Katja Langenbucher and Patrick Corcoran

Responsible AI Credit Scoring – A Lesson from Upstart.com

Abstract: Modern FinTech companies are disrupting the traditional credit scoring model for loan decision-making by turning to artificial intelligence and machine learning systems. They use those systems to assess creditworthiness based on "alternative data" like banking activity or education history. Such AI scoring has the potential to extend credit to those whose creditworthiness is not captured by standard scores. At the same time, it presents new concerns that current regulatory schemes are ill-equipped to address. This paper raises these concerns and compares the current U.S. and EU regulatory regimes insofar as they may apply to the emerging AI scoring industry. The first issue is data privacy in AI credit scoring. The EU regulates this under the omnibus approach of the General Data Protection Regulation, in the United States it implicates the Fair Credit Reporting Act. The second issue is discrimination in AI-based lending, which falls under the U.S. Equal Credit Opportunity Act and a number of European Anti-Discrimination Directives. The paper discusses the discrimination issue in the context of the U.S. Consumer Financial Protection Bureau's decision to grant the no-action letter requested by FinTech lender Upstart. In a postscript, we discuss the EU's recent proposal for an Artificial Intelligence Act and provide some preliminary thoughts on the Proposal's provisions in the context of the challenges of AI scoring regulation raised in this paper.

Table of Contents

Katja Langenbucher and Patrick Corcoran, Katja is tenured at Goethe University, House of Finance, Frankfurt, affiliated professor at Leibniz Institute for Financial Research SAFE and SciencesPo, Paris, and visiting at Fordham Law School, NYC. Patrick holds a JD of New York University School of Law, and is admitted to the bar in NYC. The authors are most grateful to feedback received during the conference to which this ECFR issue is devoted, excellently organized by Emilios Avgouelas and Heikki Marjosola. We would also like to thank Iris Chiu, UCL London, and Udo Milkau, Frankfurt, for their feedback on prior versions of this paper. Needless to say, remaining errors are ours.

 https://doi.org/10.1515/9783110749472-006

1 Introduction

"Traditional credit scores leave people behind. We use artificial intelligence to expand access to reasonably priced credit." This is how Upstart.com advertises its services to consumers. The company's website invites visitors to choose from a drop-down menu their personal credit goal (such as refinancing or making a purchase), and to "check your rate." Further questions concern the applicant's approximate credit score, details on his level of education and primary source of income, and a number of personal details. After providing this information and some supporting documentation, the applicant may choose his loan and the terms offered by Upstart.[1]

What distinguishes lenders like Upstart from a traditional bank? Traditional lenders—including, most prominently, the major banks and credit card compa-

1 For a more detailed description see Upstart's request for a no action letter, p. 1 et seq., available at: https://files.consumerfinance.gov/f/documents/201709_cfpb_upstart-no-action-letter-request.pdf (last accessed 22 January 2021) (hereinafter "Request for a no action letter").

nies—rely mainly on a predetermined set of factors when evaluating a loan applicant's creditworthiness based on his credit history. The factors determining a traditional FICO score include the applicant's history of on-time or late payments, the percentage of their available credit that they use, the length of their credit history, the variety of their credit (e. g. credit cards, mortgages, and installment loans), and the recent acquisition of new credit.[2] Upstart, on the other hand, does not rely exclusively on these factors, at least not for all applicants. Instead, it employs an artificial intelligence-based model that distills an alternative credit score from non-FICO data points looking at probability of repayment based on future salary. Upstart also operates online without brick-and-mortar locations, but partners its AI platform with a traditional bank which actually originates loans approved under Upstart's model.

