The surprising Impotence of Anti-Discrimination Law in the Age of AI
... and a comment on Art. 6 Directive 2023/2225 on Credit Agreements for Consumers

Katja Langenbucher

This paper takes up the EU's new rule on discriminatory credit underwriting. I build on earlier work, exploring how anti-discrimination law fares when fitting algorithmic credit scoring and creditworthiness evaluation into the regime of EU direct/indirect discrimination and US disparate treatment/disparate impact doctrine.[1] *I suggest that anti-discrimination law, when faced with redundant encoding, runs into doctrinal and practical problems. These manifest in proving but-for causation, in scenarios resembling algorithmic redlining, and in a more fundamental misfit between anti-discrimination law and big data analytics. The paper summarizes these challenges, addresses recent changes in US law, and submits that the new EU Directive has missed the chance to provide for a modern rule, fit to cope with algorithmic credit underwriting.*

A. Introduction

Credit scoring and creditworthiness evaluation provide excellent examples for both, the inclusive power of algorithms and the risk of algorithmic discrimination. Any decision to hand out a loan and price interest rates includes an assessment of the borrower's credit risk. Naturally, this involves a distinction among applicants to make an informed choice. Advanced

1 Parts of the following text are based on Katja Langenbucher, 'Consumer Credit in the Age of AI – Beyond Anti-Discrimination law' (2022) ecgi working paper 663/2022. In that (much longer) paper, written prior to new EU rule, I explore some of the arguments put forward here in more depth. In the following text, I occasionally use sentences and whole paragraphs with identical wording to my earlier paper.

statistics has been a classic tool for that task.[2] Today, big data and machine learning algorithms promise a disruption in how creditworthiness is evaluated.[3] The popular remark "All data is credit data. We just don't know how to use it yet"[4] suggests that leveraging a loan applicant's digital footprint has considerable potential to produce more accurate credit risk assessments. Online payment history, performance on lending platforms, age or sex, job or college education, ZIP code, income or ethnic background can all have predictive force. Additionally, consider preferred shopping places, social media friends, political party affiliations, 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. All these "input variables" include information about the applicant that is potentially relevant for computing his creditworthiness. Many of these are *redundantly encoded* in several data points: The fact that the applicant is female might be encoded in her preferred shopping places, her social media friends, her first name or her college. This paper submits that redundant encoding of this type causes a fundamental problem to received anti-discrimination law.

At first glance, the availability of big, so-called "alternative data", coupled with AI-based scoring promises to enhance access to credit markets. Mostly, this is due to lower search costs for lenders.[5] It is important to remember that lenders have long been aware that a low credit score is not necessarily an accurate reflection of an applicant's creditworthiness. However, in consumer credit markets it has not been cost-efficient for the lender to invest in locating "invisible prime"[6] applicants: While many applicants will be good

2 Josh Lauer, *Creditworthy, A History of Consumer Surveillance and Financial Identity in America* (CUP 2017) 200.

3 J Burrell and M Fourcade, 'The Society of Algorithms' (2021) Annual Review of Sociology Vol. 47 213, 222; D Citron and F Pasquale, 'The Scored Society": Due Process for Automated Predictions' (2014) Washington Law Review, Vol. 89 1, 4.

4 Quentin Hardy, 'Just the Facts. Yes, All of Them.' New York Times (New York, 25 March 2012) discussion at Emily Rosamond, '"All Data is Credit Data": Reputation, Regulation and Character in the Entrepreneurial Imaginary' (2016) Paragrana, Vol. 25, No. 2 112.

5 Katja Langenbucher, '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' (2022) ECB Legal Conference 2021 362.

6 Term proposed by M Di Maggio and D Ratnadiwakara, 'Invisible Primes: Fintech Lending with Alternative Data' (2021) 1 (2) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3937438> accessed 24 June 2025.

credit risk and present an attractive business case, search costs to identify them may far outweigh the expected return.[7] Algorithmic scoring models have changed that equation and, in that sense, promise inclusion.

At the same time, 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.[8] If the training data reflect past inequality, 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*).The fact that training data are, in this way, shaped by history has direct implications for how the AI builds its model.[9] Variables that the AI finds for most candidates who were successful in the past will be accorded most weight, for instance a specific job, sex or race. Candidates whose profile does not include the relevant positive variable will face a risk premium (*majority bias*).[10] The same logic applies to variables that send a negative signal. The AI learns from historical data and singles out variables that 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

7 Lauer (n 2) 210.

8 Deborah Hellman, 'Measuring Algorithmic Fairness' (2020) Virginia Law Review Vol. 106 811, 841; Langenbucher, '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' (n 5) 372 et seq.; Sandra Mayson, 'Bias In, Bias Out' (2019) The Yale Law Journal Vol. 128 2218, 2251: "The premise of prediction is that, absent intervention, history will repeat itself".

9 L Blattner and S Nelson, 'How Costly is Noise? Data and Disparities in Consumer Credit' (2021) 1 (12), "model bias" <https://www.researchgate.net/publication/351656623_How_Costly_is_Noise_Data_and_Disparities_in_Consumer_Credit> accessed 24 June 2025.

10 S Barocas and A Selbst, 'Big Data's Disparate Impact' (2016), California Law Review, Vol. 104 671, 689; Talia Gillis, 'The Input Fallacy' (2022), Minnesota Law Review, Vol. 106 1175; Jennifer Graham, Risk of discrimination in AI systems, Evaluating the effectiveness of current legal safeguards in tackling algorithmic discrimination in Alison Lui, Nicholas Ryder (eds), *FinTech, Artificial Intelligence and the Law* (2021), 211, 211; Katja Langenbucher, 'Responsible A.I. credit scoring – a legal framework' (2020), European Business Law Review, Vol. 31 527; Dan L Burk,: Algorithmic Legal Metrics (2021), Notre Dame Law Review, Vol. 96 1147, 1163; Antje von Ungern-Sternberg, 'Diskriminierungsschutz bei algorithmenbasierten Entscheidungen' in Anna Katharina Mangold and Mehrdad Payandeh, *Handbuch Antidiskriminierungsrecht* (Mohr Siebeck 2022) § 28, note 15 et seq.

true if the observed risk-variable is less informative for some applicants if compared with others.[11]

Algorithmic biases of that type matter even more when combined with concerns about data quality, incorrect labelling,[12] 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.[13] Additionally, the use of certain alternative data, for instance stemming from social networks, increases the risk of inaccuracies. This concerns individual loan applicants if the data used to evaluate them is inaccurate. It also impacts the entire AI model that learns from (partially) inaccurate training data. The more inaccuracies are hidden in big datasets, the more the AI model is shaped by a world that does not even adequately reflect yesterday's world, much less today's.

In what follows, I briefly summarize the US and EU legal framework for anti-discrimination (below II). I move on to highlight core challenges this framework faces when dealing with redundant encoding (below III.). The paper closes with a comment on Article 6 of the new EU Directive on Credit agreements for Consumers (below IV).

B. The Legal Framework

The observation that credit scoring sometimes produces unfair results is neither a novel concern nor a worry that is specific to AI-based underwriting. Traditional scoring models with their limited number of input variables necessarily provide a crude picture of an individual applicant.[14] Many are shaped by path-dependent historical choices of what is deemed relevant for a score. Lenders enjoy discretion to strike a balance between predictive accuracy, costs for model and data, and market expansion.[15] Furthermore, financial stability concerns provide a reason to err on the side

11 European Data Protection Board/European Data Protection Supervisor (EDPB/EDPS) (2021): Joint Opinion 5/2021; Gillis (n 11) 1178; Burk (n 11) 1164.

