probability standard for causation[174] is the more fitting legal tool to help with burden of proof.[175]

Both legal orders grapple with hard cases of that genre, blurring the bright line between using a protected attribute and a neutral one. For the US, Barocas and Selbst discuss a decisionmaker who "likes" the effects he produces[176] and label such behavior as "masking".[177] "The entire idea of masking", they explain, "is pretextual".[178] The suggestion to carve out cases where the lender deliberately chooses an only seemingly neutral variable, despite being familiar with the

---

166 Kim (2022), p. 15.

167 https://www.federalreserve.gov/boarddocs/supmanual/cch/fair_lend_fhact.pdf with exceptions for redlining to identify an area on a fault line or a flood plain.

168 See https://www.ffiec.gov/PDF/fairlend.pdf, p. iv for the OCC, the FDIC, the Fed Board, the Office of Thrift Supervision and the National Credit Union Administration; similarly Prince/Schwarcz (2020), p. 1257; Campbell/Smith (2022), p. 9 present redlining first as an example for indirect discrimination, however, at p. 14 the authors claim it is direct discrimination if the protected characteristic was considered in deciding.

169 See for an understanding along indirect discrimination Dzida/Groh (2018), p. 1917; German law prohibits redlining (§ 32 Bundesdatenschutzgesetz, § 10 Kreditwesengesetz), but it is unclear whether this is in line with the EU GDPR which does not allow for gold-plating, see von Ungern-Sternberg in Mangold/Payandeh (2022), note 53.

170 James v Eastleigh Borough Council (1990) 2 AC 751.

171 I read this as a hard case which blurs the bright-line distinction, not as supporting the broader view of Adams-Prassl et al. (2022) that EU law's scope of direct discrimination is broader than US law's disparate treatment.

172 Adams-Prassl et al. (2022) (but without a clear new benchmark).

173 See the UK case of Lee v. Ashers Baking Company Ltd and others [2018] UKSC 49 ("exact correspondence test"); hesitant Sacksofsky in Mangold/Payandeh (2022), note 53 et seq.; for different reasons: Campbell/Smith (2022), pp. 10 et seq.

174 Above D.II.2.a.

175 Discussing the hypothetical of a bouncer asking guests queuing to enter a club for a college degree: Rödl/Leidinger in Mangold/Payandeh (2022), note 57 et seq.

176 In an employment context, they suggest that "intent is clear, if the employer continues *because* he liked the discrimination produced", Barocas/Selbst (2016), p. 699.

177 Barocas/Selbst (2016), p. 699; for German law see Rödl/Leidinger in Mangold/Payandeh (2022), note 57 ("cover-up").

178 Barocas/Selbst (2016), p. 699.

discriminatory output it produces has intuitive appeal, arguably because it rings of circumventing legal rules. Following this line of thought, the second lender could not escape liability. He deliberately picks variables which correlate narrowly with the protected attribute, which makes his argument a pretext. At the same time Barocas and Selbst are aware of the slippery slope they face: "Deciding to follow through on a test with discriminatory effect", they reason, "does not suddenly render it disparate *treatment*".[179] This seems to suggest that sheer awareness of a correlation between a neutral variable and the protected attribute alone is not enough to trigger disparate treatment liability. As Hellman explains: "The Supreme Court has insisted that a screening tool must have been adopted "because of" the disparate impact and not merely "in spite of" these foreseeable consequences in order to give rise to strict scrutiny".[180] Against this background, the third hypothetical lender escapes liability. He knows that the model will find narrow correlations but refrains from deliberately training it to do so. In practice, this means that only a disparate impact/indirect discrimination case is available.[181]

Hard cases are neither a novel phenomenon nor necessarily a reason to revisit the starting point of doctrinal distinctions. So far, relevant cases have been sparse, and the debate has profited from the implicit understanding that there is a very limited number of attributes which are identical to the protected trait. Additionally, most hypotheticals concern deciders which deliberately seek out attributes generally understood to be identical to the protected characteristic – such as redlining and race. Imposing liability made sense to close an apparent gap for circumventing the rule. A defendant who could show that other motives drove the decision and he only acted "in spite of" the consequences escaped liability.

With the advent of big data and AI models, these implicit assumptions seem less compelling. Instead of a few rare variables mirroring the protected trait, the number of variables the AI identifies as stand-in proxies will grow immensely. This is true for individual variables an AI finds to correlate with a protected attribute. It is even more evident for bundles of variables which in their combination allow to precisely predict the probability of a person to share a protected trait. [182] Lenders can obtain a very precise picture of applicants for a loan. What is more, blocking a short list of protected input variables will often not lead to a different result. Many other variables redundantly encode the information contained in the protected attributes. In that way, the outlier case of finding a neutral attribute which works as a proxy for a protected characteristic becomes the new standard. One of the cornerstones of anti-discrimination law, the distinction between protected and neutral attributes, will be blurred because a neutral attribute or bundles of these can fulfill the same function as a protected one. Put differently: There will be many more hard cases.

[179] Barocas/Selbst (2016), p. 699.
[180] Hellman (2020), p 852.
[181] For EU law see Hacker (2018) p. 55; Wachter et al. (2021), p. 41; Zuiderveen Borgesius (2020), p. 31; for a broader understanding under EU law see Adams-Prassl et al. (2022); on terminology Sacksofsky in Mangold/Payandeh (2022), note 12.
[182] To illustrate, consider marketing researchers who have found numerous strategies to predict age or sex, based on online behavior, mobile phone services, natural language processing or twitter usage, see e.g. Al-Zuabi et al. (2019).

At first glance, it would be a natural consequence for the distinction between acting "because of" a negative impact on protected communities or "in spite of" that impact, to fill the doctrinal gap. If decisions cannot be adequately judged looking to protected or neutral building blocks, motives and intent might work as benchmarks. However, reasoning along those lines would be hard to square with the principles traditionally underlying US and EU anti-discrimination law. The core tenet of anti-discrimination law would then revolve around the question whether a defendant acted "because of" the negative impact on protected groups. This stands in stark contrast with received doctrine which has for good reason opposed requiring intent or specific motives when establishing an anti-discrimination case.[183]

Asking plaintiffs which are faced with algorithmic decision-making to prove intent is not only, doctrinally speaking, a step backwards. Additionally, it will often be complicated to pin down the relevant decision-maker and to prove intent. Algorithmic models are about outsourcing the identification of a specific attribute. In many ways, "the model" is the decision-maker. However, as long as there are no legal rules tailored to this situation, the plaintiff is left with going back to the human decision-makers involved in making a lending decision. Traditionally, plaintiffs investigated the decision to pick a facially neutral variable. Instead, they must now explore the state of mind of the lender who realizes the discriminatory potential of his model. If the plaintiff is lucky, he faces the second hypothetical lender who deliberately chooses a stand-in proxy. More realistically, the plaintiff will confront a situation I described as the third hypothetical lender. This lender will be both, the most frequent and the most troubling case. His motivation is business profit. This lender points his fingers at "the model" and claims to just go along with whatever it suggests. The Fintech lender Upstart provides a good illustration for a lender who, arguably, was aware of a troubling correlation and still wished to escape rather than cement discrimination.[184] Upstart's underwriting model uses various variables. Educational background is one of the variables used and Upstart claims that this allows to grant more loans to minority groups than lenders working only with traditional FICO scores. This inspired the mystery shopping exercise I described above.[185] The authors found that under Upstart's model borrowers with an educational background in a historically Black or a Hispanic/Latinx institution paid a significant penalty on both, origination fee and interest rates, if compared with borrowers who attended NYU. While Upstart did not use race as a variable, it would be surprising for a US lender to be unaware of a correlation between race and education.[186] Still, there was no indication of discriminatory intent on the side of Upstart. Rather, including education (among other variables) contributed to a more granular prediction of credit default risk, to finding more "invisible primes" and to (in the aggregate) originating more loans.[187]

Should a lender of this genre choose to run output tests, he might become aware of inequality across applicants. However, awareness alone is what the US Supreme Court addresses as deciding "in spite of" the impact on a protected characteristic. The fact alone that he is aware of an asymmetric

---

[183] See above D.II.2.b.
[184] Langenbucher/Corcoran (2022), p. 152.
[185] See above B.
[186] This assumes that Upstart was not hiding its true motivational element of discriminating on the grounds of race. If this were the case, we would be faced with a masking situation as explained in the preceding section.
[187] CFPB (2017), p. 6.

distribution of an AI model's output does not entail a disparate treatment case.[188] Additionally, even if we were prepared to accept awareness of inequality across groups alone as triggering disparate treatment/direct discrimination liability, this would bring about a somewhat paradoxical result, especially in the case of scoring credit default risk.[189] The better and more granular the model, the more likely it would be considered illegal.

b) The second Hard Case: A Facially Neutral Attributes implies a Protected Characteristic

Algorithmic decision-making based on big data is likely to shed an unwelcome light on another doctrinally hard case. It revolves around attributes which not only correlate with a protected trait but are understood to be somehow implied by it.[190] Pregnancy is a classic example.[191] Neither US nor EU law had originally listed the term "pregnancy" as a protected characteristic. A textualist reading would expect courts to address the issue as one of proxies, facially neutral variables which correlate with a protected characteristic. This is indeed what US courts did in the 1970s. In *Gilbert*, the US Supreme Court held that exclusion of pregnancy from a disability benefits payments plan was not based on sex.[192] Congress had to amend Title VII to extend its protection to pregnancy, closing an apparent gap. By contrast, the European Court of Justice, following more purposive principles of interpretation, found that pregnancy is "inextricably linked" to the female sex. A refusal to employ an applicant due to pregnancy, so the Court reasoned, can only concern women.[193] Rather than have plaintiffs wait for a legislative amendment, the Court proceeded with a broad reading of the term sex. Pregnancy was addressed as an attribute inextricably linked to sex. Half a century after *Gilbert*, the US decision in *Bostock* arrived at a similar result. The court applied Title VII's prohibition of discrimination on grounds of sex to discrimination because of sexual orientation. Claiming that this followed from a textualist interpretation of the term sex,[194] the Court argued that it is impossible to discriminate against homosexual or transgender persons without first ascertaining their sex. Decisions based on homosexuality or transgender identity were held to "necessarily entail" sex.[195]

Once again, the traditional debate profits from the fact that one attribute implying another one is a rare occurrence. Relatively few court cases speak about neutral variables which "necessarily entail" a protected characteristic. Consequently, the conceptual fuzziness these cases introduce to the

[188] See above D.II.1; Adams-Prassl et al. (2022), p. 8 claim that this is a "startling conclusion". I disagree. Credit underwriting presents a good example: Lenders will often be aware of differences in credit default risk across protected groups. This can be a result of statistical discrimination, if protected groups, statistically, present a higher credit default risk see above B. Understanding this fact alone as troubling conflates statistical and taste-based discrimination.
[189] But see above C for different a definition of success, such as allowing to find especially vulnerable borrowers.
[190] See Adams-Prassl et al. (2022), p. 12: "inherently discriminatory"; Krishnamurti/Salib (2020): "Conceptual Causation"; referring to the latter: Berman/Krishnamurthi (2021), pp. 88 et seq.; discussing "being a mother" as a "true subset of one sex" on p. 105; Sacksofsky in Mangold/Payandeh (2022), note 58 et seq.
[191] For the sake of this example, I do not address the situation of transitioning persons where a man might become pregnant, see Sacksofsky (2017).
[192] General Electric Co. v. Gilbert, 429 U.S.125 (1976), p. 149.
[193] ECJ ECLI:EU:C:1990:383 (note 2).
[194] The debate whether the Court's result in Bostock can indeed be explained under a textualist approach is outside the scope of this paper; for an overview see Berman/Krishnamurthi (2021).
[195] See also United States v. Sineneng-Smith, 590 U.S. (2020), p. 19.

separation between prohibited and neutral attributes has not yet led scholars to revisit the starting point of the distinction.[196]

Mirroring the first hard case, big data and AI alter these implicit assumptions. We will face many novel proxies: single variables or bundles of those which allow to predict protected characteristics with a very high probability.[197] Often, the AI will find unexpected correlations with enormous predictive power. If these predictions are as good as the predictive power of pregnancy for sex, can we then say the variable the AI found "implies", "necessarily entails" or is "inextricably linked" to a protected attribute? If a specific bundle of variables allows to predict a protected characteristic, does the bundle in its entirety "necessarily entail" a protected attribute? To be sure, algorithmic decision-making with big data multiplies the fuzziness which variables implying a protected attribute have brought about. So far, narrow precedents on pregnancy and sex identity provide rough guidance at best. What is more, the rationale underlying these cases has little to do with the unanticipated correlations AI unearths. Instead, it is evidence of another poor fit of received anti-discrimination law for algorithmic decision-making.