At the heart of what distinguished Upstart from traditional lenders is its AI scoring model, which focuses on borrowers' level of education and high-income potential in order to predict future salary, hence, probability of repayment.[3] If an applicant's credit score is below Upstart's minimum credit underwriting requirements, Upstart will accept him only if he has graduated from or is currently enrolled in an associate, four-year bachelor, or more advanced degree at an accredited school.[4] The underlying business idea seems compelling: Instead of narrowing the assessment of a future borrower to FICO-score criteria and past credit history, further variables are taken into account. These alternative data give a richer picture of financial capacity and likelihood to repay a loan, especially for applicants with short credit histories.[5] Young borrowers or recent immigrants enrolled in school or with a job offer present an attractive market, undertargeted by traditional lenders because they lack the history of interaction with credit markets that is required to achieve an adequate FICO score. At the same time, specific groups of potential borrowers are deliberately left out.[6] Worried that this might raise concerns of direct discrimination or disparate impact, Up-

2 Fair Isaac Corporation (FICO), What's in my FICO Scores?, (22 November 2020), https://www.myfico.com/credit-education/whats-in-your-credit-score (last accessed 22 January 2021).

3 On the potential of digital data to more accurately predict future events see: Communication from the Commission to the European Parliament, the Council, the ECON and the Committee of the Regions, 24.9.2020, COM (2020) 591 final, p. 3.

4 See Request for a no action letter (fn. 1), p. 2.

5 See ibid., p. 3.

6 For a critique of bias in consumer lending, see Student Borrower Protection Center, Educational Redlining (February 2020), https://protectborrowers.org/wp-content/uploads/2020/02/Education-Redlining-Report.pdf (last accessed 22 January 2021).

start applied for a no-action letter in 2017 which we will discuss in more detail below.

The CFPB's decision to issue a no-action letter to Upstart is both an indication that major changes to the consumer lending industry are inevitable, and a reminder that such change will create new regulatory challenges as existing rules are applied to technologies that their drafters did not anticipate. The Bureau's acknowledgement of alternative credit scoring opens the door for companies like Upstart to fulfill their promises of more efficient and inclusive lending. At the same time, the CFPB's acceptance of alternative data and AI modelling could prove a difficult fit with current regulations including the Fair Credit Reporting Act (FCRA) and Equal Credit Opportunity Act (ECOA).

This paper will examine the regulatory scheme formed by these and other statutes in the U.S. and EU to better understand how the fintech companies embracing alternative credit scoring fit into those schemes, and where these regulations may need to be adjusted to account for AI scoring methods. Part 2.1 summarizes and compares the relevant consumer lending and data privacy regulations of the U.S. and EU as they apply to lenders and scorers. Part 2.2 analyzes these jurisdictions' anti-discrimination regulations. 3 identifies questions arising from the Upstart no-action letter specifically and discusses how alternative scoring models may implicate discrimination. Part 4 analyzes how the Upstart case may apply generally and suggests how data and model quality may be improved as a result. Part 5 concludes, and Part 6 revisits these questions in the context of the EU Proposal for an Artificial Intelligence Act of April 21, 2021.

2 Which Regulatory Framework for Non-Traditional Data?

Algorithmic scoring models have started to attract regulatory scrutiny for two main reasons. The first has to do with the data collected, processed and transferred to third parties. The second concerns the hidden potential for discriminatory outcomes when using alternative data.

2.1 Data Privacy Regulation

2.1.1 The Lender

One immediate application of algorithmic scoring models is for the lender itself to apply his model to proprietary data it has already collected about the borrower. Such data may stem from a prior contractual relationship with the borrower, like a previous loan or existing bank account. In the course of this relationship, the borrower will have provided the lender with data about himself. This may include, for instance, data submitted in past applications to take out loans or open accounts. Such data encompasses names, addresses, phone numbers, and sometimes credit card account and social security numbers, income and credit histories. It may also extend to information about what kinds of stores the borrower shops at, how much he borrows, his account balance or the dollar value of his assets, what the borrower has purchased with a debit or credit card,[7] whether credit card applications have been denied, or his traditional credit score.

Lenders have naturally decided to use this type of data in the past when making a creditworthiness assessment, and several data privacy regulations apply to such use. In the U.S., the Gramm-Leach-Bliley Act (GLBA) provides the main regulatory framework, and the Financial Credit Reporting Act (FCRA) includes data-sharing rules for those who receive credit reports. State laws such as the California Consumer Privacy Act and the proposed New York Privacy Act impose additional obligations on companies handling consumer data. However, the California law makes exception for data shared among consumer reporting agencies and their furnishers, as that information is already subject to FCRA regulation.[8] In the EU, the General Data Protection Regulation (GDPR) has so far constituted the only relevant regime. The credit reporting infrastructure as such differs from country to country in the EU and there is no EU legislation in place. The proposed AI Act offers a regulatory framework for the use of AI scoring applications. Most of the Act's requirements concern the developer of the application, not necessarily the lender that utilizes the AI system (see *infra* Part 6).