12 Von Ungern-Sternberg (n 11) § 28 note 16 (sampling bias), note 17 (labelling bias), note 18 (feature selection bias).

13 Mayson (n 9).

14 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.

15 Burrell and Fourcade (n 3) 217 et seq.

of caution, rather than hand out a loan to an applicant with a high credit default risk.

I. The US legal framework

In the US, anti-discrimination law has been navigating this space since the late 1960s. Discriminatory lending is addressed by the Equal Credit Opportunity Act (ECOA) and the Fair Housing Act (FHA). Both, the ECOA and the FHA prohibit decisions that are – loosely speaking – caused or motivated by a protected characteristic such as race, gender, age, or similar protected attributes. The FHA prohibits discrimination in the context of mortgages, the ECOA concerns access to credit more generally. The Trump Administration's Executive Order on "Restoring Equality of Opportunity and Meritocracy" of April 23, 205 has stipulated that agencies roll back disparate impact doctrine for both scenarios.[16]

1. Disparate Treatment

Disparate treatment involves a lender who denies credit "because of" an applicant's protected characteristic. Key questions have to do with discriminatory intent, with (conscious or subconscious) motives, and with burden of proof. In the US, many of the finer doctrinal details of anti-discrimination law have not been developed for credit underwriting, but in Title VII cases, addressing employment discrimination.

In the context of this paper, it is of particular interest to understand how courts have established discriminatory motives, which often rest on circumstantial evidence. In *McDonnell Douglas*, a Title VII case[17] that was partly overruled by *Ames v. Ohio Dept. of Youth Services*,[18] the US Supreme Court established a three-step strategy for individual inferential proof.

First, a plaintiff bears the initial burden of establishing a *prima facie* case by producing enough evidence to support an inference of discriminatory motive.[19] He might, for instance, show that he is a member of a protected group, was qualified for a position, was rejected by the potential employer,

16 See below B.I.2.
17 *McDonnell Douglas Corp. v. Green*, 411 U.S. 792 (1973).
18 *Ames v. Ohio Dept. Of Youth Services*, 605 U.S. ___ (2025).
19 *McDonnell Douglas Corp. v. Green*, 411 U.S. 792 (1973) 802.

and the position remained open, suggesting discriminatory motives. In *Ames*, the Court clarified that members of a majority group cannot be asked to satisfy a heightened evidentiary standard than members of a minority group.[20] Additionally, Justice Thomas, concurring, invited the Court to, in the future, consider "whether the *McDonnell Douglas* framework is an appropriate tool to evaluate Titel VII claims at summary judgment".[21]

Second, if the plaintiff succeeds, the burden shifts to the defendant to articulate a legitimate, nondiscriminatory reason.[22] She does not have to prove that the reason she advances did in fact drive her decision. Instead, it is only a burden of production.

Third, the plaintiff must have a fair opportunity to show that the reasons the defendant has proffered are pretextual.[23] In *Ames*, Justice Thomas, concurring, more generally criticized the criteria set forth in *McDonnell Douglas*, submitting that they demand more from the plaintiff than the text of Title VII.[24] *McDonnell Douglas* requires a plaintiff to prove that the justificatory reasons the defendant offered were but a pretext. By contrast under Title VII it suffices to prove that a protected characteristic as a motivating factor, even though other factors also motivated the practice.[25]

Another core aspect of how US law approaches burden of proof concerns situations where it is not in doubt that a discriminatory element contributed to the decision, but the defendant disputes causation. In *Manhart*, the Supreme Court held that an employer's policy of requiring women to make larger pension fund contributions than men violated Title VII. There was no doubt that an unlawful factor was at play, given that the policy

20 *Ames v. Ohio Dept. Of Youth Services*, 605 U.S. ___ (2025) 9.

21 *McDonnell Douglas Corp. v. Green*, 411 U.S. 792 (1973), Justice Thomas, concurring 7.

22 *McDonnell Douglas Corp. v. Green*, 411 U.S. 792 (1973) 802; For a similar strategy in Germany see Ute Sacksofsky, 'Was heißt: Ungleichbehandlung „wegen"?' (Mohr Siebeck 2017) Simon Kempny and Philipp Reimer (eds), Gleichheitssatzdogmatik heute 63, 73.

23 *McDonnell Douglas Corp. v. Green*, 411 U.S. 792 (1973) 804.

24 *Ames v. Ohio Dept. Of Youth Services*, 605 U.S. ___ (2025), Justice Thomas 11.

25 If the plaintiff fails to show individual inferential proof, the Court has in Title VII cases accepted a showing of intentional discrimination through circumstantial evidence (group or systemic inferential proof). Plaintiffs use statistics to prove a "pattern and practice" revealing that their group is underrepresented. "Such imbalance", the Court held in *International Brotherhood of Teamsters v. United States*, 431 U.S. 324, 339 (1977), "is often a telltale sign of purposeful discrimination". Defendants may bring in different statistics or put forward a legitimate non-discriminatory explanation for the underrepresentation of the plaintiff's group.

specifically targeted women. Still, the employer had argued that he had no *discriminatory* intent and did not treat women differently *because of* their sex. Rather, he suggested, actuarial logic dictated a "life-expectancy adjustment".[26] It is a claim any economist would have embraced, pointing to the logic of statistical discrimination. Arguably, in a credit underwriting context, a similar logic applies. In the same way as sex influences life expectancy, hence, is relevant for pension contributions, sex will often influence creditworthiness. Still, in *Manhart*, the US Supreme Court did not follow the defendant's statistical defense. Instead, it stressed that the impermissible attribute (sex) was a but-for factor for the employer's decision. Removing the attribute "female", so the Court held, would have led to a different, non-discriminatory outcome. In a similar case, the European Court of Justice also rejected the claim of insurance companies that had argued statistics and actuarial logic required higher fees for women.[27]

2. Disparate Impact

The important challenges that a disparate treatment plaintiff faces when providing discriminatory evidence have encouraged the development of a second line of anti-discrimination doctrine. This doctrine focuses on facially neutral variables or practices. If a neutral attribute or practice consistently triggers less favorable treatment of protected communities, this makes it "suspicious", as it were. Possibly, one line of argument goes, a decision-maker has found an (only seemingly) neutral attribute or practice to hide his true discriminatory motivations. In the words of the US Supreme Court, disparate impact doctrine works as "an evidentiary tool used to identify genuine, intentional discrimination – to 'smoke out,' as it were, disparate treatment".[28]

26 *City of Los Angeles v. Manhart*, 435 U.S. 702 (1978); the ECJ followed the same logic Case C-54/07 *Feryn* [2008] ECJ; Sacksofsky (n 23) 73.

27 Case C- 236/09 *Association Belge des Consommateurs Test-Achats and others* [2011] ECJ.

28 *Ricci v. DeStefano*, 557 U.S. 557 (2009); discussed at Gillis (n 11) 1200; overview at Langenbucher, 'Responsible A.I. credit scoring – a legal framework' (n 11) 554; on EU law's trajectory from a formal to a more substantive approach of indirect discrimination doctrine see R Rebhahn and C Kietaibl, 'Mittelbare Diskriminierung und Kausalität, Recht der Internationalen Wirtschaft' (2010) Rechtswissenschaft: Zeitschrift für rechtswissenschaftliche Forschung 373, 384 et seq.; further Ute Sacksofsky, 'Unmittelbare und mittelbare Diskriminierung' in Anna Katharina Mangold

Disparate impact doctrine has not stopped there. In many cases, the doctrine has been understood as going beyond a mechanism that only serves to uncover hidden disparate treatment. Especially when faced with the government discriminating against a private citizen, most courts and scholars have so far followed some version of a "substantive" approach.