III. US Disparate Impact and EU Indirect Discrimination Law on Algorithmic Credit Underwriting

The doctrinal challenges a disparate treatment/direct discrimination case faces for algorithmic decision-making explain why current debates instead center on disparate impact/indirect discrimination.[198] This doctrine allows for a more intense normative control of contracting choices. Under both, US and EU law, disparate impact/indirect discrimination deals with facially neutral variables or practices. As a starting point, it is worth remembering that it is (of course) not prohibited *per se* to use neutral variables or practices as building blocks for a decision. However, if a specific neutral characteristic or practice consistently triggers less favorable treatment of protected communities, this makes it "suspicious", as it were. One explanation might be, that we are looking at a variation of the first hard case discussed above.[199] A decision-maker might have found a seemingly neutral attribute or practice to hide his true discriminatory motivations. In the words of the US Supreme Court, the doctrine then works as "an evidentiary tool used to identify genuine, intentional discrimination – to 'smoke out,' as it were, disparate treatment".[200] Similarly, in an older decision the ECJ held that "rules based on other criteria such as residence abroad, language, place of birth, descent

---

[196] But see the intense debate on what constitutes the counterfactual when discussing sex in the wake of *Bostock*, Berman/Krishnamurthi (2021); Dembroff/Kohler-Hausmann (2022), pp. 60 et seq.; Eidelson (2022), pp. 788, 794 et seq.; Koppelman (2022).
[197] See Adams-Prassl et al. (2022), pp. 14 et seq. for what they call the "perfect proxy".
[198] But see Adams-Prassl et al. (2022) with the attempt to reorient the discussion towards disparate treatment/ indirect discrimination.
[199] See above D.II.3.a.
[200] Ricci v. DeStefano, 557 U.S. 557 (2009), discussed at Gillis (2022), p. 25 ("intent-based"); overview at Langenbucher (2020), p. 554; on EU law's trajectory from a formal to a more substantive approach of indirect discrimination doctrine see Rebhahn/Kietaibl (2010), p. 384 et seq.; further: Sacksofsky in Mangold/Payandeh (2022), note 105; Ungern-Sternberg in Mangold/Payandeh (2022), note 91; for EU law see Mangold/Payandeh in Mangold/Payandeh (2022), note 109, listing the prohibition to circumvent anti-discrimination law as well as shifting the burden of proof.

or performance of military service in the country may in fact conceal discrimination on the basis of nationality".[201]

Today, disparate impact/indirect discrimination principles are overwhelmingly understood as going beyond a mechanism to discover covert discrimination, at least if a constitutional law context is at stake.[202] When faced with the government discriminating against a private citizen, most courts and scholars follow some version of a substantive, anti-subordination approach. They are interested in the effects and "the consequences of [...] practices, not simply the motivation".[203] It follows from this understanding that plaintiffs do not have to prove discriminatory intent or motives.[204] However, if private parties are litigating, this move from formal to substantive anti-discrimination theories is considerably less pronounced. Most start from the ground rule that private parties enjoy free contracting choices.[205] Faced with disparate impact on protected groups, the focus of the discussion shifts to whether there are other, equally useful but less discriminatory strategies for the defendant to achieve his goal. If this cannot be established, US and EU law allow for variations of a business necessity defense to justify the practice despite the disproportionate output. These are the core differences if compared with disparate treatment/direct discrimination, where no justificatory reason regime applies.[206]

1. US Disparate Impact Doctrine and Credit Underwriting

There is no comprehensive US Supreme Court guidance as to whether disparate impact doctrine extends to access to credit. *Ricci* seemed to limit the doctrine,[207] but in *Inclusive Communities* the Court held that "disparate impact claims are cognizable under the Fair Housing Act (…) when their text refers to the consequences of the actions".[208] The ECOA lacks a results-oriented language of this type. Still, the CFPB and some courts seem open to applying disparate impact in that area.[209]

2. EU Indirect Discrimination Doctrine and Credit Underwriting

---

[201] ECJ ECLI:EU:C: 1974:13; on this point see Rebhahn/Kietaibl (2010), p. 385.

[202] Mangold/Payandeh in Mangold/Payandeh (2022), note 110; Sacksofsky (2017); Sacksofsky in Mangold/Payandeh (2022), note 20 et seq., 107 et seq., stressing the link between equality and dignity/human rights; broad understanding even for private law at Grüneberg (2013) p. 1004 et seq.

[203] Griggs v. Duke Power Co., 401 US 424 (1971), p. 432; Smith v. City of Jackson, 544 U.S. 228 (2005), p. 236; Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc., 135 S. Ct. 2507 (2015), p. 2522; discussed at Gillis (2022), p. 26 ("effect-based"); see Mayson (2019), p. 2241 for understanding any output-control as aligning with anti-subordination; Sunstein (2019), p. 506; on EU law Rebhahn/Kietaibl (2010), p. 389.

[204] Rebhahn/Kietaibl (2010), p. 389; Sacksofsky in Mangold/Payandeh (2022), note 108.

[205] Below D.III.1. As an illustration see Art. 3 para. 2 EU Directive 2004/113: "This Directive does not prejudice the individual's freedom to choose a contractual partner as long as an individual's choice of contractual partner is not based on that person's sex". For European and German law see the different approach at Grüneberg (2013) p. 1004 et seq., claiming that a more general equal protection principles applies between private citizens. He moderates this claim by introducing justificatory reasons; a short version of his argument is to be found at Grünberger (2017), p. 16 et seq.

[206] See, however, Sacksofsky in Mangold/Payandeh (2022), note 96 et seq, explaining that distinctions can still be legitimate if (i) exceptions to the rule apply or (ii) the relevant persons are not similarly situated.

[207] Ricci v. DeStefano, 557 U.S. 557 (2009).

[208] Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc., 135 S. Ct. 2507 (2015), p. 2519.

[209] Ramirez v. Greenpoint Mortgage Funding, Inc., 633 E. Supp. 2d 922 (2008), pp. 926–927; Gillis (2022), p. 23.

All EU anti-discrimination Directives have so far explicitly covered both, direct and indirect discrimination,[210] however, their impact on a loan context is limited.[211] By contrast, the reformed EU Consumer Credit Directive/2021, the first to specifically prohibit discrimination in the "conditions to be fulfilled for being granted a credit",[212] does not explicitly say whether it covers indirect discrimination.

While the earlier Directives included a definition of direct and indirect discrimination, the Consumer Credit Directive/2021 does not. Arguably, this does not warrant a departure from established practice. Had the EU legislator wanted to limit the new rule to cases of direct discrimination, one might have expected a clear indication. Recitals (25) and (26) provide an additional, albeit vague argument for this reading of the Directive. They reference Art. 21 of the Charter of Fundamental Rights which some understand as extending to indirect discrimination.[213] However, even under this understanding it is important to bear in mind that the Charter explicitly addresses only the institutions of the Union and its Member States when they are implementing Union law, Art. 51 para. 1 s. 1 of the Charter.[214] The text of the Charter is silent as to bringing it to bear on private law relations between citizens.[215] However, the ECJ decisions in *Egenberger*,[216] *Max-Planck-Gesellschaft*[217] and *Cresco Investigation*[218] hint at implications along those lines[219] and in employment law cases, Art. 157 TFEU has since the 1970s been understood to prohibit sex discrimination between private parties.[220] Still, the extent to which fundamental human rights impact dealings between private citizens is today not fully established.[221] Even assuming an interpretation along those lines, the fundamental rights of the discriminated community must be balanced against the fundamental rights of the discriminator.[222] In a credit underwriting context this covers the lender's right to make an unencumbered contracting choice. Additionally, there are macro-concerns of financial stability and responsible lending to consider.[223]

---

[210] See Sacksofsky in Mangold/Payandeh (2022), note 5 distinguishing this EU secondary law strategy (explicit mentioning of indirect discrimination) from human rights texts which list protected categories and proscribe discrimination "on grounds of" these categories. The US ECOA (albeit statutory law) falls in that second category.
[211] See Directive 2000/43 and Directive 2003/124 and above at A.
[212] Art. 6.
[213] Jarass (2021), Art. 21 note 33 (however, referencing the ECJ in the context of existing Directives where indirect discrimination is explicitly covered).
[214] Gärditz in Mangold/Payandeh (2022), note 66.
[215] The same is true for Human Rights under the German Basic Law, see Gärditz in Mangold/Payandeh (2022), note 2 et seq.; however, German law accepts indirect effects of Human rights between private citizens. Some explain this as a requirement for courts to respect the Basic Law when they interpret open-ended legal terms, others understand this as a more general obligation of the State to protect its citizens, see *ibid.* note 21 et seq.; sex, ethnicity, race, language, homeland and origin, faith, religious or political opinion under Art. 3 para. 3 Basic Law are widely considered as holding not only the government, but also private citizens to anti-discriminatory standards, see *ibid* note 58.
[216] ECLI:EU:C:2018:257.
[217] ECLI:EU:C:2018:874.
[218] ECLI:EU:C:2019:43.
[219] Overview at Mörsdorf (2019).
[220] ECJ ECLI:EU:C:1978:130; ECLI:EU:C:1996:170; ECLI:EU:C:2005:709; ECLI:EU:C:2010:21. Overview at Schrader/Donath in Langenbucher (2022a), § 7 note 35.
[221] For a short introduction see Langenbucher/Donath in Langenbucher (2022a), § 1 note 53; Schrader/Donath in Langenbucher (2022a), § 7 note 125; further see Gärditz in Mangold/Payandeh (2022), note 67 et seq.
[222] See Jarass (2021), Art. 21 note 4, 18, 39; Gärditz in Mangold/Payandeh (2022), note 35, justifying anti-discrimination law (among other things) as prohibiting contractual choices which perpetuate illegitimiate stereotypes and biases; overview at Grünberger (2017), p. 16 et seq.
[223] See below E. III., IV.

3. Facial Neutral Building Blocks and Unequal Output across Groups

Under the assumption that disparate impact/indirect discrimination doctrine does apply to credit underwriting, a first test prong US[224] and EU law[225] look to are the decision's building block. It is useful to note that this should not be misunderstood as a traditional causation requirement to explain why the facially neutral attribute brought about the decision. Quite to the contrary, it is one of the characteristics of disparate impact/indirect discrimination that the neutral variable *correlates* with a protected attribute but does not cause the disparate outcome.[226] Still, what this test prong achieves is to exclude a showing of statistical disparity only.[227] Put differently: the plaintiff must identify a distinct neutral but suspicious variable or practice as one of the building blocks towards the discriminatory output.[228]

Paradigm examples for a suspicious variable used in EU employment law cases concern different working conditions for employees. If they distinguish between full-time and part-time work, there can be a *prima facie* case established if most part-time workers are female.[229] The plaintiff must identify the variable "part-time work". 'She does not have to show that every part-time worker is female. However, it is not enough to provide statistics on how female employees are treated differently than their male counterparts without identifying a variable or practice, such as part-time work, which drove the decision.

The US Supreme Court decision on Title VII in *Griggs* illustrates how a specific practice caused disparate impact across racial groups. The employer had used the score in an intelligence test as decisive for the position as a manual laborer, thereby discriminating against Black employees. Title VII prohibits employment tests which are not a reasonable measure of job performance. Justice Burger stressed that "the Act does not command that any person be hired simply because he was formerly the subject of discrimination, or because he is a member of a minority group". Bringing in statistics to show a significant underrepresentation of Black employees, without identifying the intelligence test used, would not have been a successful strategy. However, the Justice continued,

---

[224] The US Supreme Court has at times stressed the "consequences of actions". This is not to be understood as establishing a purely output-oriented control, doing away with the causation requirement. Rather, when speaking of "the consequences of actions" the Court was concerned with delineating disparate impact doctrine from "the mindset of actors", see Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc., 135 S. Ct. 2507 (2015), p. 10. See further below at D.III.3.a. on Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc., 135 S. Ct. 2507 (2015) not allowing liability imposed solely based on a showing of a statistical disparity, id. p. 18 and stressing the importance of a „robust causality requirement" id., p. 20.
[225] See ECJ ECLI:EU:C:1993:859 on burden of proof after a plaintiff has established statistical proof; see also ECJ ECLI:EU:C:2019:828 (note 56); ECLI:EU:C:2013:122 (note 42 et seq.); Mangold/Payandeh in Mangold/Payandeh (2022), note 107.
[226] Rebhahn/Kietaibl (2010), p. 390.
[227] Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc., 135 S. Ct. 2507 (2015), p. 2523; for an EU version of this argument see Grünberger (2013), p. 660 et seq.; Mangold/Payandeh in Mangold/Payandeh (2022), note 110; von Ungern-Sternberg in Mangold/Payandeh (2022), note 91.
[228] Gillis (2022), p. 27; Rebhahn/Kietaibl (2010), p. 390; von Ungern-Sternberg in Mangold/Payandeh (2022), note 92 ("relevante Wirkvariable").
[229] See ECJ ECLI:EU:C:1989:383 on bonus payments in an employment context; ECJ ECLI:EU:C:1993:859; ECJ ECLI:EU:C:1995:155; Muthorst in Mangold/Payandeh (2022), note 31; Sacksofsky in Mangold/Payandeh (2022), note 31.

"the Act proscribes not only overt discrimination, but also practices that are fair in form, but discriminatory in operation."[230] Pointing to the disparate outcome alone was not enough for the plaintiff's case. Instead, he had to identify the intelligence test as the suspicious practice.

For the plaintiff, this requirement often turns on the relevant standard of proof. In the US, a disparate impact case requires a *prima facie* showing of a disparate outcome for a protected group. Cases can be made under the burden-shifting framework developed in Title VII cases, often by bringing in statistical evidence. For an outcome to be disparate, a relevant set of persons must be identified and the outcome for these persons must be compared to the rest of the relevant sample. If the protected group is faced with a less favorable outcome, a *prima facie* case is established.[231] Along similar lines, EU Directive 2000/43 – while referring the matter back to the Member States – mentions that indirect discrimination may be established "by any means including on the basis of statistical evidence".[232] Member States law differ on who must prove facts which make a disparate output across groups more likely than not[233] and on the burden of proof to establish a justificatory reason.[234]

4. Justificatory Reasons

Disparate impact/indirect discrimination doctrine is triggered by neutral, yet suspicious characteristics or practices. The prima facie case the plaintiff establishes to show disparate output makes the characteristic or practice suspicious. Defendant must demonstrate that there was a justification which explains the disproportionate outcome across groups. Defendants can deny that there was a disproportionate outcome by discussing group membership. Assume, for example, that a lender's practice results in denying loans to 70% of female and 20% of male applicants. Given that (roughly) 50 % of the population are female, this looks like a disproportionate outcome across the sexes. However, the lender might claim that in the credit underwriting context, group membership cannot be limited to sex alone. Instead, he might suggest that only similarly situated sets of applicants ought to be compared. To decide which set is similarly situated to another set, he could propose to look at variables such as net worth, income, or credit history, all of which influence credit default risk. The effect might not be disproportionate if, for similarly situated sets of loan applicants, no sex discrimination shows. It is obvious that many cases will turn on building and comparing such sets of loan applicants. The narrower the group which serves as benchmark for a disparate impact comparison, the more difficult for a plaintiff to establish a case.