7 Credit and debit purchases are considered nonpublic personal information under the Gramm-Leach-Bliley Act, and therefore can only be shared with nonaffiliated third parties if the consumer is given clear and conspicuous notice and an opportunity to opt out of the disclosure, 15 U.S.C. § 6802 (a)–(b).

8 See, e. g., California Consumer Privacy Act (CCPA), Transunion, https://www.transunion.com/consumer-privacy (noting that "personal information related to your credit report is not subject to the CCPA") (last accessed 22 January 2021).

The GLBA requires financial institutions to safeguard certain sensitive data.[9] To comply, financial institutions have to "develop, implement, and maintain a comprehensive information security program that […] contains administrative, technical, and physical safeguards."[10] Additionally, financial institutions have to explain their data-sharing practices to their customers.[11] If they share information with certain third-party non-affiliates, i.e. companies which are not part of the same corporate group, customers must be notified.[12] There are some disclosures for which financial institutions are not required to provide the consumer with notice and an opportunity to opt out, such as when no customer relationship has been established or the information is being shared with an affiliate entity.[13]

Going beyond the GLBA, under the FCRA a financial institution which has received information from a consumer reporting agency and intends to share that information with an affiliate becomes a credit reporting agency (CRA) for FCRA purposes and is subject to the same notice and information-sharing requirements as CRAs.[14] That said, under the FCRA a financial institution can still share information relating to transactions between the consumer and that institution,[15] and may share consumer report information with entities with which it is affiliated or shares common ownership provided that consumers are provided with notice and opportunity to opt out.[16]

A core feature of both the FCRA and GLBA is the somewhat permissive (as compared to EU regulation) approach towards lenders who wish to utilize consumer data they have collected. While it has to safeguard certain data, the lender is at liberty to use consumer data as it sees fit to evaluate applicants' creditworthiness (subject, of course, to the antidiscrimination regulations discussed below). This approach goes hand in hand with placing the burden to take initiative on the borrower. Lenders who plan to share data with affiliates or non-affiliates must notify the borrower, but it is the borrower who has to speak up and opt out – if a right to opt out exists, that is. As of now, the regulatory regime is the

9 Safeguards Rule, 16 Code of Federal Regulations (CFR) § 314, implementing sections 501 and 505(b)(2) of the Gramm-Leach-Bliley Act.
10 16 C.F.R. §314.3.
11 16 C.F.R. § 313.3(b).
12 16 C.F.R. § 313.4 – 313.6.
13 16 C.F.R. § 313.4(b).
14 *Chris Brummer*, Fintech Law in a Nutshell, 2020, p. 320 et seq.
15 15 U.S.C. § 1681a(d)(2)(A) ("the term 'consumer report' does not include […] information solely as to transactions or experiences between the consumer and the person making the report").
16 12 C.F.R. § 1022.20 et seq.

same regardless of whether a lender is using traditional or alternative, AI-driven scoring models.

EU law follows a considerably less liberal regime as to data protection. Art. 6 GDPR requires there to be a legitimate reason for any form of data collection or processing. A lender who makes use of data, even if it is proprietary data he has collected about the borrower, qualifies as a "data processor" under Art. 4 para. 2 GDPR: "any operation [...] which is performed on personal data [...] such as collection, recording, organization, structuring, storage." Instead of requiring the borrower to take the initiative to opt out, it is the lender who must show that its handling of data is legal under Art. 6 para. 1 (a) GDPR ("Processing shall be lawful only [...]"). Additionally, if employing an algorithm entails profiling, special safeguards apply under Art. 22 para. 1 GDPR. The GDPR provides for a general prohibition on decisions based solely on automated processing ("The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her [...]"), allowing for exceptions in its para. 2 (necessity to enter into a contract, authorization under Union or Member State law, explicit consent by the data subject). For specially protected categories of data, even fewer exceptions apply.[17]