Those who follow a substantive approach understand disparate impact doctrine not as a tool to unearth covert motives, hidden behind a facially neutral attribute. Instead, they investigate whether available legislation – such as the FHA – prohibits discriminatory consequences, irrespective of motives.[29] If this is the case, the fact that the net result of a defendant's decisions consistently plays out worse for a minority group if compared to the majority *prima facie* can trigger a prohibition, even if the defendant had no discriminatory intent or motive. A defendant must then put forward a business defense and show that there was no less discriminatory strategy available.

The Executive Order of April 23, 2025 represents a move away from disparate impact doctrine. Its section 1 stresses that equality under the US Constitution refers to "equality of opportunity, not equal outcomes". Section 4 requests agencies to deprioritize enforcement based on disparate impact liability. Section 6 requests the Attorney General, the Secretary of Housing and Urban Development, the Director of the Consumer Financial Protection Bureau, the Chair of the Federal Trade Commission and other relevant agency heads to "evaluate all pending proceedings that rely on theories of disparate impact liability" as far as the Equal Credit Opportunity Act and the Fair Housing Act are concerned.

---

and Mehrdad Payandeh (eds), *Handbuch Antidiskriminierungsrecht* (Mohr Siebeck 2022), § 14 note 105; von Ungern-Sternberg (n 10) § 28 note 91; for EU law see A Mangold and M Payandeh, 'Antidiskriminierungsrecht – Konturen eines Rechtsgebiets' in Anna Katharina Mangold and Mehrdad Payandeh (eds.), *Handbuch Antidiskriminierungsrecht* (Mohr Siebeck 2022), § 1 note 109, listing the prohibition to circumvent anti-discrimination law as well as shifting the burden of proof.

29 *Griggs v. Duke Power Co.*, 401 US 424 (1971) 432; *Smith v. City of Jackson*, 544 U.S. 228 (2005) 236; *Texas Department of Housing and Community Affairs v. Inclusive Communities Project Inc.* 135 S. Ct. 2507 (2015) 2522; discussed at Gillis (n 10); see further Mayson (n 9); Cass R Sunstein,: Algorithms, Correcting Biases (2019) Social Research: An International Quarterly, Vol. 86 499; Rebhahn and Kietaibl (n 28) 389.

3. Discriminatory credit underwriting

While there is US Supreme Court guidance as to disparate impact doctrine under the FHA, [30] it has been unclear whether disparate impact doctrine extends to access to retail credit. In *Inclusive Communities* the Court held that "disparate impact claims are cognizable under the Fair Housing Act (…)", the reasoning stressing that its "text refers to the consequences of the actions". The ECOA, by contrast, lacks a results-oriented language of this type. While the FDIC and the Fed have in the past seemed generally open to considering disparate impact in their supervisory activities, both agencies have stressed that "the fact that a policy or practice creates a disparity on a prohibited basis is not alone a proof of a violation." They require an agency that finds a lender's practice to have a disparate impact to determine whether it is justified by a manifest business necessity and whether there was an alternative practice serving the same purpose with less discriminatory results. [31] It is an open question whether the Fed and the FDIC will change course after the Executive Order of April 23, 2025. The same goes for courts.[32]

For the plaintiff, disparate impact cases often turn on the relevant standard of proof. A disparate impact case has so far required a *prima facie* showing of an outcome that is disproportionate for a protected group. The Executive Order of April 23, 2025 is critical of this practice, understanding it as an "insurmountable presumption of unlawful discrimination".

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.[33] Defendants can, among other defenses, deny that there was a disproportionate outcome by questioning group membership.

30 *Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc.*, 135 S. Ct. 2507 (2015) 2519.

31 FDIC(Federal Deposit Insurance Corporation), 'Consumer Compliance Examination Manual' – December 2024 <https://www.fdic.gov/resources/supervision-and-examinations/consumer-compliance-examination-manual/documents/4/iv-1-1.pdf> accessed 24 June 2025; Federal Fair Lending Regulations and Statutes Overview <https://www.federalreserve.gov/boarddocs/supmanual/cch/fair_lend_over.pdf> accessed 24 June 2025.

32 *Ramirez v. Greenpoint Mortgage Funding, Inc.*, 633 F. Supp. 2d 922 (2008) 926 et seq.; Gillis (n 11) 1198.

33 See Pauline Kim, 'AI and Inequality' (2021) Washington University in St. Louis Legal Studies Research Paper No. 21-09-03 on difficulties in practice to collect data about outcomes across the applicant pool.

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.[34] 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 that serves as benchmark for a disparate impact comparison, the more difficult for a plaintiff to establish a case.

If private parties are litigating, the move towards substantive anti-discrimination theories is considerably less pronounced than if the government is involved. Most start from the ground rule that private parties enjoy free contracting choices. Against this background, disparate impact on protected groups comes across as an unwelcome, but usually legal, side effect. Various business necessity defenses are available to justify the practice despite the disproportionate output.[35] The most natural defense for a lender is that he is required to carefully assess credit default risk. Statistical evaluation and scoring procedures have developed as a sensible and legitimate tool over the last century. AI-based scoring will provide another tool. The burden then shifts back to the plaintiff to show that there was a less discriminatory way to achieve that same goal.

II. The EU framework

EU law has largely followed similar doctrinal patterns as the US. However, courts have interpreted anti-discrimination rules more broadly in two ways. Firstly, a rule that prohibits direct discrimination has so far been read as prohibiting indirect discrimination as well. Secondly, the courts have

34 Langenbucher, 'Consumer Credit in the Age of AI – Beyond Anti-Discrimination law' (n 1) 28.

35 Noting that there is little guidance on this question under US law: Gillis (n 11) 1249; critical on the vagueness of the (English) concept Sacksofsky (n 30) § 14 note 45, 129 et seq.

often been open to applying anti-discrimination rules not only between government and citizen, but also between private citizens.

Art. 51 para. 1 s. 1, 21 of the Charter of Fundamental Rights protect against discriminatory treatment based on 15 enumerated attributes or characteristics. While the text of the Charter is silent as to bringing it to bear on private law relations between citizens, the ECJ has in many situations interpreted it along those lines.[36] In the area of employment, Art. 157 TFEU's guarantee of equal pay between the sexes has since the 1970s been understood to prohibit discrimination between private parties.[37]

Various EU directives address specific relationships between private citizens. Some focus on employment,[38] two directives concern access to publicly available goods or services.[39] All of these explicitly cover both, direct and indirect discrimination. In these directives, indirect discrimination is understood as a situation "where an apparently neutral provision, criterion or practice would put persons "who share a protected characteristic" 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".[40]

These directives leave details of private enforcement and litigation to Member States' national law, as long as Member State law respects *effet*

36 Overview at Andrea Edenharter, 'Wie argumentieren EuGH und BVerfG in Grundrechtsfragen?' (2022), EuR 2022 302; Oliver Mörsdorf, 'Europäisierung des Privatrechts durch die Hintertür? Einige Gedanken zum Einfluss der Grundrechte-Charta auf das nationale Privatrecht in der jüngeren Rechtsprechung des EuGH' (2019) JZ Juristenzeitung Vol. 74 Issue 22 1066.