[230] Griggs v. Duke Power Co., 401 US 424 (1971), p. 430 et seq.
[231] See Kim (2022), p. 17 on difficulties in practice to collect data about outcomes across the applicant pool.
[232] Recital (15); see Sacksofsky in Mangold/Payandeh (2022), note 126 et seq.; in an employment context see ECJ ECLI:EU:C:2019:828 (note 56); for German law transposing EU law see Hoffmann (2014); Thüsing (2021), § 22 AGG note 17.
[233] Accepting general data on labour markets but stressing the discretion of national courts: ECJ ECLI:EU:C:2019:828 (note 56); see Muthorst in Mangold/Payandeh (2022), note 57, 64; critical as to statistical data: Thüsing (2021), § 22 AGG note 17.
[234] See above D.II.1.b. on burden of proof, further: ECJ ECLI:EU:C:1993:859 on burden of proof after a plaintiff has established statistical proof; see also ECJ ECLI:EU:C:2019:828 (note 56); ECLI:EU:C:2013:122 (note 42 et seq.); on justificatory reasons: Reimer in Mangold/Payandeh (2022); *ibid* note 70 et seq. on probabilistic assessments as a justificatory reason.

More defense strategies are available if defendants can establish why their legitimate outweigh the interests of the members of the protected group. How to do this varies significantly across context. A case of employment discrimination calls for different reasons than a housing case, and a credit underwriting case calls for still different justificatory reasons. In a loan context, US law asks the lender to demonstrate a business necessity for treating the members of the protected group less favorably.[235] More granularly, EU law first asks for a legitimate aim and a legitimate tool, to then focus on the variables or practices employed by the discriminator. These must pass a test for being fit for purpose, necessary, and overall proportionate.[236]

The most natural aim for a lender to pursue is assessing credit default risk. Statistical evaluation and scoring procedures have developed as a legitimate tool over the last century.[237] Put differently: the statistical discrimination these produce does not automatically translate into discrimination under the law.[238] Instead, the lender must provide justificatory reasons.[239] The burden then shifts back to the plaintiff for another strategy: He might show that there was a less discriminatory way to achieve that same goal. He will succeed if there is an alternative regime which is equally fit for purpose, if this regime produces less discriminatory output and if requiring the lender to use this alternative regime is overall proportionate if balanced against the interests of the applicants.[240]

On a side note, it might be worth noting that strategic business goals of the lender which go beyond credit default risk pose more complicated normative questions. Personalized pricing of a loan based on its value for the specific borrower provides an especially troubling example.[241] Borrowers in urgent need of a loan or with less sophisticated knowledge when evaluating and comparing interest rates will often be a vulnerable target for predatory lenders,[242] a practice sometimes dubbed as "reverse redlining".[243] In the US, some federal action has been taken through the Secure and Fair Enforcement for Mortgage Licensing Act[244] or the CFPB's qualified mortgage rule.[245] The DOJ, the CFPB and the OCC have in October 2021 announced the launch of a "Combatting Redlining Initiative" which includes "modern-day redlining" and discriminatory algorithms.[246] At the same time, strategic pricing is not illegal *per se*. The debate on this strategy in the context of algorithmic

---

[235] Noting that there is little guidance on this question under US law: Gillis (2022), p. 72; critical on the vagueness of this (English) concept Sacksofsky in Mangold/Payandeh (2022), note 45, 129 et seq.
[236] ECJ ECLI:EU:C:2015:480; ECLI:EU:C:2012:657.
[237] See above B.
[238] See above B.
[239] Hurlin et al. (2021) offer a methodology to distinguish whether a lender discriminates only for creditworthiness.
[240] See von Ungern-Sternberg in Mangold/Payandeh (2022), note 68, 75.
[241] See below E.III.2.
[242] DeYoung/Philipps (2006).
[243] Fisher (2009) pp. 126 et seq.
[244] Fisher (2009), p. 153.
[245] O'Keefe (2016).
[246] https://www.justice.gov/opa/pr/justice-department-announces-new-initiative-combat-redlining (last accessed 27 October 2022).

credit-underwriting is still in its infancy. The EU Consumer Credit Directive/2021 seems rather open towards personalized pricing.[247] The amendment proposed by the EP is somewhat more restrictive.[248]

5. The Third Hard Case: Neutral Building Blocks in Algorithmic Credit Underwriting

Above, I have identified the third hypothetical lender as the most realistic narrative.[249] He uses an algorithm and big data without deliberately choosing proxy variables that stand in for protected attributes. He might be aware of unequal output across groups but sticks with the algorithm "in spite of", not "because of" its disparate impact.[250] There are two potential avenues for a disparate impact/indirect discrimination case against a third hypothetical lender of this genre. A plaintiff could work her way through every single facially neutral variable the AI has access to and show how one – or many – of these variables, or various variables in their combination, trigger the disparate output. Instead, a plaintiff might claim that using the algorithmic model as such is a facially neutral practice.[251]

a) Bundles of Variables

The first strategy, looking at every single variable, can make sense in a limited-input model, illustrated, for instance, by the e-commerce model mentioned above which worked with only ten variables.[252] If the lender uses a model of this type, the plaintiff might be able to identify the suspicious variable. In a litigation context this presupposes that the plaintiff gets access to the variables the lender uses. To the extent that courts understand these as proprietary business information of the lender or the scoring agency, plaintiffs will face significant hurdles.

It is less clear whether this strategy will be successful if not one variable, but various variables in their combination produced the disproportionate output. Traditional doctrine has not dealt with these cases.[253] For credit underwriting, the Fintech lender Upstart provides an illustration.[254] To decide on a loan application, it uses a bundle of variables, such as education, employment history and more. Importantly, it only processes variables in concert, not in isolation.[255] The mystery shopping exercise

---

[247] See recital (40) referencing the AI Act and stating that "creditors, credit intermediaries and providers of crowdfunding credit services should be allowed to personalise the price of their offers for specific consumers or specific categories of consumers based on automated decision-making and profiling of consumer behavior allowing them to assess the consumer's purchasing power"; see further recital (39) and Art. 10 para. 3 (t) requiring transparency; in more detail below at E.III.

[248] See amendment 16, European Parliament, Report – A9-0212/2022, available at https://www.europarl.europa.eu/doceo/document/A-9-2022-0212_EN.html, deleting the reference to profiling of consumer behavior and adding the requirement for creditors, credit intermediaries and providers of crowdfunding credit services to inform consumers about the sources of data used for the personalisation of the offer.

[249] See above D.II.3.a.

[250] Hellman (2020), p. 852.

[251] See von Ungern-Sternberg in Mangold/Payandeh (2022), note 67 requiring justificatory reasons for both, the decision taken and the tool employed.

[252] Berg (2018), see above B.

[253] Doubting that indirect discrimination adequately covers these cases: von Ungern-Sternberg in Mangold/Payandeh (2022), note 92.

[254] See above B.

[255] CFPB (2017), p. 4.

on Upstart described above[256] suggested a disparate impact case along traditional lines. It held all other inputs constant and varied only educational background. Its ensuing report found significant differences between Black, Latinx and white persons as to both, loan origination fee and interest rate, depending on which college a loan applicant had attended. However, the problem with the mystery shopping exercise's methodology is that Upstart uses a bundle of variables which redundantly encode the same information. To give a comprehensive answer, the mystery shoppers would have needed to show the result the algorithm would have arrived at when eliminating the variable "college attended". The more sophisticated Upstart's algorithm and the broader its data base, the higher the probability that the model would have arrived at the same conclusion, even without including the college the applicant attended.[257] In theory, a plaintiff could go through many rounds of eliminating variables and stand-in variables. But the more variables are eliminated, the more we face a stripped-down version of the model which has little to do with the one the lender used.

b) The Algorithm as a Facially Neutral Practice

A second strategy to make a disparate impact/indirect discrimination case is to understand the entire algorithmic model as a facially neutral practice which leads to the discriminatory result. There are two concerns with this understanding. The first has to do with the counterfactual when eliminating the model as the building block. The second addresses the role of causation and correlation.

Plaintiffs must identify a building block which led to the decision and show that the outcome would have been different if the defendant had not based his decision on the relevant variable or practice. In line with disparate treatment/direct discrimination, this does not require the variable or practice to be the sole building block.[258] EU law even embraces a probability standard under which it must be more likely than not that the building block triggered the discriminatory output.[259] Still, neither the ECJ nor the US Supreme Court have accepted a showing of statistical disparity only.[260] This serves to rule out situations where random sampling caused the disparity.[261] It also rules out cases where the same disparate impact would have existed without the defendant engaging in the challenged practice.[262]

---

256 See above B.

257 Importantly, bundles of variables pose a different problem than mixed motives (on those see above D.II.2.a.). Mixed motives, and their treatment as a problem of but-for causation, provide an answer to the question whether a decision must be triggered *solely* by one variable (the answer is: no). With bundles of variables, it is the combination of the various variables which triggers the offending result.

258 Above D.II.2.a.

259 Above D.II.2.a

260 Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc., 135 S. Ct. 2507 (2015), p. 2523: this would be incompatible with a "robust" proof of causation, meant to ensure that defendants are not "held liable for racial disparities they did not create"; for an EU version of this argument see the discussion at Grünberger (2013), p. 660 et seq.; Mangold/Payandeh in Mangold/Payandeh (2022), note 110; von Ungern-Sternberg in Mangold/Payandeh (2022), note 91. A "robust" proof must show that statistical significance is "sufficiently substantial", see further Watson v. Fort Worth Bank & Trust, 487 U.S. 977 (1988), p. 995; providing a hypothetical under German law transposing EU law and suggesting a reversal of burden of proof: Grünberger (2021), p. 233; see further Hoffmann (2014).

261 Groves v. Alabama State Bd. of Educ., 776 E.Supp. 1518 (1991), pp. 1527–1528.

262 Ellston v. Talladega Cty. Bd. of Ed., 997 E.2d 1394 (1993), p. 1415.

To judge whether the same result would have been reached without the lender employing the algorithm, one needs to establish a hypothetical counterfactual. If the underwriting process was entirely automated, it is not clear how to proceed. Removing the AI model leaves us without a guideline for assessing the counterfactual. By contrast, if a human credit officer is involved, the plaintiff faces the challenge of proving that the credit officer would have come to a different, less discriminatory conclusion without the model. While not unthinkable, in practice this will amount to a very challenging task, even under the EU's probability standard.

One might instead be tempted to dispense with identifying relevant building blocks altogether when dealing with algorithmic decision-making. Each time an algorithm was involved in a decision, it would be sufficient to point to the output. If it proved disproportionate across groups, we would have to find a justificatory reason. This would be true even if it could not be established that the decision would have looked differently without the algorithm. Put differently: the mere fact of employing an algorithm would intensify the normative control of contracting choices.

Despite its short-cut appeal, a proposal along those lines faces several concerns. There are good reasons for received doctrine to require distinct building blocks to counterbalance normative control mechanisms. One of these is to publicly identify suspicious variables or practices which courts suspect of allowing to deliberately hide discriminatory intentions.[263] Another reason has to do with limiting responsibility. While disparate impact/indirect discrimination is not fault-based, a defendant has so far not been held responsible for consequences of past discrimination, understood as a wrong he did not create.[264] Without requiring to identify distinct building blocks, this line is blurred.[265] Yet another reason to require such building blocks is to incentivize defendants who become aware of their action's discriminatory potential to look for alternatives which are less discriminatory. (Only) if this cannot be established, a justificatory reason is available. This opens the door to discussing the quality of the algorithmic model employed, an avenue I will pursue below.[266] Furthermore, without identifying distinct building blocks and, instead, simply pointing to the algorithmic model, we would hold algorithmic decision-making to a higher normative standard than human decision-making, even if the discriminatory output was lesser than what a human decider would have produced.[267]

c) Beyond Discrimination: Output Only?

This last thought might have persuaded the CFPB to choose an entirely different path in its no-action letter for Upstart. The Bureau did not follow received doctrine of pointing out a neutral variable or practice, establishing disparate impact and looking for justificatory reasons. From what can be gathered based on publicly available information, the Bureau simulated outcomes under Upstart's

---

[263] See the "smoking out disparate treatment" argument in Ricci v. DeStefano, 557 U.S. 557 (2009).
[264] In the words of US Supreme Court Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc., 135 S. Ct. 2507 (2015), p. 2523 on racial disparities.
[265] See Wachter (2022), pp. 31–32 on explaining that "discriminatory behaviors carry with them an assumption of moral superiority" and claiming that "these notions make a lot of sense from a human lens (…) however, when algorithms being used it is not fully clear if their grouping invokes the same moral wrong".
[266] Below E.I. and in this context von Ungern-Sternberg in Mangold/Payandeh (2022), note 68.
[267] Critical as to output controls more generally: Mayson (2019), p. 2249: "imposing certain metrics of output equality will therefore have a cost in accuracy"; see further: Sunstein (2019), p. 509: "tradeoff between accuracy and fairness".

proprietary model and compared them with outcomes under a hypothetical model using FICO scores.[268] The simulation, based on data provided by Upstart, saw this lender approving 27 % more borrowers than traditional lending models. Personal loan interest rates were 16 % lower on average.[269] The CFPB found no disparities for minorities, females, or applicants 62 years or older. Put differently: Minority borrowers had better chances to be eligible for a loan under Upstart's model than under the hypothetical traditional model. The Bureau understood these findings to exclude disparate impact concerns, because borrowers fared better under Upstart's model than under a (hypothetical) traditional model.[270] The Bureau did not comment on the fact that, zooming in on one subset of minority borrowers (Black and Black-Hispanic applicants), the distribution was still skewed. Minority borrowers which were eligible under Upstart's model were facing disadvantages when compared with the subset of white and white-Hispanic persons eligible under Upstart's model. This was true as to relative numbers of access to credit, origination fees and interest rates. It is also consistent with the empirical studies described above.[271]

It is hard to say whether one should applaud the CFPB for this unusual strategy or criticize it for overreaching its authority. The Bureau's thinking might have been: If in absolute numbers more protected-group-borrowers have access to loans than under a hypothetical FICO score, this provides for more inclusion. Against that background, the Bureau might have claimed, it does not matter if the surplus is unequally distributed, everyone is better off.[272] But is this a convincing argument under the angle of anti-discrimination doctrine?