2.1.2 The Scoring Agency

Algorithmic credit scoring, especially when based on non-traditional data, will often be done not by traditional lenders, but by third party FinTech companies. Traditionally, credit bureaus have delivered credit scores (e. g. by the German SchuFa) or credit reports (e. g. by Experian, TransUnion and Equifax in the U.S.), the latter of which form the basis for the applicant's FICO[18] score. FinTech companies such as ZestFinance[19] and Underwrite.ai[20] offer novel scoring models which go beyond the traditional variables underlying the FICO score. AI, machine learning, and related technologies enable these companies to model and predict creditworthiness based on a more complex analysis of relevant consumer data. AI-based scorers of this type may rely exclusively on the proprietary data of

17 Art. 22 para. 2, Art. 9 para. 2 (a), (g) GDPR.
18 FICO is a leading analytics software company that delivers the software to compute credit scores to many of the largest U.S. lenders, but is not itself a credit reporting agency, FICO, About Us, https://www.fico.com/en/about-us#our-company (last accessed 22 January 2021).
19 https://www.zest.ai (last accessed 29 January 2021).
20 https://www.underwrite.ai (last accessed 29 January 2021).

the lender itself. Such seems to be the case for Underwrite.ai, which adds value by applying more sophisticated analysis to the data contained in a lender's pre-existing data on cured loans. Underwrite.ai's approach has no need for the collection of additional data, while other companies, like Upstart, rely on new sources of data to supplement traditional FICO elements.

Beyond delivering novel scoring models to be applied to lenders' data, FinTech companies like Upstart collect their own data in addition to running it through their AI-based models to compute a score. Often, these companies source borrowers, but a bank originates the loan. In the case of Upstart, Cross River Bank, operating under a New Jersey charter, is the originator. FinTech business models vary in detail. Some have the issuing bank take care of the entire process of debt collection while others are involved in servicing, funding and debt collection and may even buy back the loan.

Algorithmic scoring models rely heavily on data. Of course, all statistical credit scoring uses data to some degree, but AI scoring is unique in the sheer volume of data processed and the number of variables that may be analyzed in creating and applying models. While companies like Underwrite.ai and GiniMachine[21] use only the lender's own historical lending data, others acquire explicit permission from customers to access more data.[22] Petal, for example, requires that applicants with little to no credit history link their bank accounts in order to apply for certain products.[23] Models may include variables which the user does provide, but where he does not necessarily understand the way in which they are important in a credit context. An often-cited example concerns a specific font found on a user's electronic device which correlated with the use of an online gambling site.[24] Others ask potential borrowers to grant broad access to some form of digital footprint, like a PayPal or Amazon account, a mobile phone or a fitness tracking app. They then correlate such data points with their proprietary data on probability of repayment. This is where machine learning comes in to analyze the relationships and interactions between hundreds of potentially relevant variables, and thereby discover the predictive power of data

21 https://ginimachine.com (last accessed 29 January 2021).
22 See GiniMachine, "How it works: An End-to-End Scoring Platform", https://ginimachine.com/product/ (explaining that models are based on records of previously issued loans) (last accessed 22 January 2021).
23 Petal, "What do you do with my bank information?", https://support.petalcard.com/hc/en-us/articles/360012518794-What-do-you-do-with-my-bank-information- (last accessed 22 January 2021).
24 On Kreditech see the report on p. 23 at https://www.european-microfinance.org/sites/default/files/document/file/Inclusive-credit-scoring-Final.pdf (last accessed 22 January 2021).

points that might otherwise never be realized as predictive of a likelihood to repay.[25]

EU regulation will usually understand scoring agencies as "data processors."[26] Under the GDPR's omnibus regime, this includes "data collection" as well as "disclosure by transmission, dissemination or otherwise making available." In order for data processing to be legitimate, it must qualify under one of the GDPR's exceptions. The most natural exception is under Art. 6 para. 1 (a): if the data subject gave his consent. Consent has to be in the form of a "freely given, specific, informed and unambiguous indication of the data subject's wishes."[27] If the non-traditional data involves protected categories, Art. 9 GDPR lays down a stricter regulatory framework, asking for explicit (rather than "freely given") consent.