37 Case C - 149/77 *Defrenne / Sabena* [1978] 130 ECJ; Case C-13/94 - *P / S and Cornwall County Council* [1996] 170 ECJ; Case C-144/04 - *Mangold* [2005] 709 ECJ; Case C-555/07 - *Kücükdeveci* [2010] 21 ECJ; Case C-414/16 - *Egenberger* [2018] 257 ECJ; Case C-684/16 - *Max-Planck-Gesellschaft zur Förderung der Wissenschaften* [2018] 874 ECJ; Case C-193/17 - *Cresco Investigation* [2019] 43 ECJ; Brief overview at P Donath and D Schrader, 'Arbeitsrecht' in Katja Langenbucher (ed), *Europäisches Privat- und Wirtschaftsrecht* (Nomos 2022), § 7 note 35.

38 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).

39 Directive 2000/43/EC implementing the principle of equal treatment between persons irrespective of racial or ethnic origin; Directive 2004/113/EC implementing the principle of equal treatment between men and women in the access to and supply of goods and services.

40 See for instance Art. 2 para. 2 lit. b Directive 2000/43/EC implementing the principle of equal treatment between persons irrespective of racial or ethnic origin.

utile. One example of *effet utile* reigning in on Member State's discretion concerns burden of proof. The European Court of Justice (ECJ) has, in the context of employment discrimination, required Member States to adjust their rules on burden of proof to guarantee a minimum standard of enforceability.[41] Similarly to US law, 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".[42] A classic argument for the defense, like under US law, concerns group membership. The defendant may show that the plaintiff is not similarly situated to the set of persons that are treated more favorably.[43]

Prior to October 2023, there was no explicit directive addressing discriminatory credit underwriting under EU law, leaving the issue to Member State law. While several directives have dealt with instances of discrimination between private citizens, credit underwriting fell outside of their scope. Two directives had to do with access to publicly available goods or services.[44] However, the relevance of personal attributes in a credit underwriting context excluded an understanding of loan contracts as a publicly available good.[45]

Art. 6 of Directive 2023/2225 on Credit Agreements for Consumers changes this.[46] The rule protects consumers who are legally resident in

41 See Case C-109/88 - *Handels- og Kontorfunktionærernes Forbund i Danmark / Dansk Arbejdsgiverforening, agissant pour Danfoss* [1989] 383.

42 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.

43 Case T-473/12 - Aer Lingus v Commission [2015] 473 ECJ; Case C-356/09 - Kleist [2010] 703; Case C-366/99 - Griesmar [2001] 648 ECJ discussing "comparable situations".

44 Directive 2000/43/EC implementing the principle of equal treatment between persons irrespective of racial or ethnic origin; Directive 2004/113/EC implementing the principle of equal treatment between men and women in the access to and supply of goods and services.

45 For examples see F Rödl and A Leidinger, 'Diskriminierungsschutz im Zivilrechtsverkehr' in Anna Katharina Mangold and Mehrdad Payandeh (eds.), *Handbuch Antidiskriminierungsrecht* (Mohr Siebeck 2022), § 22 note 42 et seq.

46 Available at <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202302225> accessed 24 June 2025.

the EU against discrimination “on ground of their nationality or place of residence or any other ground as referred to in Article 21 of the Charter of Fundamental Rights of the European Union”. The drafting of the new Directive, if compared to other anti-discrimination directives, includes a surprising feature. All EU anti-discrimination directives have so far explicitly defined and prohibited both, direct and indirect discrimination. By contrast, Directive 2023/2225 follows a strategy that is more common in human rights texts such as Art. 21 of the Charter. It mentions protected attributes and prohibits discrimination “on grounds of” these.[47] Departing from earlier directives, Directive 2023/2225 neither defines direct and indirect discrimination, nor explicitly proscribes the latter.[48]

C. The Shortcomings of Received Doctrine

I. The Metaphor of Building Blocks

So far, we have seen that anti-discrimination law is triggered by input to a decision-making process, namely attributes or characteristics of a person.[49] I refer to these as “building blocks” of a decision. There are outright prohibited building blocks, and facially neutral ones. Direct discrimination/disparate treatment prohibits the use of certain building blocks, for instance race or sex, even if they are of direct empirical relevance. By contrast, facially neutral building blocks can be lawfully used unless they trigger indirect discrimination/disparate impact. A paradigm case is the 1970s US Supreme Court decision in *Griggs*. It illustrates how a specific practice caused disparate impact across racial groups. An employer had used the score in an intelligence test as decisive for the position as a manual laborer. This practice statistically discriminated against minority

47 See Sacksofsky (n 28) § 14 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, not a human rights text) falls in that second category.

48 See below D.I.

49 On input see M Berman and G Krishnamurthi, ‘Bostock was Bogus: Textualism, Pluralism, and Title VII’ (2021) Notre Dame Law Review Vol. 97 67, 98; Andrew Koppelman,: ‘Bostock and Textualism: A Response to Berman and Krishnamurthi’ (2022) Notre Dame Law Review Reflection Vol. 98 89, 98 (the latter criticizing the former, but in agreement about this basic point).

employees, however there was no evidence of discriminatory intent on the side of the defendant.[50] In *Griggs*, 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". Hence, 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, "the Act proscribes not only overt discrimination, but also practices that are fair in form, but discriminatory in operation."[51] To establish this, the plaintiff successfully made two showings: the disparate outcome across black and non-black job applicants and the identification of the intelligence test as one necessary building block of the employer's decision.

II. Building blocks and redundant encoding

This paper puts a spotlight on the role of "building blocks" such as the IQ test in *Griggs*. Metaphorically speaking, traditional doctrine requires one building block with discriminatory potential to be involved in the decision. The law's role is, first, to carefully examine such building blocks and to determine whether the decision would have looked differently if the unlawful building block was removed: Would the same employees have been hired if the IQ test was not run? A second legal requirement, the availability of justificatory defenses for using the relevant building block, is beyond the scope of this paper.

With improving technology, the core notion of anti-discrimination law to require a specific building block faces a novel challenge.[52] Big data furnishes a universe of different data points. Machine learning algorithms unearth innumerable correlations between those data points. Depending on the AI model used, the lender may be unable to identify salient data points, their weight, or core correlations. For the law this translates into a tricky problem: In many cases, the building block, such as the IQ test, will influence the outcome, but be unknown to the decision-maker. The same

50 Today, following a Title VII amendment of 1991, the law explicitly prohibits employment tests that are not a reasonable measure of job performance.

51 *Griggs v. Duke Power Co.*, 401 US 424 (1971), p. 430 et seq.

52 A Fuster et al., 'Predictably Unequal? The Effects of Machine Learning on Credit Markets' (2022) Journal of Finance Vol. 77 5, 8; along similar lines: Gillis (n 11).

goes for the plaintiff. He can show a statistically disparate outcome, but not the individual building blocks that might have triggered it.