Looking to a hypothetical simulation with FICO as the counterfactual is not entirely unreasonable in a country where access to loans follows a standardized routine. It rewards lenders who offer an advantage, at least for some groups and at least if compared with the current situation.

However, two downsides are apparent. The current (in the US: FICO-based) standard remains the benchmark. This can hurt innovation and it can defeat the purpose of those FinTechs who wish to offer access to credit for borrowers who do not perform well under the traditional metric. Additionally, the CFPB's aggregate-view test is hardly compatible with received disparate impact doctrine. Anti-discrimination law is about relative disadvantages of one group when compared to another group. The Bureau focused instead on the surplus produced by Upstart's model, irrespective of the relative composition of the group of borrowers. This is not to say that such criteria are irrelevant or should not be pursued. They just cannot be justified under received anti-discrimination law.

In its 2020 renewal of the no-action letter, the CFPB even more clearly reoriented its investigation away from disparate impact doctrine. It did not ask for distinct building blocks, nor did it try to establish a hypothetical counterfactual to understand what would happen if the model were removed.

---

[268] Critical as to this method: Student Borrower Protection Center (2020), p. 21 fn. III but see above D.III.2.a. for a critique of the mystery shopping exercise.
[269] CFPB (2019); Upstart Blog: An Update from CFPB on Upstart's No-Action Letter, available at: https://www.upstart.com/blog/an-update-from-cfpb-on-upstarts-no-action-letter (last accessed 27 October 2022).
[270] CFPB (2019).
[271] See above D.III.2.
[272] On this argument see Langenbucher/Corcoran (2022), p. 156.

Rather, the CFPB mentioned a Model Risk Assessment Plan which Upstart is required to follow.[273] This includes model documentation as well as monitoring how Upstart's customer population and model performance change over time. The CFPB explicitly asked for "access-to-credit testing". What is left from received doctrine is the Bureau testing the "model and/or variables or groups of variables" for disparate impact and predictive accuracy by group as well as "research approaches that may produce less discriminatory alternative algorithms that meet legitimate business needs". The Bureau's strategy when dealing with Upstart points beyond received anti-discrimination doctrine and instead aims at controlling the quality of algorithmic decision-making via monitoring its use. In this form, it can be one of the cornerstones of a future regulatory design.

E. Next Steps: Towards a Regulatory Design for Consumer Credit in the Age of AI

The CFPB has not published data on the reasons which explain the asymmetry in origination fees and pricing across protected groups in Upstart's underwriting model. As described above, there are many possible explanations.[274] Disproportionate effects can go back to real differences in credit default risk which the model found. In this case, the model reflects existing inequalities in the world and helps with efficient statistical discrimination to allow calculating a risk premium. However, the inequality in output across groups could also be triggered (or compounded) by bias in the data or by bias in the model.[275] In that case, the model presents an inadequate understanding of the credit default risk of some groups. Yet another explanation looks to Upstart's business model. Maybe Upstart developed a profile which works especially well for white persons with an educational background in a predominantly white college who do not perform well under standard FICO-metrics. This could explain the surplus in borrowers the CFPB found with its FICO-simulation. Alternatively, Upstart might engage in strategic pricing. Its model could have figured out that the probability to accept less favorable terms was higher in protected communities.

All these explanations evidence how the shortcomings of received anti-discrimination law may be mitigated by an institutional design of consumer credit markets regulation which addresses the novel concerns brought about or compounded by algorithmic decision-making. EU and US law provide model outlines of a future design, each with a different focus. The EU AI Act includes requirements to carefully examine the quality of models and data the lender employs to guarantee that flaws of either do not produce skewed output.[276] Both, EU and US law give borrowers a chance to rectify incorrect data, given that the lender will not always be able to. The EU GDPR works across the board, the US FCRA targets the underwriting context. Arguably, both regimes need an update.[277] Additionally, US law provides disclosure in the form of specific notices to borrowers concerning adverse action on a credit application. This includes some information on credit scoring.[278] Eventually, legislators and regulators will have to address more far-reaching issues such as consumer

[273] CFPB (2020a).
[274] See above D.II.3. and D.III.2.
[275] See Gillis (2020), p. 18.
[276] See below E.I.
[277] See below E.II.
[278] See below E.III.

protection against personalized, at times predatory, pricing.[279] While this paper highlights these areas of future regulation, a detailed discussion is outside its scope.

I. Quality and Governance Control at the Lender

Quality issues of the AI model or the data can hurt both sides of the transaction. The borrower pays too much interest or is not eligible for a loan. The lender leaves money on the table if he denies a loan because the model incorrectly sends the applicant to a statistical bucket, he does not belong in. The EU AI Act[280] tries to respond to this by treating AI models as high-risk products in need of regulation.[281] The Act requires risk management systems which include continuous processes and regular updating.[282] Data governance and management practices look at training, validation, and testing data,[283] models must be regularly re-trained, and human oversight by "natural persons who have the necessary competence, training and authority"[284] must be ensured. Along similar lines, authors such as Citron and Pasquale have suggested licensing and audit requirements,[285] the FTC has initiated an advanced notice of proposed rulemaking and the US AI Bill of Rights talks about safe and effective systems.[286]

1. The EU AI Act

The EU AI Act introduces a risk-based approach for what it calls "AI systems".[287] A small number of AI systems are impermissible. Many face minimal or no compliance requirements and some are considered high risk. AI underwriting and scoring models fall in the high-risk category. Interestingly, the reason for placing them in the high-risk category is that these systems "determine (…) access to financial resources" and their use "may lead to discrimination of persons or groups and perpetuate historical patterns of discrimination (…) or create new forms of discriminatory impacts".[288] Still, the AI Act does not engage with a prohibition of discriminatory lending practices but starts from the assumption that it is regulated elsewhere.[289]

---

[279] See below E.IV.
[280] Proposal for a Regulation of the European Parliament and the Council Laying down harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) of 21 April 2021, COM(2021) 206 final. The text is based on the 4th Presidency Compromise Text of 10 October 2022. For better readability, I refer to this text as: AI Act.
[281] See Langenbucher (2022); Langenbucher/Corcoran (2022).
[282] Art. 9 AI Act.
[283] Art. 10 AI Act.
[284] Art. 29 para. (1a) AI Act.
[285] Citron/Pasquale (2014), p. 21 for employment, insurance and health care; id., pp. 24 et seq. on the FTC's statutory authority to combat unfair trade practices as to scoring.
[286] FTC (2022) and above D. II.1.b.
[287] Art. 1(a) AI Act; for credit scoring, German law asks for its own form of quality control, requiring a scientifically approved mathematical-statistical regime which can be shown to be relevant for assessing the probability of relevant behavior, see § 31 Bundesdatenschutzgesetz (but see D.II.3.a. for this rule's conflict with maximum harmonization of the GDPR).
[288] Recital (37) AI Act.
[289] Langenbucher (2022) p. 368; see below E.III.2. on the interplay between the AI Act and the Consumer Credit Directive/2021.

For high-risk systems, the Act follows the logic of regulating dangerous products, similar to the US Blueprint for an AI Bill of rights which lists the need for "safe and effective systems".[290] The AI Act (roughly) distinguishes five categories of compliance requirements which focus on data and data governance, technical documentation and record-keeping, transparency, human oversight, and checks on robustness, accuracy and cybersecurity. All these concern professional developers and users only. By contrast, the Act does not address the situation of end-consumers but instead delegates it to the private law of the EU Member States. Looking ahead, the EU Consumer Credit Directive/2021 proposes an anti-discrimination rule.[291]

Training, validation, and testing data sets are to undergo data governance checks. The drafters aim at data which is "to the best extent possible free of errors and complete".[292] They are aware of the pitfalls of collecting alternative data and ask developers to evaluate "availability, quantity and suitability of data sets" [293]. Developers must make sure their data set has the "appropriate statistical properties"[294] and reflects specifics of "the geographical, behavioral or functional setting within which the high-risk AI system is intended to be used"[295]. Additionally, they must identify "data gaps or shortcomings"[296] as well as "possible biases (…) that lead to discrimination prohibited by Union law".[297] To detect these, the proposal permits processing of sensitive data on protected characteristics if this is "strictly necessary for the purpose of bias monitoring".[298] The Act requires "appropriate safeguards (…) including technical limitations on the re-use and use of state-of-the-art security and privacy-preserving measures, such as pseudonymisation, or encryption" [299].

Developers must run compliance checks on their AI systems before putting them on the market. If they provide high-risk AI systems they must ensure that these systems undergo the relevant conformity assessment procedure and draw up an EU declaration of conformity.[300] Following the EU's new legislative framework (NLF),[301] the developers of the product carry out these conformity assessments.[302] For some products this involves a conformity assessment body, a private entity which Member States designate to run conformity assessments.[303] Developers of AI systems which are

---

[290] Available at https://www.whitehouse.gov/ostp/ai-bill-of-rights/ (last access 27 October 2022).
[291] Art. 6 Proposal for a Directive of the European Parliament and of the Council on consumer credits, COM(2021) 347 final.
[292] Art. 10 para. (3), recital (44) AI Act.
[293] Art. 10 para. (2) AI Act.
[294] Art. 10 para. (3) AI Act.
[295] Artt. 10 para. (4), 13 para. (3) lit. b no. i AI Act.
[296] Art. 10 para. (2) lit. g AI Act.
[297] Art. 10 para. (2) lit. f AI Act.
[298] Art. 10 para. (5) AI Act.
[299] Art. 10 para. (5) AI Act.
[300] Artt. 43 et seq. AI Act.
[301] See Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products and amending Directive 2004/42/EC and Regulations (EC) No 765/2008 and (EU) No 305/2011, 2019 O.J. L (169) 1; Regulation (EC) No 765/2008 of the European Parliament and of the Council of 9 July 2008 setting out the requirements for accreditation and market surveillance relating to the marketing of products and repealing Regulation (EEC) No 339/93, 2008 O.J. L (218) 30.
[302] For market surveillance see: Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20. June 2019 on market surveillance and compliance of products and amending Directive 2004/42/EC and Regulations (EC) No 765/2008 and (EU) No 305/2011, 2019 O.J. L (169) 1.
[303] See https://ec.europa.eu/growth/tools-databases/nando/ (last access 27 October 2022).

financial institutions under Union financial services legislation already follow a special compliance regime of regulated industries. A financial institution which uses a high-risk AI system fulfills various monitoring obligations of the AI Act by complying with the relevant Union financial services legislation.[304]

To ensure adequate market surveillance, Member States designate national competent authorities.[305] The AI Act's institutional design combines a sectoral with an omnibus oversight approach,[306] depending on the entity which uses the AI model or puts it on the market. As an integral part of financial services oversight, the banking regulator will supervise compliance with the AI Act if there is "a direct connection with the provision of those financial services"[307] and unless the Member State has identified another relevant authority.[308] However, not all entities which are involved in scoring or in extending credit of some sort are financial institutions. Many Fintech platforms are not,[309] nor are scoring agencies, insurance companies[310] or companies which the AI Act refers to as offering "essential private services" such as housing, electricity, and telecommunication. These non-banks fall under the jurisdiction of newly to be established regulatory agencies, entrusted with monitoring use and development of high-risk AI. Additionally, AI regulatory sandboxes are geared towards promoting innovation.[311]

The AI Act does not address private litigation.[312] In line with its spirit of product regulation, it speaks to developers and professional users, not to retail consumers or borrowers. A right to complain to market surveillance authorities is included,[313] but private rights of action for damages of retail borrowers fall under EU and Member State law. Additionally, a novel proposal for a EU AI Liability Directive[314] takes up some of these claims. While the plaintiff must establish defectiveness of the product, the damage suffered and causation, the Directive includes a presumption of defectiveness and of causation in certain situations. The drafters of the reform start from the assumption that fault-based liability, when faced with the complexity, autonomy, and opacity of AI, makes it impossibly hard for plaintiffs to establish a case. For non-contractual liability, the proposal shifts the burden of proof to the defendant and includes discovery provisions for plaintiffs seeking damages.

2. Challenges for Quality Control

[304] Artt. 17 para. (3), 18 para. (2), 20 para. (2), 29 para. (4) subpara. (2), para. (5) subpara. (2) AI Act.
[305] Art. 59 AI Act.
[306] See Langenbucher (2020) advocating a sectoral approach.
[307] Art. 63 para. (4) subpara. (1) AI Act.
[308] Art. 64 para. (4) subpara. (2) AI Act.
[309] Art. 37 Consumer Credit Directive/2021 requires Member States to ensure that Fintech platforms which match lenders and borrowers ("credit intermediaries" and "providers of crowdfunding services") fall under licensing and supervision by an independent competent authority.
[310] See recital (37), Annex III para. 5 lit. e AI Act.
[311] Artt. 53 et seq. AI Act.
[312] See recital (5a) AI Act.
[313] Art. 63 para. (11) AI Act.
[314] Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive), COM(2022) 496 final.

One of the starting points of the EU AI Act is the assumption that an AI scoring outfit that uses historically biased training data will often come up with underwriting models which present a snapshot of reality at some point in time.[315] Its value for assessing borrowers then depends on the match between the snapshot and today's world. If the way in which today's world is different from yesterday's world does not matter for credit default risk, using old data might not hurt much. But if the training data implies, for instance, restrictions on taking out loans or an income distribution across sex or race which do not correctly represent today's world, the model will come up with skewed scores. The same is true if the score attributed to applicants who share features of historically privileged communities is higher than what their actual situation suggests.