The European Court of Justice (ECJ) has just started to specify what it considers necessary features for consent. A pre-checked box on a website does not meet the court's standard of "active" consent. In an *obiter dictum*, the ECJ raised doubts whether behavioral nudges, such as making continuation in an online gambling game dependent on giving consent to the processing of one's data, are legal.[28] If there is no consent, processing may be legitimate if it "is necessary for the purposes of the legitimate interests pursued by the controller or by a third party," as long as these legitimate interests outweigh the interests and fundamental rights and freedoms of the data subject.[29] However, it is certainly doubtful that the legitimate interests of a scoring agency will ever outweigh the interest of the data subject in preventing access to their private data without their consent, likely rendering this exception ineffectual in this context.

In a 2020 Whitepaper on artificial intelligence, the EU Commission began to outline a "European approach to excellence and trust" that addresses privacy protection when employing AI, among other concerns.[30] The report highlights

25 On the basis of (limited, non-representative) empirical research the authors of this paper have done, scoring agencies always ask for consent. Put differently: we have not seen agencies which scrap the internet for publicly available information on potential borrowers. Of course this is not to say that such business models do not exist.

26 Art. 4 para. 2 GDPR.

27 Art. 4 para. 11 GDPR.

28 ECJ, 1 October 2019, Planet49, C-673/17, ECLI:EU:C:2019:801, mn. 64.

29 Art. 6 para. 1 (f) GDPR.

30 White Paper on Artificial Intelligence – A European approach to excellence and trust, p. 10 et seq., available at: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf (last accessed: 14 February 2021); the Public Consultation on the AI White Paper, Final Report on the public consultation is available at: https://ec.europa.eu/digital-that, while a regulatory framework is already in place, regulators must continue to consider proper enforcement and, possibly, the need for adjustments to existing regulations. As mentioned above, the Whitepaper did not specifically address AI scoring.

FinTech companies equipped with a banking license and originating the loan themselves will have to comply with the GDPR's regulation on "profiling" and on "decisions based solely on automated processing." Recital 71 of the GDPR explicitly refers to a prohibition of (fully) automated refusals of an online credit application on the basis of profiling, unless Union or Member State law allows for them. Companies which are involved in scoring only, i. e. whose models propose to issue the loan which is then granted by an originating bank, will usually still be involved in automated processing under the GDPR.

Under U.S. law, the FCRA applies to entities which qualify as a "consumer reporting agency" (CRA) and to data which can be considered a "consumer report." CRAs are agencies that compile and maintain public information and credit account information "for the purpose of furnishing reports to third parties bearing on a consumer's credit worthiness."[31] A consumer report is any communication "bearing on a consumer's credit worthiness [...] which is used or expected to be used [...] as a factor in establishing the consumer's eligibility for" credit, insurance, or employment.[32]

Under the FCRA, a CRA may not report information adverse to the consumer if that data is over seven years old.[33] The CRA may only furnish a credit report for certain enumerated purposes, including the evaluation of applicants for credit, insurance, and employment.[34] To ensure they comply with this requirement, CRAs must require their clients (i. e. the lender, insurer, or employer) to identify themselves and their purposes for requesting the consumer report.[35] The fees a CRA charges for reports must be reasonable.[36]

The statute also imposes responsibilities on the entities that furnish the CRA with consumer information. These 'furnishers' constitute a wide variety of good

single-market/en/news/white-paper-artificial-intelligence-public-consultation-towards-european-approach-excellence (last accessed: 14 February 2021), see pp. 12, 14 on privacy.

31 15 U.S.C. § 1681a(p).

32 Id. at § 1681a(d)(1).

33 15 U.S.C. § 1681c(a).

34 15 U.S.C. § 1681b(a)(3).

35 15 U.S.C. § 1681e(a).

36 15 U.S.C. § 1681 g(f)(8).

and service providers with whom consumers directly interact.[37] Furnishers must notify consumers when negative information is sent to a CRA,[38] and must not furnish information that a consumer has told them or they otherwise have reason to know is inaccurate.[39] Consumers are entitled to know the sources of the information in their credit report,[40] and to dispute the accuracy of information directly with the entity that furnished it.[41]

Summing up, artificial intelligence models' reliance on big data and FinTech lenders' interest in a wider array of data points to inform alternative credit scoring models will inevitably bring those lenders within the scope of various data privacy regulations. Though such regulations will generally require the consent of and disclosures to the borrower when lenders access and use data to make credit decisions, data privacy regulations alone cannot guarantee that borrowers understand the scope of any alternative data accessed, and how nontraditional data points might affect their credit decision.