Sometimes, work-arounds for the problem exist. The first is using an explainable AI (or simple regression models)[53] in combination with a limited number of input data points to train the model. However, limiting input data for training purposes seriously compromises on predictive accuracy, foregoing the added predictive force that alternative data offers.[54]

A second work-around might be offered by rules such as Art. 18 para. 3 EU Consumer Credit Directive. The rule requires lenders to use "relevant" information which is "necessary and proportionate to the nature, duration, value and risks of the credit for the consumer" and excluding protected categories of data. The Directive lists evidence on income, financial assets and liabilities, or information on other financial commitments as examples.[55] In that way, the Directive excludes traditionally offensive building blocks like the IQ test in *Griggs*. However, it is unclear whether this also provides a satisfying solution if the applicant delivers "relevant", small data only, but the AI is trained on big, alternative data. Due to redundant encoding, sophisticated models will sort applicants based on learned patterns, even if individual applicants only deliver limited data points.

A third work-around are explainable AI methods that work with counterfactuals such as DiCE.[56] The lender might run the model multiple times, and at each step change one data point, for instance sex, or race, or religious belief of the applicant. If changing one data point triggers a different outcome, the lender will assume that the removed building block has a meaningful impact on the decision. Note, however, that this strategy only

53 The German credit scoring company SCHUFA has decided to use such regression models, not more sophisticated AI models.

54 Additionally, the explanation the AI delivers will not necessarily help. Especially if prohibited characteristics, such as race or sex, are not included in the AI's training data, its explanation will, at best, produce a first step towards evaluating individual loan decisions. Take, for instance, an explainable AI that tells us core data points were first name and height. This explanation might raise the suspicion that (indirect) sex discrimination is going on. However, to confidently say so, we would need to establish a necessary correlation between sex and those variables.

55 For more detail see Annex 2 of EBA Guidelines on loan origination and monitoring, p. 71 et seq.

56 On DiCE and other methods of explainable AI see: Katja Langenbucher, 'Explainable AI as a Component of Building Trust The Case of Regulating Creditscoring' in S Bücker et al. (eds), *Digital Governance* (2025) (forthcoming).

works smoothly if the data points do not correlate, hence, that there is no redundant encoding involved.

Against this background, redundant encoding poses a fundamental challenge to received anti-discrimination doctrine. The model often learns the information that is embodied either in the protected characteristic or in the facially neutral attribute from other data points. Traditionally, a plaintiff litigates to have the discriminator remove the building block from his decision-making practice. In *Griggs*, this meant: not running the IQ test. In the case of algorithmic scoring and creditworthiness evaluation, removing one such building block is mostly unhelpful. If the plaintiff succeeds with his case, the lender deletes one building block from the training data (or from the applicant's digital profile). This could be, for instance, the building block "first name". The outcome will remain unchanged if a pattern, which the AI has detected, still emerges, now on the basis of other data points.

III. Two hypothetical Lenders to Illustrate the Challenges

To illustrate the challenge that anti-discrimination doctrine faces, I introduce two hypothetical lenders.[57] The first hypothetical lender reasons as follows: "The training data for my AI model includes data points of all kinds, including protected attributes. I include these data points because, statistically, sex is an important indicator when calculating credit risk. Women have a higher default risk. However, this is just one of the many observable variables I use. Of course, I do not want to discriminate against anyone, and it is not the only data point I use!"

The second hypothetical lender claims: "I understand that denying credit because of a protected characteristic is impermissible. I use a data filtering method that makes sure that the AI is not trained with data on sex, race, religion, or any other protected characteristic. The model still works fine." [58]

57 See Langenbucher, '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' (n 5) 16 et seq.

58 See for this strategy in Fintech lending: Di Maggio and Ratnadiwakara (n 6) 4; see further: On this argument K Langenbucher and P Corcoran, 'Responsible AI Credit Scoring – A Lesson from Upstart.com' in Emilios Avgouleas and Heikki Marjosola (eds), *Digital Finance in Europe: Law, Regulation, and Governance* (DeGruyter 2022) 143.

1. Causation, But-For Standard, and Proof

The first hypothetical lender submits that sex was but one of the many data points his model was trained on. His underwriting decision on female applicants, so he suggests, includes their sex, but only as one among various other factors. At first glance, this is not a valid defense. Neither US nor EU law require the protected characteristic to be the *sole* building block towards the decision.

As explained above, under US law relevant doctrine has revolved around employment discrimination under Title VII. In *Price Waterhouse v. Hopkins*, the court rejected the argument that a discrimination case requires the contested decision to be triggered *only* by a protected characteristic such as sex. It pointed to the text of the statute that did not read "solely because of".[59] Instead, established practice requires plaintiffs to prove that the protected attribute was one out of various but-for factors that caused the decision.[60] Events can have multiple but-for causes of this type. To win their case, plaintiffs must show that removing the protected attribute would have changed the outcome.[61]

If US courts adopted that standard for the ECOA and the FHA, a lender could not escape liability if a plaintiff can show that a protected attribute was one but-for cause. 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. In *Bostock*, the Court has invoked this terminology to apply what it understands as the but-for standard of causation. Similarly, the FHA speaks of discrimination "because of" protected characteristics. The ECOA makes it unlawful to discriminate "on the basis of" certain protected attributes. [62] None of these statutes require that the outcome was reached "solely" because of the protected attribute, [63] suggesting that both statutes can be read along the

59 *Price Waterhouse v. Hopkins*, 490 U.S. 228 (1989).

60 Berman and Krishnamurthi (n 51) 100 et seq.; Benjamin Eidelson, 'Dimensional Disparate Treatment' (2022) Southern California Law Review Vol. 95 785, 797 et seq.; on a narrower reading of the term "because of" as "by reason of"; see R Dembroff and I Kohler-Hausmann, 'Supreme Confusion About Causality at the Supreme Court, City University of New York' (2022) Law Review Vol. 25, 74.

61 *Bostock v. Clayton County*, 590 U.S. (2020) 6; Dembroff and Kohler-Hausmann (n 62) 58.

62 EU law's close analogue reads: "on grounds of".

63 15 USC Chapter 41 § 1691; 42 USC § 3604.

same lines as Title VII. EU law reaches the same result via its *conditio sine qua non* test:[64] Under Art. 6 of the 2023 Directive, discrimination is prohibited "on ground of" a protected attribute, not "solely on grounds of".[65]

2. The challenge of producing a counterfactual

Following these standards, a hypothetical female plaintiff must show that the first hypothetical lender would have reached a different outcome if the AI had not had access to her sex. In practice, this will be a *probatia diabolica*.[66] The plaintiff would have to run the lender's model (to which he rarely has access) twice: First, she needs to repeat the procedure that the lender followed. Second, she needs to come up with a suitable counterfactual. Maybe there is a way to omit her sex in her application for credit. She could, for instance, change it to male but leave everything else intact. For her case to succeed, the counterfactual would need to look different. However, the more data points the AI model has been trained on, the less likely this is. Omitting her sex entirely or changing it to male, leaves all other data points of her digital footprint untouched. The plaintiff still has a female first name, her height, taste in music or preferred shopping place, the college she attended, or her medical bills might "give her away" to the AI.

The Fintech lender Upstart provides an illustration. To decide on a loan application, it uses bundles of data points, such as education, employment history and more. Upstart only processes variables in concert, not in isolation.[67] Working with the services offered by Upstart, a 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 a discriminatory result. A hypothetical applicant who attended Howard University, a historically black college, paid higher origination fees and higher interest rates over the life of their loans than

64 Rebhahn and Kietaibl (n 30) 378.
65 Rödl and Leidinger (n 47) § 22 note 52; Sacksofsky (n 30), § 14 note 43 et seq.
66 von Ungern-Sternberg (n 11) § 28 note 27.
67 Consumer Financial Protection Bureau (CFPB), Consumer Response Annual Report (2017), 4.

an applicant who was construed as having attended New York University.[68] Similar results were observed for New Mexico State University, a Hispanic Serving Institution.[69] There is a variety of hypotheses to explain these empirical results. We might be looking at taste-based discrimination of the lender, 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.[70] Yet alternatively, historic data might have trained the model to discount applicants from certain colleges.