Against this background, there seems to be a straightforward solution: Why not run compatibility checks between yesterday's and today's world? Unfortunately, a core problem for controlling for biased data is not only about identifying the extent of that match. In many cases, the lack of compatibility is not apparent, or the problem extends to choosing variables. An often-cited example for this latter concern has to do with a decision-making algorithm used by US hospitals. The algorithm allocated patients to programs improving care for those with complex medical needs.[316] A machine-learning (ML) research team at UC Berkeley received data from a hospital to work on ML and health care services. The researchers were surprised to find lower risk scores for Black persons which were equally sick as white persons who were assigned higher scores. They found that the algorithm looked to total health-care costs per year as the variable to assess risk scores. However, as an indication of how sick a person is, this proved wrong for Black persons. Health care administered to them cost considerably less per year than health care provided to a white person with a similar health profile. As a result, Black persons had to be much sicker than white persons to be allocated the same risk score. Put differently: They had to wait much longer to receive the personalized care for patients with complex medical needs.

The example illustrates the plethora of problems. One problem has to do with awareness of data or model quality. Without the researchers and their statistics, the problem with the model's choice of variable might not have been detected at all.[317] A possible fix for this are regular model audits which include a thorough examination of the underlying variables and assumptions.[318] Variables can be fitted to subgroups, if the traits "are more predictive for one race than for another", as Hellman suggests.[319] This strategy can help to cope with the example discussed earlier where non-performance was used as a core variable but implied different things across groups.[320] If a model attributes equal weight to non-performance across groups of borrowers, this might not adequately reflect credit default risk of each person in the population. If one group of the population consistently faces higher

[315] See above C.I. and Hurlin et al. (2021), p. 3 on using training data based on past decisions made by biased loan officers.
[316] Obermeyer et al. (2019); on this example Burrell/Fourcade (2021); Langenbucher (2020), p. 555; Ledford (2019).
[317] On a related point see von Ungern-Sternberg in Mangold/Payandeh (2022), note 20.
[318] Ledford (2019); see the "equity assessments" which the US Blueprint for an AI Bill of Rights suggests under Algorithmic Discrimination Protections, available at https://www.whitehouse.gov/ostp/ai-bill-of-rights/ (last access 27 October 2022).
[319] See Hellman (2020), pp. 853 et seq, providing examples which have been discussed in the literature.
[320] See above C.I.

interest rates despite being similarly situated, the probability of default in that group will be higher.[321] The reason for this is not the initially lower creditworthiness of each member of this group, but the higher burden to pay back. To raise awareness in those and similar situations, diverse coding teams and awareness trainings have been suggested as a step towards better model quality control.[322] The AI Act requires that "high-risk AI systems that continue to learn after being placed on the market or put into service shall be developed in such a way to eliminate or reduce as far as possible the risk of possibly biased outputs influencing input for future operations ('feedback loops')".[323]

Still, there are no easy and straightforward fixes. Better variables to predict risk might be hard to find. The users of the model might prefer to go ahead with an imprecise model, rather than with no model at all. This is especially true if the imprecise model still performs better than biased and cognitively limited human credit officers.[324]

Additionally, the example of the hospital algorithm concerns a conscious choice of variable (total health care costs per year) by the coders. If ML algorithms take over the process of choosing variables and attributing weight to them, quality checks are much more challenging.[325] Reverse engineering to identify a set of core variables has been proposed as a model quality check. Competition between AI models is another approach. Data audits can help if some groups are penalized because of a lack of relevant data on the members of this group. None of these strategies will necessarily help with historical bias when conditions change.[326] Take, for instance, a rule under which married women were not eligible for credit unless their husband signed the loan contract. It is one thing to have the model integrate the change between yesterday's and today's world, once legal reform allows unmarried women to sign a loan on their own. But (like the opaque bundles of proxies discussed above) the status as a married woman will be encoded redundantly in many other variables. Tweaking the model by fitting it to the subgroup of unmarried women will, for some time, make the model less precise for lack of data on that subgroup.

II. Consumer Rights to Rectify Incorrect Data

The EU AI Act has focused on the quality of algorithmic decision-making at the lender. However, information gathered from sources such as social networks, internet usage or behavioral tests is considerably more prone to mistakes and misunderstandings than traditional credit reporting data.[327] Lenders are not able to remedy these flaws if they outsource data collection. It follows from there, that regulators and legislators must provide tools for borrowers to rectify flawed data, possibly also to incentivize them to use these tools.

---

[321] See O'Neill (2016), p. 144: "nasty feedback loop".
[322] Benjamin (2019).
[323] Art. 15 para. (3) subpara. (3) AI Act.
[324] Ledford (2019); reaching the same conclusion for the criminal justice system: Mayson (2019), p. 2277; however, see Kim (2022), p. 5 in the context of the inability of AI underwriting models to learn from false negatives (see C.III above): "one of the claimed benefits of AI – its ability to learn over time – is far more limited when used to make decisions about people".
[325] Citron/Pasquale (2014), p. 5.
[326] See above C.
[327] See above C.III.

In the US[328] credit reporting agencies collect various data on consumers in a credit report. This is mostly financial data such as payment history, loans, and current debt, additionally personal data on place of work, home, prior arrests, lawsuits, or bankruptcy.[329] Credit reporting agencies furnish data to lenders or scoring agencies, but do not themselves develop a score.[330] The most well-known credit scoring agency is FICO.[331]

Credit reporting in the EU varies across Member States.[332] The German SchuFa, as an illustration, stores personal information on place of residence, along with financial data such as bank accounts, mobile phone contracts, payments in instalments, or collection procedures.[333] SchuFa does not collect the data itself. Instead, companies which are in long-term contractual relationships with customers, such as banks, telecommunication companies, or mail-order companies, furnish SchuFa with this data.[334] Based on that data, SchuFa develops a credit score which SchuFa's contractual partners may access.

US and EU law in that space diverge much more markedly than in anti-discrimination. The US follow a sectoral approach, targeting the role of credit reporting agencies and providing certain rights for consumers to access and rectify data and get information on their credit score.[335] The EU has not yet explicitly harmonized credit reporting or credit scoring across Member States. However, collection, processing, and furnishing of personal data[336] fall under the GDPR. Credit reporting agencies and lenders must provide justificatory reasons to lawfully collect and use the data. Additionally, they must inform consumers (the "data subjects") when they collect data and what its purpose is. Consumers have rights to access personal data and demand rectification.[337] The Consumer Credit Directive/2021 outlines which data may be used for an underwriting decision and includes some general rules on credit reporting.

1. US Law

The US have since the 1970s a statutory framework in place, regulating certain consumer rights as to credit reporting and scoring. The FCRA provides safeguards for borrowers who want to dispute completeness or accuracy of a credit report.[338] Additionally, under section 615(a), (h), lenders must provide a risk-based pricing notice. It discloses that the lender used a credit report, which furnished information beyond the lender's own files or information the applicant provided. Its purpose is to

---

[328] The most well-known ones are Equifax, Experian and Transunion.
[329] For the US see: https://www.usa.gov/credit-reports.
[330] For the US see: https://www.usa.gov/credit-reports.
[331] https://www.fico.com/en/products/fico-score.
[332] Recital (49) Consumer Credit Directive/2021, ensuring cross-border access to private or public databases across Member States.
[333] https://handbookgermany.de/en/schufa.
[334] https://handbookgermany.de/en/schufa.
[335] Comparative overview at Langenbucher (2020), pp. 534 et seq.
[336] Defined in Art. 4 para. (1) GDPR (requiring information relating to an identifiable natural person); see exception under Art. 2 para. (2) GDPR.
[337] Langenbucher (2020), pp. 536 et seq.
[338] 15 U.S.C. § 1681(i)(a)(1)(A); Langenbucher/Corcoran (2022), p. 162.

allow the applicant to verify the information in the credit report.[339] An applicant can notify the FTC which will conduct a reasonable reinvestigation to determine the accuracy of the information and, if incorrect, have it deleted.[340] In that way, the statute strikes a balance between the interests of lenders in statistical discrimination and the consumers' right to accuracy of information and, to some extent, financial privacy.[341]

Consumer rights under the FCRA turn on a data aggregator qualifying as a "consumer reporting agency". Only then does the data aggregator face compliance requirements, enforced by the FTC and the CFPB.[342] Solely consumer reporting agencies must comply with a consumer's request to access data which the data aggregator has furnished to third parties[343] and to rectify incorrect entries.[344]

However, many data aggregators which collect and process the vast input of non-financial information currently operate outside that regulatory perimeter.[345] To qualify as a consumer reporting agency under the FCRA, the person must regularly engage in the practice of assembling and evaluating consumer reports.[346] The FTC has understood evaluation as "appraising, assessing, determining or making a judgment on such information". "An entity that performs only mechanical tasks in connection with transmitting consumer information is not a consumer reporting agency".[347] Instead, persons which perform mechanical tasks of this type are qualified as a "conduit" only.[348] Various data aggregators have claimed that they should be understood as a conduit, and the FinTech lending platforms qualified as evaluating the consumer report.[349] The CFPB seems more open to bringing data aggregators under its jurisdiction. Section 1033 of the Dodd-Frank-Act is cited towards that end[350] and competitors of Fintech firms from the banking industry are urging the Bureau to do so.[351]

On the one hand, this suggests that even the minimum requirement of being able to access data and rectify flawed data is for alternative data not guaranteed under US law.[352] Arguably, AI underwriting models and big data suggest an overhaul of the statutory rules. To that end, many US states have started to introduce privacy laws or are in the process of doing so. The FTC has published an advanced

---

339 On the combination with an ECOA notice see below E.III.
340 Citron/Pasquale (2014), pp. 14 et seq.; Langenbucher (2020), p. 535.
341 Langenbucher (2020), p. 534.
342 Section 1033 of the Dodd-Frank Act transferred rulemaking responsibilities under FCRA from the FTC to the newly established CFPB. Enforcement authority rests with the FTC.
343 Langenbucher (2020), pp. 535.
344 See 15 U.S.C. § 1681i § 611 (a)(1)(A).
345 Citron/Pasquale (2014), p. 20; Langenbucher/Corcoran (2022), p. 151.
346 Kim/Hanson (2016), pp. 21 et seq.
347 FTC (2011) p. 29; see for a narrow reading of the LexisNexis product "Accurint" which was not considered delivering "credit reports": Kim/Hanson (2016), p. 28.
348 Id., p. 29.
349 NCLC (2020), p. 8; see the Ninth Circuit on a similar argument when it decided that *Fannie Mae* was not a consumer reporting agency, Zabriskie v. Federal National Mortgage Association 912 E.3d 1192 (9th Cir. 2019); but see Kim/Hanson (2016), pp. 30 et seq. for other courts reaching a different conclusion.
350 CFPB (2020b), pp. 71009 et seq.
351 ABA (2022).
352 On further concerns (difficulty for furnishers to determine whether information they have provided is negative and therefore requires notice to be sent to the consumer): Langenbucher/Corcoran (2022) p. 163.

notice of proposed rulemaking,[353] and the US Blueprint for an AI Bill of Rights includes a right to privacy.[354] On the other hand, the FCRA risk-based pricing notice provides rough contours of a future tool to encourage borrowers to verify data underlying a credit decision,[355] even if further work is required to understand what verifying alternative data could look like.[356]

2. EU Law

EU law does not face the problem of having to fit data aggregators under a narrow statutory definition such as the US FCRA's "credit reporting agency". The GDPR allows to qualify traditional credit reporting agencies as well as novel data aggregators and Fintech platforms as data processors.[357] Irrespective of the type of entity involved in collecting or processing personal data, any such action is only lawful according to a list of justificatory reasons.[358] At first glance, consent by the borrower seems a natural justificatory reason.[359] However, the standard for freely given consent is strict. Additionally, consent is under the GDPR revocable. Hence, data aggregators typically look for different justificatory reasons. The most relevant reasons ask for a request of the data subject prior to entering into a contract,[360] or for legitimate interests of the lender.[361] Should the data collected involve protected categories, the GDPR additionally requires explicit consent.[362] Under the EU Digital Markets Act (DMA),[363] entered into force on 1st November 2022, data aggregators can not necessarily rely on consent if they qualify as a gatekeeper.[364] This new restriction for gatekeepers covers the combination of personal data which the gatekeeper obtains as part of its core service with data from other platform services it provides. It also rules out "legitimate interest" as a justification for combining or cross-using data.

The EP's amendment to the Commission's proposal for the Consumer Credit Directive/2021 introduces a right for consumers to be notified of any negative credit data[365] along with a requirement

353 FTC (2022).
354 Available at https://www.whitehouse.gov/ostp/ai-bill-of-rights/ (last access 27 October 2022).
355 Below E.III.
356 Below E.II.2.
357 Art. 4 para. (2), (8); Langenbucher/Corcoran (2022), p. 149, on FCRA id., p. 150, and Langenbucher (2020), p. 534.
358 In more detail Langenbucher (2020), pp. 534 et seq.
359 Art. 6 para. (1) lit. a GDPR.
360 Art. 6 para. (1) lit. b GDPR.
361 Art. 6 para. (1) lit. f GDPR.
362 Art. 9 GDPR.
363 Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act), 2022 O.J. L (265) 1.
364 Art. 2 para. (1), (3), (5) DMA.
365 Art. 19 para. (4a) final draft of the European Parliament, available at https://www.europarl.europa.eu/doceo/document/A-9-2022-0212_EN.html..

for national competent authorities to conduct regular audits on databases to assess their compliance with the GDPR[366] and to ensure that complaint procedures are in place for consumers.[367]

Along those lines, the GDPR and the Consumer Credit Directive/2021 provide borrowers with robust rights to access personal data a credit reporting agency holds on them.[368] This covers both, traditional and alternative data. Any data controller must inform the person from whom it collects data and to explain how to get access. Applicants can request confirmation on whether an entity processes personal data on them and what the purpose of such processing is.[369] Additionally, consumers have rights to rectification and erasure.[370]

In theory, these are core tools, permitting consumers to verify that credit is not denied due to incorrect data. Lenders profit to the extent that their models are trained on verified data. In practice, it is often hard to incentivize consumers to use these tools, especially as far as alternative data is concerned. The US FCRA-notice explained above[371] and the EU rights under the GDPR share this concern. Depending on the type of alternative data, it might not be evident what rectification of flawed data would look like. While this is evident for facts such as name or address, it is less clear when data on taste in music, number of typos in text messages or speed when filling out an online form is at stake. The number of collectors of alternative data is as vast as the potential purposes for which they collect data. To comply with the requirement to inform consumers, data aggregators use highly standardized tools which often amount to a tick-the-box-exercise.[372] Rational apathy of consumers is a natural consequence.[373] This might be mitigated if consumers were alerted again when data is used in a credit context. However, once information is lawfully collected by the data aggregators,[374] there is under the GDPR no additional requirement to inform the consumer again when he furnishes the data to a lender, to a credit reporting or scoring agency. In contrast with the EU, this is an advantage of the US institutional design. As discussed in the following section, it provides disclosure in the form of specific notices on the use of data in an underwriting situation.