2.2 Anti-Discrimination Regulation

A perhaps less evident concern when dealing with P2P lending and AI-based credit scoring is the regulation of discriminatory lending practices. Withholding credit solely on the basis of certain characteristics of the borrower, such as gender, race or religious affiliation, is prohibited in both the U.S. and the EU. These regulations obviously rule out AI-based models which explicitly make their credit decision dependent on these characteristics. However, the problem with AI-based models is a more complicated one. The larger the data pool from which machine learning algorithms pull the correlations they use, the higher the risk that a correlation indirectly discriminates or – in U.S. terminology – has a dispa-

37 Furnishers mostly consist of "automobile dealers; banks, clothing, department, and variety stores; finance agencies; grocery and home furnishing dealers; insurers; jewelry and camera stores; contractors; lumber, building materials, and hardware suppliers; medical-care providers; national credit card companies and airlines; oil companies (credit card divisions); personal services other than medical; mail-order houses; real estate agents; hotel keepers; sporting goods and farm and garden supply dealers; utilities; fuel distributors; government agencies (e. g. the Federal Housing Administration and the Veterans Administration); wholesalers; advertisers; and collection agencies." Frederick H. Miller/Alvin C. Harrell/Daniel J. Morgan, Consumer Law: Cases, Problems, and Materials, 1998, 296.
38 15 U.S.C. § 1681s-2(a)(7).
39 15 U.S.C. § 1681s-2(a)(1).
40 15 U.S.C. § 1681 g(a).
41 15 U.S.C. § 1681s-2(a)(8)(E).

rate impact on protected groups. This is because an AI system can find correlations between a high likelihood of debt repayment and complex combinations of input variables, some of which may have no obvious relationship to a person's financial tendencies or responsibility. For example, an algorithm might recognize that applicants who shop online at Website X and communicate with Messenger App Y are less likely to stay on top of credit card payments. However, the intersection of those two variables may well serve as a proxy for a particular race even if race itself is not being considered as a standalone variable, and the scorer may not even realize that race or other protected classes are indirectly influencing their model's calculations in this way.

Upstart provides a compelling illustration of this concern. In the U.S., education is significantly correlated with race. A 2015 report by the U.S. Census Bureau indicated that among people aged 25 and older, 36 % of white people had attained at least a bachelor's degree, compared to 23 % of Black people and 16 % of Hispanic people. For the same population, 14 % of white people held advanced degrees, compared to 8 % of Black people and 5 % of Hispanic people.[42] A study by the Student Borrower Protection Center shows that Upstart's education-dependent model leads to higher costs (e. g. interest rates and origination fees) for students of Historically Black Colleges and Universities and Hispanic-Serving Institutions than for students of non-minority serving institutions.[43] When refinancing student loans with Upstart, this study found, a hypothetical Howard University graduate[44] is charged almost $3,500 more over the life of a five-year loan than a NYU graduate, all other inputs held constant.[45] A hypothetical graduate with a B.A. from New Mexico State University, a Hispanic-Serving Institution, is charged almost $1,800 more over the life of a five-year loan than a NYU graduate.

The approaches to anti-discrimination regulation in the U.S. and EU mirror the two jurisdictions' approaches to data protection: U.S. law relies on a sectoral legal regime prohibiting discriminatory lending practices, while the EU provides for more general anti-discrimination rules.

---

42 *Camille L. Ryan/Kurt Bauman*, Educational Attainment in the United States: 2015, U.S. Census Bureau (March 2016), https://www.census.gov/content/dam/Census/library/publications/2016/demo/p20-578.pdf (last accessed 22 January 2021).