Either way, the problem with the mystery shopping exercise's methodology is that Upstart uses a bundle of data points that redundantly encode the same information. Giving a comprehensive answer that a plaintiff could have used to make a court case would have required the NGO to run another counterfactual. It would have had to eliminate the variable "college attended" entirely and retrain the model. If Upstart uses a small set of training data only, this might have been the core variable that determined the result. By contrast, the broader Upstart's trainingdata base, the higher the probability that the model would have arrived at the same conclusion without including the college the applicant attended. The AI would have redundantly encoded the same information in other data points – for instance first name, geographical location during term time, friends on social media or taste in music.

3. The challenge of gaining access to the AI

The example has so far assumed that the plaintiff can get access to the lender's AI. In practice, this is not necessarily the case. US, but not EU law allows for pretrial discovery. Even if EU courts are open to changing this, lenders will claim a business defense to keep their AI secret. [71]

68 Student Borrower Protection Center, 'Educational Redlining' (2020), methodology described at 16.

69 Student Borrower Protection Center (n 68), findings described at 17 et seq.

70 Along those lines, R Bartlett et al., 'Consumer-Lending Discrimination in the FinTech Era' (2022) Journal of Financial Economics Vol. 143 30; Gills (n 11), 1188, 1252 et seq.; Hurlin et al., 'The Fairness of Credit Scoring Models' (2021)., HEC Paris Research Paper No. FIN-2021-1411.

71 Note, however, that in a different context, the ECJ did not accept that argument. An employer had used a non-transparent bonus payment system that yielded disparate outcome across men and women. The ECJ required the employer to show that the

Even if courts adjust the plaintiff's burden of proof, a further problem emerges. The lender himself might not have all the relevant data. A recent ECJ case has highlighted the somewhat paradoxical situation of the loan applicant, if a scoring agency is involved.[72] A German loan applicant was denied a loan by a lender, due to a low credit score. Turning to the scoring company to explain the score was not an option: Prior to that ECJ decision, courts and scholars had read the EU GDPR as excluding a private right to be informed on details of a scoring model against the scoring agency. For the applicant, it would not have been helpful to ask the lender to furnish this information either: The lender did not have access to the scoring agency's model. The ECJ remedied the situation, allowing for a private right of action against the scoring agency, at least if the score is of "paramount importance" for a decision on credit. However, even if courts adjusted the burden of proof and are ready to support plaintiffs in access to information regarding lender, credit reporting and scoring agency, redundant encoding still poses a problem. If plaintiffs must show that removing one protected attribute, for instance "sex", triggers a different outcome, they will fail if various data points encode sex.

4. Hard Cases

The second hypothetical lender mentioned above[73] claims to escape liability because the training data for his model does not include any data points that qualify as protected characteristics. It is impossible, so he submits, that he discriminated "on the basis of sex", because he filtered the data his model is trained on. No protected characteristics are used as training data

system was not discriminating against women, rejecting the employer's argument that this would make the bonus system overly transparent. Case C-109/88 - *Handels- og Kontorfunktionærernes Forbund i Danmark / Dansk Arbejdsgiverforening, agissant pour Danfoss* [1989] 383 ECJ, note 15; similar decisions concern Art. 157 AEUV: Case C-170/84 - *Bilka / Weber von Hartz* [1986] 204 ECJ, note 31; Case C-228/89 - *Farfalla Flemming v Hauptzollamt München-West* [1990] 265 ECJ, note 16; Case C-184/89 - *Nimz / Freie und Hansestadt Hamburg* [1991] 50 ECJ, note 15; see Olaf Muthorst, 'Beweisrecht' in Anna Katharina Mangold and Mehrdad Payandeh (eds.), *Handbuch Antidiskriminierungsrecht* (Mohr Siebeck 2022), § 19, notes 30 et seq.

72 Case C-634/21 - *SCHUFA Holding (Scoring)* [2023] 957 ECJ; Katja Langenbucher, 'Die Schufa vor dem EuGH' (2024) BKR-Zeitschrift für Banken-und Kapitalmarktrecht 66; Katja Langenbucher, 'Wirschaft – Medien – Digitalisierung' in E Mand et al. (eds), *Festschrift für Georgios Gounalakis* (Nomos 2025), 715.

73 See C.III.

– like Art. 18 of the 2023 EU Consumer Credit Directive suggests. Faced with cases such as these, courts and scholars in the EU and in the US have so far followed the same ground rules. Either, a protected variable, such as sex, appears as one building block of the decision or else the plaintiff must show a case of disparate impact/indirect discrimination. However, in both jurisdictions there have been hard cases, blurring this bright-line distinction. A common characteristic of these hard cases is that they start from a facially neutral attribute. While this would – in theory – call for the disparate impact/indirect discrimination standard, some attributes stand out as especially "suspicious", distinguishing them from "truly neutral" attributes. As such, they suggest that a decisionmaker might hide his true discriminatory intent behind an attribute that is neutral in form only.[74]

Both jurisdictions have paradigm examples for hard cases like these. In the US, redlining is a classic illustration of a practice that differs in name only from racial discrimination.[75] It refers to the practice of denying an applicant a mortgage in a predominantly minority-owned neighborhood, even though the applicant may generally be eligible for a loan. This form of redlining has been held by courts and regulatory agencies to be an illegal practice[76] entailing disparate treatment (not disparate impact) liability.[77] In this context, it does not matter that the lender does not explicitly refer to "race". Redlining is understood to correlate closely good with race to serve as a proxy. Courts will treat the lender as if he had used the protected attribute itself.

In the EU, there is no comparable history of redlining,[78] but Europe has its own hard cases. An example is provided by a UK court case concerning sex and age discrimination. The defendant was a town that lowered

74 "Covert discrimination", Sacksofsky (n 30) § 14 note 54.
75 Kim (n 33) 15.
76 See <https://www.federalreserve.gov/boarddocs/supmanual/cch/fair_lend_fhact.pdf> accessed 24 June 2025 with exceptions for redlining to identify an area on a fault line or a flood plain.
77 See <https://www.ffiec.gov/PDF/fairlend.pdf> accessed 24 June 2025, iv for the OCC, the FDIC, the Fed Board, the Office of Thrift Supervision and the National Credit Union Administration; similarly A Prince and D Schwarcz, 'Proxy Discrimination in the Age of Artificial Intelligence and Big Data' (2020) Iowa Law Review Vol. 105 1257, 1257; C Campbell and D Smith, 'Distinguishing Between Direct and Indirect Discrimination' (2023) The Modern Law Review, Vol. 86, 307, 315.
78 See for redlining as indirect discrimination B Dzida and N Groh, 'Diskriminierung nach dem AGG beim Einsatz von Algorithmen im Bewerbungsverfahren' (2018) NJW-Neue Juristische Wochenschrift Vol. 71 Issue 27 1917.

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. Hence, for access to the swimming pool, there was sex and age discrimination against men involved.[79] 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 proxy for a protected characteristic.[80]

Similar hard cases revolve around attributes that not only correlate with a protected trait but are understood to be somehow implied by it.[81] Pregnancy is a classic example.[82] 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 facially neutral variable that correlates 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.[83] 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.[84] 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.