III. Consumer Rights to be Informed About Scoring and the Reasons for Denying Credit

---

366 Art. 19 para (4b) final draft of the European Parliament, available at https://www.europarl.europa.eu/doceo/document/A-9-2022-0212_EN.html..

367 Art. 19 para (4c) final draft of the European Parliament, available at https://www.europarl.europa.eu/doceo/document/A-9-2022-0212_EN.html. Interestingly, the word "efficient" before "complaint procedures" is deleted in its latest version, see https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/IMCO/DV/2022/07-11/18-CCD-FinalCAs_EN.pdf.

368 Art. 13 GDPR regulates data the entity collects, Art. 14 GDPR includes data received from third parties, Art. 15 provides a right of access, see Langenbucher (2020), pp. 539 et seq; von Ungern-Sternberg in Mangold/Payandeh (2022), note 93.

369 Art. 15 GDPR.

370 Artt. 16, 17 GDPR.

371 E.II.1.

372 On US law see Awrey/Macey (2022); FTC (2022), pp. 51274 et seq.; see Art. 34 EU Consumer Credit Directive/2021 on requiring Member States to promote financial education.

373 Langenbucher (2020), pp. 535–536.

374 Art. 13 para. (1) lit. e GDPR requires this information only when collecting the data. If the creditor processes data collected by a data aggregator, the creditor must inform the applicant under Art. 14 para. (2) lit. f GDPR. Banks typically provide for such information in their general terms and conditions.

Credit scores are one of the main building blocks of an underwriting decision.[375] Their highly standardized traditional metrics, but also their algorithmic version, driven by big data, have both been blamed for discriminatory lending practices.[376] Against this background, consumers have an interest in understanding how their score is built. Lenders, by contrast, will be reluctant to sharing proprietary information. Especially if alternative data are included, there is an additional risk of gaming the system.[377] A scoring agency that found attributes which are particularly predictive of credit default risk, for instance having a dating or a finance app installed, has little interest in disclosing that. The fear is that as soon as potential borrowers become aware, they will react and delete (or install) the relevant app. For the scoring agency, this means losing a variable which was easy to establish and had good predictive force.[378]

1. US Law

Under US law, applicants for a loan receive specific information on any adverse action of the lender. This serves two policy goals: First, the notice enables the applicant to check that no discriminatory reasons underly the adverse action. Second, it allows to make sure that no incorrect data is used in the credit report.[379] Section 202.9(a)(2)(i), (b)(2) of Regulation B, implementing the ECOA's anti-discriminatory goals, gives consumers a right to a statement of specific reasons for adverse action on a credit application ("ECOA notice"). According to the official interpretation of this rule,[380] if failure to meet the lender's credit score threshold is the reason, he must disclose "those factors actually scored in the system". "No factor that was a principal reason for adverse action may be excluded from disclosure. The creditor must disclose the actual reasons for denial (for example, "age of automobile") even if the relationship of that factor to predicting creditworthiness may not be clear to the applicant". As noted above,[381] section 615(a), (h) of the FCRA requires lenders to provide a risk-based pricing notice. Its goal is to disclose that the lender used a credit report and allow the applicant to verify the information in the credit report. If the lender combines both notices, disclosing that a credit report was used does not satisfy the ECOA notice requirement. The lender must identify which reason(s), revealed in the credit report, drove the adverse action.

The requirement to provide specific reasons was developed against the background of standard information contained in credit reports and the experience of discriminatory lending practices. It remains to be seen how regulators will interpret these rules when faced with algorithmic scores, based on bundles of big data variables and black-box algorithms. However, like the FCRA-notice discussed above,[382] it is worth noting both, the anticonsumer-friendly spirit of that law and the attempt to provide a tool, targeting the underwriting situation. This latter aspect distinguishes the US regulatory

---

[375] Above A.
[376] Above B.
[377] See Citron/Pasquale (2014), pp. 29 et seq.
[378] Langenbucher (2020), pp. 542 et seq.
[379] Above E.II.
[380] Quotes in what follows are available at: https://www.consumerfinance.gov/rules-policy/regulations/1002/9/#9-b-2-Interp-2.
[381] Above E.II.
[382] E.II.1.

design from its EU counterpart which follows an omnibus approach across all instances of data collection.

2. EU Law

a) The GDPR

The EU GDPR does not provide a distinct tool requiring lenders to disclose their reasons for denying credit.[383] Under its general rights to access personal data, the borrower must be informed about some of the ingredients of an underwriting decision, as it were. However, courts have so far not extended this to disclosing the proprietary models lenders or scoring agencies use to evaluate creditworthiness.[384] Having said that, there is a case pending at the ECJ, involving the German credit scoring agency SchuFa.[385] It concerns denial of credit to an applicant after the SchuFa had produced a negative credit score to the lender. The applicant requests detailed information on individual variables used and on their weight in its scoring model. The court referred the case to the ECJ for an interpretation of Art. 22 GDPR. This rule prohibits decisions based solely on automated processing, subject to a list of justificatory reasons.[386] If scoring qualifies as automated decision-making under Art. 22, the consumer's right to access his data extends to "meaningful information about the logic involved" (in automated processing) "as well as the significance and the envisaged consequence of such processing".[387] A similar requirement concerns the moment of data collection.[388] It remains to be seen how the ECJ will balance access to data against the lender's rights concerning proprietary business secrets.[389] Recital (15) explains the rule on automated processing as an attempt at being technology neutral and preventing circumvention. However, recital (71) goes beyond that limited goal. It is the only time where the GDPR explicitly mentions the underwriting context, referencing the "automatic refusal of an online credit application" and linking it to profiling[390] based on various variables. Recital (71) does not prohibit automated decision-making, but mentions justificatory reasons, including consent and "suitable safeguards". Among these, the recital lists "human intervention, to express his or her point of view, to obtain an explanation reached after such assessment and to challenge the decision".

The GDPR tool on automated processing differs from both, the US rule requiring an ECOA notice and the FCRA. The ECOA has discriminatory lending practices in mind and aims at furnishing the borrower with a tool to detect such practices. The FCRA ensures access to data for the purposes of

---

[383] But see below E.III.2.b. on the Consumer Credit Directive/2021.
[384] Langenbucher (2020), p. 542.
[385] C-634/21, Request for a preliminary ruling, 15 October 2021; on SchuFa see above E.II.
[386] These include: the necessity for entering into or performance of a contract, Art. 22 para. (2) lit. a, authorization by the Member State, lit. b, or explicit consent, lit. c. For the exceptions under lit. a and c, the data controller must provide a right to obtain human intervention and to contest the decision, Art. 22 para. (3). Additional requirements concern decisions which involved protected categories, Art. 22 para. (4).
[387] Art. 15 para. (1) lit. h GDPR.
[388] Art. 13 para. (2) lit. f GDPR.
[389] See recital (4) on "the right to the protection of personal data (…) considered in relation to its function in society and (…) balanced against other fundamental rights, in accordance with the principle of proportionality"; further see below E.III.2.b. on a similar provision in the Consumer Credit Directive/2021.
[390] Defined in Art. 4 para. (4) GDPR.

verification. The GDPR, more broadly, puts data protection in the context of fundamental rights[391] and, against that background, tries to capture the situation of natural persons faced with a machine.[392] Adopted almost 50 years after the FCRA and the ECOA, the GDPR is already familiar with the risks now materializing in the form of big data and algorithmic decision-making. Still, the spirit of both laws is similar in one respect, namely the attempt to grant the consumer a right to get meaningful information. The challenge for both is to understand what this entails for algorithmic underwriting practices.

b) The AI Act and the Consumer Credit Directive/2021

The two pieces of EU legislation which are currently in the making, the AI Act and the Consumer Credit Directive/2021, engage with scoring and credit underwriting when faced with algorithmic decision-making. As noted above, while the AI Act takes the risks of discriminatory practices as the policy goal for placing algorithmic credit scoring and underwriting in the high-risk category, consumer rights are outside its scope.[393] Instead, the Consumer Credit Directive/2021 explicitly prohibits discriminatory lending practices.[394]

Somewhat similar to the ECOA notice, the Consumer Credit Directive/2021 requires the lender to inform the consumer of a denial of credit if "the application is rejected on the basis of a consultation of a database".[395] The rule mentions automated processing being involved. However, in contrast to the ECOA, it does not include an obligation to give specific reasons. Furthermore, the policy goal of this rule does not seem to have discriminatory lending practices in mind. Instead, it aims "to prevent irresponsible lending practices and overindebtedness".[396]

Linking the Consumer Credit Directive/2021 to the GDPR, the Directive requires lenders who use profiling or automated processing to make sure that consumers have the right to obtain human intervention.[397] Additionally, the Directive takes first steps at engaging with the balance between the rights of consumers to be informed against the interests of lenders in their proprietary trade secrets.[398] The Commission has included a right to "request a clear explanation of the assessment of creditworthiness, including on the logic and risks involved in the automated processing of personal

---

[391] See recitals (1), (2).
[392] See below E.III.2.b.
[393] Above E.I.
[394] Above D.III.
[395] Art. 19 para. (4) final draft of the European Parliament, amendment 160, available at https://www.europarl.europa.eu/doceo/document/A-9-2022-0212_EN.html.; includes the categories of data taken into account. Additionally, the EP proposed to amend recital (50) to require information of consumers on new negative data being entered into a credit reporting database and to include procedures for consumers to challenge the content of credit databases and the outcome of database searches.
[396] Art. 18 para. (1). The amendment proposed by the EP seeks to entrust EBA with the power to develop guidelines on the type of data to be used for creditworthiness assessments, see amendment 135 to Art. 18 para. (2) subpara. (1a), new, available at https://www.europarl.europa.eu/doceo/document/A-9-2022-0212_EN.html.
[397] The EP's amendment restricts this to a denial of credit, see amendment147 available at https://www.europarl.europa.eu/doceo/document/A-9-2022-0212_EN.html.
[398] Above E.III.2.a. on this question under the GDPR, pending at the ECJ.

data as well as its significance and effects on the decision". Member States must ensure that consumers have a right to contest the decision.[399]

The EP aims to provide even more detailed insight into the scoring process. It proposes to include a requirement to inform on "the categories of data processed as part of the assessment and the weighting of each category in the decision".[400] This responds to the concern that, today, we can only speculate which sets of applicants will in the long-term profit from AI underwriting. We might eventually be looking at outcomes which do not reflect what the current legal framework of anti-discrimination laws has in mind.[401] Consider, for instance, AI underwriting models which lead to unanticipated imbalances in the distribution of loans outside the definition of protected groups: persons who do not take care of regularly updating their software, who do not have a social media presence, whose IoT fridge often signals a lack of alcoholic beverages or who regularly google information on moving houses. They might find it hard to get a loan approved, irrespective of their sex, race, or religion. However, the compatibility of a right along these lines with black-box AI and neural networks is unclear, if these make it hard or altogether impossible to establish which one of these variables drive an AI's score. The question is not addressed in the amendment proposed by the EP.

Furthermore, the Directive directly targets the use of alternative data. It starts from the assumption that the data lenders look to when checking credit default risk should be financial data. These concern "the consumer's income and expenses and other financial and economic circumstances".[402] This information must be "relevant and accurate" as well as "appropriately verified".[403]Additionally, the EP's proposal introduces a prohibition for creditors[404] to use and providers of crowdfunding services[405] to hold sensitive data[406] and "data collected from digital social networks".[407] Credit intermediaries[408] are not addressed, hence, it remains to be seen whether these will include social networks when scoring. Arguably, given redundant encoding and flexibility of multivariate regressions, there are limits to meaningful input control along the lines proposed by the EP.

The proposal seems somewhat undecided how to best design a framework taking moral concerns into account. It refrains from explicitly addressing the underlying tension between the chance for more granular predictions at the expense of sensitive personal data nor does it explicitly engage with the potential inclusionary consequences of using big data in credit underwriting. Where the Directive will

---

[399] Art. 18 para. (6), the EP's amendment changes the wording. Instead of contesting the decision, it proposed to include a right to request a review, see amendment 149 available at https://www.europarl.europa.eu/doceo/document/A-9-2022-0212_EN.html.
[400] Amendment 148 available at https://www.europarl.europa.eu/doceo/document/A-9-2022-0212_EN.html.
[401] See Wachter (2022), pp. 15–29.
[402] Art. 18 para. (2) Consumer Credit Directive/2021, see recital (47) referencing the EBA guidelines on loan originating and monitoring.
[403] Art. 18 para. (2).
[404] Defined in Art. 3 (2).
[405] Defined in Art. 3 (4).
[406] This comprises "personal data revealing racial or ethnic origin, political opinions, religious or philosophical belief, or trade union membership, and the processing of genetic data, biometric data (…), data concerning health or data concerning a natural person's sex life or sexual orientation", Art. 18 para. (2) subpara. (2a), Art. 19 para. (3) final draft of the European Parliament available at https://www.europarl.europa.eu/doceo/document/A-9-2022-0212_EN.html.
[407] Art. 19 para. (3) final draft of the European Parliament.
[408] Defined in Art. 3 (12).

eventually land on rights for consumers to access detailed information on the scoring model remains to be seen, including the case under the GDPR which is pending at the ECJ.