43 Student Borrower Protection Center (fn. 6) p. 15 et seq.

44 Howard University is a historically Black University.

45 Student Borrower Protection Center (fn. 6) pp. 4, 7.

Setting aside fundamental human rights protections and how they impact contract law,[46] EU law provides for a number of Directives which prohibit discrimination in specific situations such as employment or social security. EU Directive 2000/43/EC is intended to implement the principle of equal treatment between persons irrespective of racial or ethnic origin. Under Art. 2 para. 2 (a), (b) and Art. 3 para. 1 (h), the Directive prohibits direct and indirect discrimination in relation to "access to and supply of goods and services which are available to the public, including housing." To consider credit scoring and loan contracts as qualifying for that rule, those agreements would have to be standardized services, rather than individualized agreements. Art. 3 EU Directive 2004/113/EC prohibits direct and indirect gender discrimination as to the offer of goods and services which are available to the public, except for some goods related to private and family life. With its broader wording, scoring and loan provision will in many cases qualify.

In its Whitepaper on AI, the EU Commission expressed its awareness of the potential for discrimination that AI presents.[47] The report highlights both "flaws in the overall design of AI systems" and issues arising "from the use of data without correcting possible bias."[48] Drawing conclusions from such preliminary work, the Proposal for an AI Act sets up a regulatory framework with the explicit goal of prohibiting harmful practices which "contradict Union values of respect for human dignity, freedom, equality, democracy and the rule of law and Union fundamental rights, including the right to non-discrimination, data protection and privacy."[49]

In the U.S., the ECOA makes it unlawful for any creditor to discriminate against any applicant on the basis of race, color, religion, national origin, sex, marital status, or age.[50] Creditors also cannot discriminate because an applicant derives all or part of their income from public assistance, or because an applicant has in good faith exercised their rights under the Consumer Protection Act.[51] The ECOA also creates a private right of action for applicants against creditors who have discriminated against them.[52] The Act includes both direct or in-

46 In more detail at *Katja Langenbucher*, "Responsible AI-based Credit Scoring – A Legal Framework", 31 European Business Law Review 2020, 527, 544 et seqq.
47 See White Paper (fn. 30), 1, 10 ("opaque decision-making"); Final Report (fn. 30), pp. 14, 16, 17.
48 Final Report (fn. 30), p. 11.
49 Proposal, Recital (15).
50 15 U.S.C. § 1691(a)(1).
51 Id. at § 1691(a)(2)–(3).
52 15 U.S.C. § 1691e(a).

tentional discrimination based on the aforementioned factors, and indirect or "disparate impact" discrimination in which the lender's practices have "a disproportionately negative impact on members of a protected class—and the lender is unable to demonstrate that the practice is justified by a legitimate business need and cannot reasonably be achieved by other less discriminatory needs."[53]

The ECOA also requires lenders to notify applicants of adverse actions (e.g. denying credit or offering credit on less favorable terms) within 30 days.[54] That notice must contain the specific reasons for which the decision was made or a promise to deliver that explanation upon the applicant's request. Broad statements that the adverse action was based "on the creditor's internal standards or policies, or that the applicant [...] failed to achieve a qualifying score on the creditor's crediting system are insufficient."[55] In other words, regardless of how complex a scoring algorithm might be, incorporating myriad variables and interactions, the decisions it recommends must be explainable in a way that is comprehensible. However, depending on the complexity of a model and the number and variety of variables bearing on its decision, an explanation comprehensible to the average consumer may necessarily fall short of identifying all the factors contributing to the decision, and a fully accurate explanation may prove too verbose and intricate to be readily understood. Even if the lender can clearly explain the workings of its particular model, such disclosures could also implicate the scorer's proprietary decision algorithm.

ECOA applies to all creditors, which the statute defines as any person who "regularly extends, renews, or continues credit; any person who regularly arranges for the extension, renewal, or continuation of credit; or any assignee of an original creditor who participates in the decision to extend, renew, or continue credit."[56] Under this definition, a company like Underwrite.ai or GiniMachine that sells its AI technology to lenders but does not extend credit itself would not face liability if its model's decisions were not sufficiently explainable. However, there would seem to be some responsibility on the scorer to provide a scoring model that can at the very least be transposed into traditional credit factors for the sake of explaining decisions to consumers. As it stands, lenders would likely have to impose this responsibility on AI scorers via contract. If such contracts shifted liability for ECOA explainability violations to the scorer, scorers

53 *Brummer* (fn. 14), p. 337.
54 15 U.S.C. § 1691(d)(1).
55 12 C.F.R. § 1002.9(b)(2).
56 15 U.S.C. § 1691a(e).

may be incentivized to develop models the decisions of which are not a "black box" to the average consumer.