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

80 I read this as a hard case which blurs the bright-line distinction, not as supporting the broader view of J Adams-Prassl et al., ‘Directly Discriminatory Algorithms’ (2023) The Modern Law Review Vol. 86 144 that EU law’s scope of direct discrimination is broader than US law’s disparate treatment.

81 See Adams-Prassl (n 82) 157: “inherently discriminatory”; G Krishnamurthi and P Salib, ‘Bostock and Conceptual Causation’ (2020) Yale Journal of Regulation Notice and Comment Blog <https://www.yalejreg.com/nc/bostock-and-conceptual-causation-by-guha-krishnamurthi-peter-salib/> accessed 24 June 2025: “Conceptual Causation”; referring to the latter:Berman and Krishnamurthi (n 51) 88 et seq.; discussing “being a mother” as a “true subset of one sex” on p. 105; Sacksofsky (n 30) § 14 note 58 et seq.

82 For the sake of this example, I do not address the situation of transitioning persons where a man might become pregnant, see Sacksofsky (n 23).

83 *General Electric Co. v. Gilbert*, 429 U.S.125 (1976), para 149.

84 Case C-177/88 *Dekker / Stichting Vormingscentrum voor Jong Volwassenen* [1990] ECJ I-3941, para 12.

Hard cases that sit in between doctrinal categories, such as the ones explained in the preceding section, are neither a novel phenomenon nor necessarily a reason to revisit the starting point of doctrinal distinctions. So far, these cases have in anti-discrimination law been quite sparse, and the debate has profited from the implicit understanding that there is a very limited number of attributes that are *de facto* identical to (or implied in) the protected trait. Additionally, most examples seem to concern deciders who deliberately seek out attributes that are generally understood to be identical to the protected characteristic – such as redlining and race or pregnancy and sex. Imposing disparate treatment liability made sense to close a gap for circumventing the rule.

With the advent of big data and AI models, these implicit assumptions seem less compelling. When discussing the first hypothetical lender, it has become apparent how difficult it was for the plaintiff to prove that omitting the building block "sex" would change the outcome. The reason for this was redundant encoding. Even if the lender omitted sex in the training data or if the plaintiff changed her sex on the credit application, many other attributes "gave her away".

What was a problem of proof in the case of the first hypothetical lender morphs into a question of distinguishing "suspicious" from "truly" neutral variables for the second hypothetical lender. Traditionally, courts had been looking at one suspicious building block ("ZIP code", "pregnancy") that stood in for a protected attribute ("race", "sex"). Now, an AI model unearths correlations between innumerable data points. What used to be an outlier case of an only seemingly neutral attribute (such as ZIP codes) to stand in as a proxy for a protected characteristic (such as race) becomes the new standard. If the correlations the AI will find are as good as the predictive power of pregnancy for sex, can we say that using an AI "implies" a protected attribute? How useful is litigation that targets individual building blocks, if big data will immediately replace one attribute with another? One of the cornerstones of anti-discrimination law, the distinction between protected and neutral attributes, is blurred, due to data analytics redundant encoding. Plaintiffs will find it increasingly impossible to identify distinct building blocks that have caused a discriminatory decision.

D. Art. 6 Directive 2023/2225 on Credit Agreements for Consumers

The EU has stood out for active rulemaking where AI is concerned. Its AI Act[85] stipulates a risk-based approach based on the legislator's perception of especially risky AI use cases. AI-based credit scoring and creditworthiness evaluation count among these, Art. 6 para. 2 AI Act, Annex III Nr. 5b. Recital (58) AI Act explains why this is the case: These AI systems determine "access to financial resources or essential services", they "may lead to discrimination (…) and may perpetuate historical patterns of discrimination (…) or may create new forms of discriminatory impacts". The AI Act's answer are product-specific compliance requirements for developers and users of these systems. By contrast, the Act does not explicitly deal with the relationship between loan applicant and lender or credit scoring company. It is the 2023 Consumer Credit Directive that zooms in on loan applicant and lender.

I. Art. 6's unhelpful text/part 1: "do not discriminate (…) on ground of"

Art. 6 of Directive 2023/2225, for the first time, includes a prohibition of discriminatory lending practices. Its unhelpful drafting style has been mentioned above.[86] Departing from what has been established practice for all EU directives that prohibit discrimination, the new Directive neither defines nor proscribes indirect discrimination. Instead, it prohibits discrimination "on ground of" various protected characteristics. In this way, the rule introduces legal uncertainty into lending practices. Additionally, it misses the chance for a modern rule that gives guidance for the challenge of redundant encoding.

Arguably, the EU legislator did not intend to prohibit only direct discrimination, despite foregoing the established explicit reference to indirect discrimination in its statutory text. It has been explained above that EU courts have read both, Art. 157 TFEU and Art. 21 of the Charter broadly, covering direct and indirect discrimination. Had the EU legislator wanted to change direction, one would have expected a clear sign. Instead, the

85 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) [2024] OJ L 2024/1689.

86 See above B.II.

Directive's recitals (29) and (31), along with the wording of Art. 6 reference the Charter without a qualifier.

Following those same lines, the unusual wording should not be understood as restricting its scope to the relationship between government and private citizens. While the Charter explicitly addresses only the institutions of the Union and its Member States when they are implementing Union law,[87] courts have extended anti-discrimination provisions to cover contracting choices of private parties.[88] The credit underwriting context will mostly concern these, and the Directive's Art. 3 para. (2) is well aware, defining "creditor" as "a natural or legal person who grants or promises to grant credit in the course of that person's trade, business or profession". Against this background, there is no reason to assume that the Directive's prohibition of discriminatory lending practices is to be read narrowly. Still, its text obscures, rather than clarifies the rule's scope.

II. Art. 6's unhelpful text/part 2: "without prejudice to the possibility of offering different conditions (…) where (…) duly justified by objective criteria"

Discrimination presupposes that like cases have not been treated alike and that there is no justificatory reason available. In a credit discrimination case, this makes for two defenses: *Either*, the plaintiff is not like the members of the group he claims he belongs to, hence, not "similarly situated", *or* there was a good reason to treat him differently. Both, EU, and US law subscribe to the admissibility of some form of business defense in disparate impact/indirect discrimination cases. However, neither US nor EU law spell out in detail what an acceptable business defense looks like. While *Manhart* and a similar decision by the ECJ, concerning an earlier anti-discrimination Directive, [89] had ruled out a defense of statistical discrimination, neither

87 Charter of Fundamental Rights of the European Union, Art. 51 para. 1 s. 1, see Klaus Ferdinand Gärditz in Anna Katharina Mangold and Mehrdad Payandeh (eds), *Handbuch Antidiskriminierungsrecht* (Mohr Siebeck 2022), § 4 note 66.

88 See above B.II.

89 Case C- 236/09 *Association Belge des Consommateurs Test-Achats and others* [2011] ECJ.

court has extended this jurisprudence to discrimination on grounds of other protected characteristics.[90]

The wording of the new rule, instead of providing a clear guideline, veers on the side of caution. Art. 6 para. 2 stresses that the non-discrimination rule “shall be without prejudice to the possibility of offering different conditions”. Recital (31) adds that “this should not be understood as creating an obligation for creditors or credit intermediaries to provide services in areas in which they do not conduct business”. Denying credit altogether or asking for higher interest is acceptable if “those different conditions are duly justified by objective criteria”. What these criteria could entail remains open. That they must be “objective” seems evident and rules out a procedure that allows for a subjective assessment of creditworthiness.