IV. Personalized Pricing

Arguably, an overhaul of the regulatory framework for consumer credit underwriting should not stop at updating the traditional toolbox outlined in the previous section but engage with further risks to fair lending which will be compounded by AI models. Their optimization goals do not only assist in finding invisible primes.[409] They can also be helpful in identifying applicants which are likely to sign a loan above market price, for instance because they do not have the time or skill for comparison shopping.[410] Vulnerable borrowers will often not be aware that the access to the data they provide hurts, rather than helps them.

An AI's optimization goal is one element of lender's proprietary trade secrets, defining what the algorithm is looking for.[411] A possible optimization goal is credit default, enabling the lender to build groups of borrowers and benchmark his offer against market prices for comparable loans Under this assumption, one would expect comparable terms and conditions across lenders when faced with similarly situated applicants. Bank regulators as well as courts are likely to encourage, rather than limit the use of AI underwriting models to that end. This seems particularly appealing if algorithms contribute to more granular risk assessment, better risk-adjusted pricing, and in this way contribute to sound and stable financial markets. However, empirical studies have identified inequality in output even for similarly situated persons.[412] Borrowers vary in their access to information, their financial literacy, and the urgency of their need for credit. As Fuster and his co-authors have speculated, this could explain unequal outcome across groups which does not track variation in credit default risk but enables pricing above market.[413]

The interaction between personalized pricing, risks of credit underwriting, predatory lending and algorithmic models sorting applicants according to vulnerability and financial literacy[414] is beyond the scope of this current paper, safe for highlighting the need for legislators and regulators to engage with this topic. The Consumer Credit Directive/2021 takes a first step by requiring lenders to inform borrowers "that the price was personalized on the basis of automated processing."[415] However, the key piece of information is not about automated processing, but about understanding that prices differ even for comparably situated borrowers. A regulatory requirement to disclose this to applicants might set some incentives for more comparison shopping, thereby enhancing competition between lenders. The Consumer Credit Directive/2021 does not go that far. To the contrary and despite its moral

---

[409] See above B.
[410] Aggarwal (2021), p. 50.
[411] O'Neil (2016), p. 21 ("definition of a success").
[412] Hurlin et al. (2021) offer a methodology to distinguish whether a lender discriminates only for creditworthiness.
[413] See above D.I.
[414] Introduction at Eidenmüller/Wagner (2021), pp. 50–54; Ernst (2017).
[415] Art. 10 para. (3) lit. t; recital (40) in the final draft of the European Parliament for informing consumers about the sources of data used for personalization.

concerns around alternative data, the Commission's draft explicitly encourages lenders to "personalise the price of their offers for specific consumers or specific categories of consumers based on automated decision-making and profiling of consumer behavior allowing them to assess the consumer's purchasing power".[416] This does not allow for discriminatory personalized pricing. Under the US FHA and the ECOA, the US DoJ has discussed this for practices such as "reverse redlining".[417] However, many targeted offers of this type will escape liability under anti-discrimination laws,[418] if they target, for example, students, recent immigrants, or refugees. These lenders do not qualify under anti-discrimination laws unless the loan portfolio is skewed towards protected characteristics.

F. Summary

The potential of big data and AI credit underwriting models to lower search costs for lenders when assessing the creditworthiness of loan applicants marks the introduction to this paper. There is considerable empirical evidence on the achievements of Fintech companies to provide access to finance for groups which have previously found this difficult. Going beyond standard metrics, algorithms and big data can help to identify invisible primes. Along those lines, they provide a useful tool to make more granular predictions than traditional metrics. However, empirical analysis has brought out more troubling findings, too. Studies suggests that Fintech algorithms tend to produce unequal output across groups of loan applicants.

Part B provides an overview on empirical studies which evidence both, the potential for inclusion and for inequality.

Part C summarizes main findings of computer scientists, pointing to different types of model biases, the fact that AI models redundantly encode information in various variables, and concerns about inaccurate data.

Against that background, part D explores how US and EU anti-discrimination laws fare when faced with algorithmic credit underwriting. Its core finding is that received anti-discrimination doctrine is ill-suited to deal with algorithmic discrimination.

Part D.I. introduces the concept of building blocks, understood as input to decision-making. Anti-discrimination law's received understanding is rooted in distinct building blocks along a chain of causation. Many of these building blocks are harmless, lawful attributes, practices or motivations. Others concern protected characteristics and are considered outright unlawful. A few are facially neutral but suspicious, standing in for a protected characteristic. Implicit is the understanding that there is a limited number of proxies available for human decision-makers. The paper submits that, in the face not only of a massive increase of potential building blocks, but also of bundles of variables

---

[416] Recital (40); the EP proposes to delete the part on profiling, see amendment 16 available at https://www.europarl.europa.eu/doceo/document/A-9-2022-0212_EN.html. On profiling under the GDPR see Kaminski (2019), p. 1551; Langenbucher (2020), pp. 538, 540.
[417] Available at https://www.justice.gov/crt/housing-and-civil-enforcement-cases-documents-231 (last access 27 October 2022).
[418] Ernst (2017), p. 1034.

which, taken together, accurately predict a protected characteristic, anti-discrimination law's received concept will lose significance.

Part D.II. discusses US disparate treatment and EU direct discrimination laws. They handle decisions which are made because of protected attributes. D.II.2. finds that, under both jurisdictions, the law does not require a protected characteristic to be the sole building block of a decision but proceeds if it is one of several building blocks. Furthermore, the law does not accept a defense if the lender claims he discriminated for a business reason such as actuarial statistics, rather than "because of" a protected characteristic.

Part D.II.3. moved on to highlighting two hard cases which anti-discrimination law has traditionally faced and which will be compounded by algorithmic credit underwriting. The first hard case concerns drawing a line between protected and neutral characteristics and looks at attempts to circumvent the law by intentionally picking facially neutral attributes which correlate with a protected attribute. The second hard case implicates neutral attributes which imply a protected characteristic, for instance pregnancy implying sex. While both situations do not raise entirely novel concerns, the advent of AI models with their multitude of variables suggests that has been a fringe situation will become the new normal.

Against that background, part D.III. understands US disparate impact and EU indirect discrimination law as the main battleground for algorithmic discrimination. D.III.1. explores whether US disparate impact law extends to credit underwriting, D.III.2. does the same for EU indirect discrimination law. The paper proceeds to establishing two test prongs of received doctrine. D.III.3. summarizes both jurisdiction's understanding of facially neutral building blocks and D.III.4. takes a brief look at justificatory reasons.

D.III.5. investigates the third hard case which involves identifying the neutral building block in algorithmic credit underwriting. The more variables an algorithm employs, the more difficult this exercise becomes. D.III.5.a. submits that eliminating a facially neutral building block will change the outcome only for limited-input models. For sophisticated algorithms, due to redundant encoding and to the flexibility of multivariate regressions, the model will use stand-in proxies to arrive at the same prediction. This makes the received test unfit to cope with many algorithmic models. D.III.5.b. moves on to explore whether a solution lies in understanding the entire AI model as the building block underlying the decision. For lack of a counterfactual, this strategy is found unhelpful for fully automated models and for human credit officers which rely primarily on the algorithmic recommendation. In these situations, removing the entire model means there are no other factors left which cause the decision. As a last resort, the paper discusses dispensing with the requirement of building blocks altogether if an algorithmic decision is at stake and brings out important concerns with this strategy. Against this background, D.III.c. investigates a US CFPB no-action letter which went beyond received anti-discrimination doctrine by focusing on discriminatory output only. While mentioning downsides of this approach, the paper accepts one element of this strategy, namely the focus on quality and governance of algorithmic decision-making.

Against this background, part E. outlines preliminary contours of a regulatory design for consumer credit in the age of AI. E.I. proposes quality control as one element. This includes technical and governance controls, both as to data and model. Flaws can hurt the borrower, if he is rejected or overpays, and the lender, if he leaves money on the table by refusing a loan which would have been attractive. The paper explores how the EU AI Act has proposed avenues towards quality control of this type.

Part E.II. discusses consumer rights to rectify incorrect data under US and EU law. Given that information gathered from sources such as social media networks are prone to mistakes and misunderstandings, these rights will gain in prominence. Current laws provide consumers with various degrees of rights to access their data, ask for rectification and erasure. E.II.1. presents the US FCRA which focuses on the underwriting context. E.II.2. looks at the EU GDPR which works with omnibus rules. The paper submits that both sets of laws need adjustment to the age of AI credit underwriting, especially with an eye on efficient enforcement.

E.III. explores consumer rights to be informed about scoring and about the reasons for a denial of credit. E.III.1. introduces the US ECOA notice, E.III.2. contrasts this with the EU GDPR and the ongoing reform of the EU Consumer Credit Directive. Part E.IV. offers concluding thoughts on personalized pricing in credit underwriting, a concern which remains outside the scope of this paper.

References

Adams-Prassl, Jeremias/Binns, Reuben/Kelly-Lyth, Aislinn (2022): Directly Discriminatory Algorithms, Modern Law Review, Vol. 85, preprint available at https://onlinelibrary.wiley.com/doi/full/10.1111/1468-2230.12759 (last access 27 October 2022).

Agarwal, Sumit/Shashwat, Alok/Ghosh, Pulak/Gupta, Sudip (2021): Financial Inclusion and Alternate Credit Scoring: Role of Big Data and Machine Learning in Fintech, Indian School of Business, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3507827 (last access 27 October 2022).

Aggarwal, Nikita (2021): The Norms of Algorithmic Credit Scoring, The Cambridge Law Journal, Vol. 80, pp. 42–73.

Al-Zuabi, Ibrahim Mousa/Jafar, Assef/Aljoumaa, Kadan (2019): Predicting customer's gender and age depending on mobile phone data, Journal of Big Data, available at https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0180-9 (last access 27 October 2022).

American Bankers Association (ABA) (2022): Petition for rulemaking defining larger participants of the aggregation services, available at https://www.aba.com/-/media/documents/letters-to-congress-and-regulators/petition-to-cfpb-for-larger-participant-rulemaking-080222.pdf?rev=c0674598829b40808a179b1f4942b591 (last access 27 October 2022).

Arrow, Kenneth J. (1971): Some models of Racial Discrimination in the Labor Market, available at https://apps.dtic.mil/sti/pdfs/AD0735068.pdf (last access 27 October 2022).

Awrey, Dan/Macey, Joshua (2022): The Promise and Perils of Open Finance, ECGI working paper No 632/2022, available at https://ecgi.global/sites/default/files/working_papers/documents/maceyawreyfinal.pdf (last access 27 October 2022).

Balyuk, Tetyana (2021): FinTech Lending and Bank Credit Access for Consumers, Rotman School of Management Working Paper No. 2802220, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2802220 (last access 27 October 2022).

Barocas, Solon/Selbst, Andrew D. (2016): Big Data's Disparate Impact, California Law Review, Vol. 104, pp. 671–732.

Bartlett, Robert/Morse, Adair/Stanton, Richard/Wallace, Nancy (2022): Consumer-Lending Discrimination in the FinTech Era, Journal of Financial Economics Vol. 143, pp. 30–56.

Becker, Gary S. (1957): The Economics of Discrimination, Chicago University Press.

Becker, Gary S. (1993): Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, Chicago University Press, available at https://doi.org/10.7208/chicago/9780226041223.001.0001 (last access 27 October 2022).

Benjamin, Ruha (2019): Race after Technology, Polity Press.

Berg, Tobias/Burg, Valentin/Gombović, Ana/Puri, Manju (2018): On the Rise of FinTechs – Credit Scoring using Digital Footprints, NBER Working Papers 24551, National Bureau of Economic

Research, Inc., available at https://www.nber.org/system/files/working_papers/w24551/w24551.pdf (last access 27 October 2022).

Berman, Mitchell N./Krishnamurthi, Guha (2021): Bostock was Bogus: Textualism, Pluralism, and Title VII, Notre Dame Law Review, Vol. 97, pp. 67–126.

Blattner, Laura/Nelson, Scott (2021): How Costly is Noise? Data and Disparities in Consumer Credit, available at https://www.researchgate.net/publication/351656623_How_Costly_is_Noise_Data_and_Disparities_in_Consumer_Credit (last access 27 October 2022).

Bordalo, Pedro/Coffman, Katherine/Gennaioli, Nicola/Shleifer, Andrei (2016): Stereotypes, Quarterly Journal of Economics, Vol. 131, pp. 1753–1794.

Brito, Dagobert/Hartley, Peter (1995): Consumer Rationality and Credit Cards, Journal of Political Economy, Vol. 103, pp. 400–433.

Bureau of Consumer Financial Protection (CFPB) (2017): Request for Information, Docket-nr. CFPB-2017-0005 (2017), available at https://files.consumerfinance.gov/f/documents/20170214_cfpb_Alt-Data-RFI.pdf (last access 27 October 2022).

Bureau of Consumer Financial Protection (CFPB) (2019): An update on credit access and the Bureau’s first No-Action Letter, available at https://www.consumerfinance.gov/about-us/blog/update-credit-access-and-no-action-letter/ (last access 27 October 2022).

Bureau of Consumer Financial Protection (CFPB) (2020a), available at https://files.consumerfinance.gov/f/documents/cfpb_upstart-network-inc_no-action-letter_2020-11.pdf (last access 27 October 2022)

Bureau of Consumer Financial Protection (CFPB) (2020b): Consumer Access to Financial Records, 85 FR 71003, pp. 71003–71011.

Burk, Dan L. (2021): Algorithmic Legal Metrics, Notre Dame Law Review, Vol. 96, pp. 1147–1203.

Burrell, Jenna/Fourcade, Marion (2021): The Society of Algorithms, Annual Review of Sociology, Vol. 47, pp. 213–237.