Much will depend on national law, regulating the burden of proof for plaintiffs. If the applicant bears the burden, he will need access to the lender’s AI model, data and business strategy. A recent ECJ decision seemed open towards an approach suggested by the referencing Austrian court: The lender was not required to deliver this information to the loan applicant but, instead, to the court.[91] However, even if the lender produced all information and even if the courts – relying on experts – was able to meaningfully assess these, the applicant would need to produce some form of counterfactual data to show that he should have been offered the loan. This might be impossible, given that the AI learns which loans it should *not* have offered because borrowers did worse than predicted (false positives). But it does not usually learn which loans would have been attractive (false negatives). Because the denial of a loan (or a certain interest rate) implies that the loan (or the interest rate) was not offered to the applicant, the AI has no chance to learn whether the applicant would have paid back.[92]

The EU legislator has not explained what counts as “duly justified objective criteria” to justify disparate impact, leaving the issue for regulators and judges to settle. The EU AI Act is more helpful than the Directive. Its recital (58) starts from the premise that “AI systems used to evaluate the credit score or creditworthiness of natural persons should be classified as high-risk”. The recital moves on to clarify that the high-risk category does not apply to AI systems used “for prudential purposes to calculate credit

90 See Antje von Ungern-Sternberg in Anna Katharina Mangold and Mehrdad Payandeh (eds), *Handbuch Antidiskriminierungsrecht* (Mohr Siebeck 2022) § 28 note 57.
91 Case C-203/22 *Dun & Bradstreet Austria* [2025] ECJ.
92 Kim (n 33) 5.

institutions' and insurances undertakings' capital requirements". It stands to reason that, in the scope of the Directive, (statistical) discrimination based on prudential purposes provides a "duly justified objective criterion". Beyond that, recitals (31) and (46) suggest leeway for the lender to structure his business model as he sees fit. It remains to be seen, how this liberal approach fits together with EU courts prohibiting discriminatory underwriting between private parties in its directly or indirectly discriminatory form.

III. The Consumer Credit Directive in the age of AI: Algorithms for inclusion?[93]

The Consumer Credit Directive regulates consumer lending more generally, remarks on AI-based creditworthiness evaluation are sparse. As is evident from recital (56), legislators are aware of (some) risks that algorithmic scoring and underwriting entail but decided to refrain from going into details. The problem of redundant encoding is not addressed nor are problems of proof. There is a right to obtain human oversight where creditworthiness assessments involve automated processing of personal data, Art. 18 para. 8, recital (56). Additionally, Art. 18 para. 3 s. 5 rules out "social networks" as a source of information for assessing creditworthiness of the consumer.

The Directive does not specify what a "social network" is, nor does it allow to use social network data if it helps the consumer or if he consents. The US Consumer Financial Protection Bureau (CFPB) offers an illustration of how this could work.[94] Early on, it had issued a no-action letter for the Fintech Upstart mentioned above, drawing a comparison of traditional and novel credit scoring methods.[95] It first simulated outcomes under Upstart's proprietary model. Then, the Bureau compared them with outcomes

93 Orly Lobel, *The Equality Machine, Harnessing digital technology for a brighter, more inclusive future* (Public Affairs New York 2022).

94 Discussion at Langenbucher, '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' (n 5) 34.

95 P Ficklin and P Watkins, 'An update on credit access and the Bureau's first No-Action Letter' (Consumer Financial Protection Bureau Blog, 6 August 2019) <https://www.consumerfinance.gov/about-us/blog/update-credit-access-and-no-action-letter/> accessed 24 June 2025.

under a hypothetical model using FICO scores (a common US scoring agency).[96] The simulation had Upstart approve 27 % more borrowers than traditional lending models. Personal loan interest rates were 16 % lower on average.[97] The CFPB found no disparities for minorities, females, or applicants who are 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.[98]

One concern the Bureau did not address is that the distribution remained skewed. Black and black-Hispanic minority borrowers, who were eligible under Upstart's model, were still facing disadvantages. These emerged when comparing their group with the group of white and white-Hispanic persons eligible under Upstart's model. The disadvantage showed as to relative numbers of access to credit, origination fees, and interest rates. 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 than before and, in that sense, we are looking at something like pareto-optimality.[99]

Comparing a hypothetical simulation with a standard credit score as the counterfactual is not entirely unreasonable 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. To be sure, this form of "pareto-optimality" test along the CFPB's lines is incompatible with received disparate impact doctrine. Anti-discrimination law tries to remedy relative disadvantages of one group when compared with another group. The Bureau focused instead on the surplus produced by Upstart's model, irrespective of the relative composition of the group of borrowers. While this makes the strategy unhelpful for private claimants, it could offer some guidance for public enforcement or supervisory oversight.

---

96 Critical as to this method: Student Borrower Protection Center (n 68) 21 (fn. III) but see above C.III.2 for a critique of the mystery shopping exercise.
97 Ficklin and Watkins (n 95).
98 Ficklin and Watkins (n 95).
99 On this argument see Langenbucher and Corcoran, (n 60) 156.

E. Take-aways

This paper has explored shortcomings of received anti-discrimination doctrine when faced with redundant encoding. Many concern the burden of proof where big data encodes the protected attribute via various not-protected attributes. Additionally, plaintiffs would need access to data and model, potentially from various actors, including the lender, scoring agencies, and credit reporting bureaus, to make their case. Should courts be ready to adjust the burden of proof, redundant encoding will still in many situations rule out proof.

For disparate impact doctrine, it has under US law been unclear whether its scope covers credit underwriting. The Executive Order of April 23, 2025 requests agencies to stop proceeding under this doctrine.

The EU Consumer Credit Directive has not been explicit on this point. Received court practice would suggest that the doctrine is applicable. Still, it suffers from the problems mentioned above: Anti-discrimination doctrine requires the plaintiff to establish a specific "building block" – an attribute, a characteristic, a criterion or a certain practice – that has triggered the disparate output across groups. If the lender uses big data, redundant encoding makes this approach toothless. As soon as one building block is removed, for instance a protected characteristic or a facially neutral attribute deleted from the data, various correlating data points take its place.

Unhelpfully, the EU Directive 2023/2225 on Consumer Credit Agreements has put the ball in the field of regulatory agencies and courts. This raises various challenges for consumers, for credit scoring agencies and for financial institutions that wish to lawfully employ AI. Among these are legal risks as to discriminatory conduct, the admissibility of business defenses, the tools supervisory agencies may employ, and burden of proof issues for plaintiffs.[100] Mostly, the Directive's anti-discrimination regime in Art. 6 seems geared towards traditional, limited-data credit scoring and underwriting. If consumer lending increasingly moves towards big data analytics EU regulators and courts will have to decide how to deal with redundant encoding.

100 An exploration in detail is beyond this current paper's scope, see Katja Langenbucher, 'Financial Profiling' in: Langenbucher (ed) *The Regulation of Credit Scoring in Europe* (forthcoming in Edward Elgar 2025) for further details.

The AI Act offers the chance to reorient the discussion, replacing the intricacies of algorithmic discrimination with internal validation, regulatory supervision, product quality checks and testing of both, AI model and data.[101] Arguably, this product-regulation approach provides a better fit with the challenges raised by AI-based credit underwriting than traditional anti-discrimination doctrine.

101 Katja Langenbucher, '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' (n 5) 370 et seq.