Campbell, Colin/Smith, Dale (2022): Distinguishing Between Direct and Indirect Discrimination, Modern Law Review, Vol. 85, preprint available at https://onlinelibrary.wiley.com/doi/full/10.1111/1468-2230.12760 (last access 27 October 2022).

Chander, Anupam (2017): The Racist Algorithm?, Michigan Law Review, Vol. 115, pp. 1023–1045.

Citron, Danielle K./Pasquale, Frank (2014): The Scored Society: Due Process for Automated Predictions, Washington Law Review, Vol. 89, pp. 1–32.

Dembroff, Robin/Kohler-Hausmann, Issa (2022): Supreme Confusion About Causality at the Supreme Court, City University of New York Law Review, Vol. 25, pp. 57–92.

DeYoung, Robert/Phillips, Ronnie J. (2006): Strategic Pricing of Payday Loans: Evidence from Colorado, 2000–2005, available at https://www.indstate.edu/business/sites/business.indstate.edu/files/Docs/2006-WP-05_Young-Phillips.pdf (last access 27 October 2022).

Di Maggio, Marco/Ratnadiwakara, Dimuthu/Carmichael, Don (2021): Invisible Primes: Fintech Lending with Alternative Data, Harvard Business School Working Paper No. 22-024, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3937438 (last access 27 October 2022).

Dobbie, Will/Liberman, Andres/Paravisini, Daniel/Pathania, Vikram (2019): Measuring Bias in Consumer Lending, Harvard Kennedy School Working Paper No. RWP19-029, available at https://www.hks.harvard.edu/publications/measuring-bias-consumer-lending (last access 27 October 2022).

Dzida, Boris/Groh, Naemi (2018): Diskriminierung nach dem AGG beim Einsatz von Algorithmen im Bewerbungsverfahren, Neue Juristische Wochenschrift, pp. 1917–1922.

Eidenmüller, Horst/Wagner, Gerhard (2021): Law by Algorithm, Tübingen.

European Data Protection Board/European Data Protection Supervisor (EDPB/EDPS) (2021): Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act), 18 June 2021, avalaible at https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf (last access 27 October 2022).

Eidelson, Benjamin (2022): Dimensional Disparate Treatment, Southern California Law Review (forthcoming), Vol. 95, pp. 785–855.

Ernst, Christian (2017): Algorithmische Entscheidungsfindung und personenbezogene Daten, Juristenzeitung, pp. 1026–1036.

Federal Trade Commission (FTC) (2011): 40 years of experience with the FCRA, 2011, available at https://www.ftc.gov/sites/default/files/documents/reports/40-years-experience-fair-credit-reporting-act-ftc-staff-report-summary-interpretations/110720fcrareport.pdf (last access 27 October 2022).

Federal Trade Commission (FTC) (2022): Advance notice of proposed rulemaking, Trade Regulation Rule on Commercial Surveillance and Data Security, 87 FR 51273, pp. 51273–51299.

Financial Stability Board (FSB) (2022), Fintech and Market Structure in the COVID-19 Pandemic, 21 March 2022, available at https://www.fsb.org/wp-content/uploads/P210322.pdf (last access 27 October 2022).

Fisher, Linda E. (2009): Target Marketing of Subprime Loans: Racialized Consumer Fraud & Reverse Redlining, Journal of Law and Policy, Vol. 18, pp. 121–155.

Fuster, Andreas/Goldsmith-Pinkham, Paul/Ramadorai, Tarun/Walther, Ansgar (2022): Predictably Unequal? The Effects of Machine Learning on Credit Markets, Journal of Finance, Vol. 77, pp. 5–47.

Gillis, Talia (2022): The Input Fallacy, Minnesota Law Review (forthcoming), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3571266 (last access 27 October 2022).

Graham, Jennifer (2021): Risk of discrimination in AI systems, Evaluating the effectiveness of current legal safeguards in tackling algorithmic discrimination, in: Lui/Ryder (ed.), FinTech, Artificial Intelligence and the Law, Routledge, pp. 211–229.

Grünberger, Michael (2013), Personale Gleichheit, Der Grundsatz der Gleichbehandlung im Zivilrecht, Nomos.

Grünberger, Michael (2017): Grundstrukturen (allgemeine Strukturmerkmale) von Gleichheitssätzen, in: Kempny/Reimer (ed.), Gleichheitssatzdogmatik heute, Tübingen Mohr Siebeck, pp. 5–34.

Grünberger, Michael (2021): Reformbedarf im AGG: Beweislastverteilung beim Einsatz von KI, Zeitschrift für Rechtspolitik, pp. 232–235.

Guseva, Alya/Rona-Tas, Akos (2001): Uncertainty, Risk, and Trust: Russian and American Credit Card Markets Compared, American Sociological Review, Vol. 66, pp. 623–646.

Hacker, Philipp (2018): Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law, Common Market Law Review, Vol. 55, pp. 1143–1185.

Hellman, Deborah (2020): Measuring Algorithmic Fairness, Virginia Law Review, Vol. 106, pp. 811–866.

Hoffmann, Jochen (2014): Die Feststellung mittelbarer Diskriminierungen, Archiv für die civilistische Praxis, Vol. 214, pp. 822–850.

Hurlin, Christophe/Pérignon, Christophe/Saurin, Sébastien (2021): The Fairness of Credit Scoring Models, HEC Paris Research Paper No. FIN-2021-1411, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3785882 (last access 27 October 2022).

Jarass, Hans C. (2021): Charta der Grundrechte der Europäischen Union, 4th ed., Beck.

Kaminski, Margot (2019): Binary Governance: Lessons from the GDPR's approach to Algorithmic Accountability, Southern California Law Review, Vol. 92, pp. 1529–1616.

Kim, Pauline (2022): AI and Inequality, in: Johnson, Kristin/Reyes, Carla (ed.), The Cambridge Handbook on Artificial Intelligence and the Law (forthcoming), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3938578.

Kim, Pauline/Hanson, Erika (2016): People Analytics and the Regulation of Information under the Fair Credit Reporting Act, St Louis University Law Journal, Vol. 61, pp. 17–34.

Kissinger, Henry/Schmidt, Eric/Huttenlocher, Daniel (2021): The Age of AI: And Our Human Future, Little Brown and Company.

Kiviat, Barbara (2019a): The Art of deciding with data: evidence from how employers translate credit reports into hiring decisions, Socio-Economic Review, Vol. 17, pp. 283–309.

Kiviat, Barbara (2019b): The Moral Limits of Predictive Practices: The Case of Credit-Based Insurance Scores, Socio-Economic Review, Vol. 84, pp. 1134–1158.

Koppelman, Andrew (2022): Bostock and Textualism: A Response to Berman and Krishnamurthi, Notre Dame Law Review Reflection (forthcoming), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3904246 (last access 27 October 2022).

Krishnamurti, Guha/Salib, Peter (2020): Bostock and Conceptual Causation, Yale Journal of Regulation, Notice & Comment, available at https://www.yalejreg.com/nc/bostock-and-conceptual-causation-by-guha-krishnamurthi-peter-salib/ (last access 27 October 2022).

Langenbucher, Katja (2020): Responsible A.I. credit scoring – a legal framework, European Business Law Review, Vol. 31, pp. 527–572.

Langenbucher, Katja (2022): AI credit scoring and evaluation of creditworthiness – a test case for the EU proposal for an AI Act, in ECB, Continuity and change – how the challenges of today prepare the ground for tomorrow, ECB Legal Conference 2021, pp. 362–386, available at https://www.ecb.europa.eu/pub/pdf/other/ecb.ecblegalconferenceproceedings202204~c2e5739756.en.pdf (last access 27 October 2022).

Langenbucher, Katja (ed.) (2022a): Europäisches Privat- und Wirtschaftsrecht, 5th ed., Nomos.

Langenbucher, Katja/Corcoran, Patrick (2022): Responsible AI Credit Scoring – A Lesson from Upstart.com, European Company and Financial Law Review, Vol. 5, pp. 141–179.

Lauer, Josh (2017): Creditworthy, Columbia University Press.

Ledford, Heidi (2019): Millions of black people affected by racial bias in health-care algorithms, Nature Portfolio, available at https://www.nature.com/articles/d41586-019-03228-6 (last access 27 October 2022).

Lucas, Robert (1976): Econometric policy evaluation: A critique, Carnegie-Rochester Conference Series on Public Policy, Vol. 1, pp. 19–46.

Mangold, Anna Katharina/Payandeh, Mehrdad (ed.) (2022): Handbuch Antidiskriminierungsrecht, Strukturen, Rechtsfiguren und Konzepte, Mohr Siebeck.

Mayson, Sandra (2019): Bias In, Bias Out, The Yale Law Journal, Vol. 128, pp. 2218–2300.

Mörsdorf, Oliver (2019): Europäisierung des Privatrechts durch die Hintertür? Einige Gedanken zum Einfluss der Grundrechte-Charta auf das nationale Privatrecht in der jüngeren Rechtsprechung des EuGH (Egenberger, Bauer, CCOO u.a.), Juristenzeitung, pp. 1066–1074.

National Consumer Law Clinic (NCLC) (2020): Written Statement for CFPB's Symposium on Consumer Access to Financial Records, available at https://www.nclc.org/resources/nclc-statement-for-cfpb-1033-symposium/ (last access 27 October 2022).

Obermeyer, Ziad/Powers, Brian/Vogeli, Christine/Mullainathan, Sendhil (2019): Dissecting racial bias in an algorithm used to manage the health of populations, Science, Vol. 336, pp. 447–453.

O'Keefe, Patrick (2016): Qualified Mortgages & Government Reverse Redlining: How the CFPB's Qualified Mortgage Regulations Will Handicap the Availability of Credit to Minority Borrowers, Fordham Journal of Corporate & Financial Law, Vol. 21, pp. 413–447.

O’Neil, Cathy (2016): Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy, Penguin Books.

Parlour, Christine/Rajan, Uday (2001): Competition in Loan Contracts, The American Economic Review, Vol. 91, pp. 1311–1328.

Phelps, Edmund (1972): The Statistical Theory of Racism and Sexism, The American Economic Review, Vol. 62, pp. 659–661.

Prince, Anya E.R./Schwarcz, Daniel (2020): Proxy Discrimination in the Age of Artificial Intelligence and Big Data, Iowa Law Review, Vol. 105, pp. 1257–1318.

Rebhahn, Robert/Kietaibl, Christoph (2010): Mittelbare Diskriminierung und Kausalität, Recht der Internationalen Wirtschaft, pp. 373–396.

Rosamond, Emily (2016): “All Data is Credit Data”: Reputation, Regulation and Character in the Entrepreneurial Imaginary, Paragrana, Vol. 25, No. 2, pp. 112–124.

Sacksofsky, Ute (2017): Was heißt: Ungleichbehandlung „wegen“?, in: Kempny/Reimer (ed.), Gleichheitssatzdogmatik heute, Mohr Siebeck, pp. 63–90.

Stiglitz, Joseph/Weiss, Andrew (1981): Credit Rationing in Markets with Imperfect Information, The American Economic Review, Vol. 71, pp. 393–410.

Student Borrower Protection Center (2020), Educational Redlining, February 2020, available at https://protectborrowers.org/wp-content/uploads/2020/02/Education-Redlining-Report.pdf (last access 27 October 2022).

Sunstein, Cass (2019): Algorithms, Correcting Biases, Social Research: An International Quarterly, Vol. 86, pp. 499–511.

Thüsing, Gregor (2021): Münchener Kommentar zum BGB, Vol 1, §§ 1–240 BGB, AllgPersönlR, ProstG, AGG, 9th. Ed., Beck.

Wachter, Sandra (2022): The Theory of Artificial Immutability: Protecting Algorithmic Groups under Anti-Discrimination Law, Tulane Law Review (forthcoming), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4099100 (last access 27 October 2022).

Wachter, Sandra/Mittelstadt, Brent/Russell, Chris (2021): Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI, Computer Law & Security Review, Vol. 41, available at: https://www.sciencedirect.com/science/article/abs/pii/S0267364921000406 (last access 27 October 2022).

Weigert, Daniel-René (2018): Die Kausalitätswahrscheinlichkeit – Das „vergessene“ Tatbestandsmerkmal des § 22 AGG, Neue Zeitschrift für Arbeitsrecht, pp. 1166–1171.

Zuiderveen Borgesius, Frederik (2020): Price Discrimination, Algorithmic Decision-Making, and European Non-Discrimination Law, European Business Law Review, Vol. 31, pp. 401–422.

european corporate governance institute

about ECGI

The European Corporate Governance Institute has been established to improve *corporate governance through fostering independent scientific research and related activities.*

The ECGI will produce and disseminate high quality research while remaining close to the concerns and interests of corporate, financial and public policy makers. It will draw on the expertise of scholars from numerous countries and bring together a critical mass of expertise and interest to bear on this important subject.

The views expressed in this working paper are those of the authors, not those of the ECGI or its members.

www.ecgi.global

european corporate governance institute

ECGI Working Paper Series in Law

Editorial Board

Editor	Amir Licht, Professor of Law, Radzyner Law School, Interdisciplinary Center Herzliya
Consulting Editors	Hse-Yu Iris Chiu, Professor of Corporate Law and Financial Regulation, University College London
	Martin Gelter, Professor of Law, Fordham University School of Law
	Geneviève Helleringer, Professor of Law, ESSEC Business School and Oxford Law Faculty
	Kathryn Judge, Professor of Law, Coumbia Law School
	Wolf-Georg Ringe, Professor of Law & Finance, University of Hamburg
Editorial Assistant	Asif Malik, ECGI Working Paper Series Manager

https://ecgi.global/content/working-papers

european corporate governance institute

Electronic Access to the Working Paper Series

The full set of ECGI working papers can be accessed through the Institute's Web-site (https://ecgi.global/content/working-papers) or SSRN:

Finance Paper Series	http://www.ssrn.com/link/ECGI-Fin.html
Law Paper Series	http://www.ssrn.com/link/ECGI-Law.html

https://ecgi.global/content/working-papers