The Truth in Lending Act, which regulates and standardizes the terms used to explain credit offerings in order to ensure consumers' understanding of the lending agreements they enter, also applies.[57] The Act's disclosure requirements are detailed and vary based on the specific type of credit or transaction at issue, but in general a credit card lender must disclose any mandatory minimum payments and the annual percentage rate,[58] and must also regularly update the consumer about their balance and charges. In order to prevent terms from being hidden in fine print, CFPB regulations require that these disclosures be "clear and conspicuous."[59] While the use of AI scoring over traditional scores does not change the terms used in the ultimate credit agreement, lenders employing alternative scoring should be aware of how that technology might affect the clarity of any routine disclosures.

3 The Upstart No-Action Letter

Worried that the use of its statistical model could violate ECOA and Regulation B or more general disparate impact principles, Upstart requested a no-action letter from the U.S. Consumer Financial Protection Bureau (CFPB) in 2017. The CFPB has primary regulatory authority over a range of consumer lending activities, including credit cards.[60] A no-action letter is a statement by the bureau that it has "no present intention to recommend initiation of an enforcement or supervisory action against the requester," and intended to prevent current regulations from "hinder[ing] the development of innovative financial products that promise substantial consumer benefit because, for example, existing laws and rules did not contemplate specific products."[61]

Insisting that its model does not lead to discriminatory lending practices, Upstart has compared applicant outcomes under its own model against outcomes that would result from a model using only traditional variables.[62] If an applicant scored well under the traditional model, Upstart's non-traditional vari-

57 15 U.S.C. § 1601.
58 15 U.S.C. § 1663.
59 12 C.F.R. § 226.5(a)(1).
60 Brummer (fn. 14), p. 30.
61 Policy on No-Action Letters; Information Collection, 81 Fed. Reg. 8,686 (22 February 2016).
62 See Request for a no action letter (fn. 1), p. 14.

ables would not affect the loan decision.[63] By contrast, if an applicant did not meet the minimum requirements for traditional creditworthiness but fulfilled Upstart's additional tests, he would be eligible for a loan that otherwise would have been denied or offered at higher costs. In this sense, Upstart claimed to offer an arguably fair regime: some people will be better off, and no one will be worse off than under a purely FICO score-based system. The CFPB granted Upstart the no-action letter.[64]

3.1 The Argument in Upstart

To understand Upstart's reasoning, it might be useful to recall its focus on education variables. Enrollment at an elite institution, so Upstart claims on the basis of its model, makes a higher-paying job more likely, and is therefore a natural variable to be considered by a lender.[65] Furthermore, Upstart argues, traditional scoring based on FICO variables also results in Black Americans qualifying for loans at higher interest rates than white Americans in comparable financial circumstances.[66] As with traditional lenders, using alternative data and AI models to inform credit decisions will see some level of disparate outcomes across protected classes, a phenomenon which is not unique to FinTech lenders. Upstart might point out that the fact that probability of repayment is statistically lower for Black and Hispanic Americans than for white Americans, while deplorable, reflects existing inequality. Furthermore, Upstart showed, all the "promising individuals with limited credit history"[67] are better off. No applicant is worse off than under a traditional scoring model, because the additional variables are used only if the traditional score is too low.[68]

63 Ibid., p. 3 et seq.

64 For more details see the Bureau's request for information, available at: https://files.consumerfinance.gov/f/documents/20170214_cfpb_Alt-Data-RFI.pdf (last accessed 22 January 2021) (hereinafter "Request for information").

65 A counterargument has been put forward by the Student Borrower Protection Center which claims that there is only a slight correlation between institutional selectivity and increased earnings Student Borrower Protection Center (fn. 6) p. 10.

66 See Student Borrower Protection Center (fn. 6), p. 6.

67 See Request for a no action letter (fn. 1), p. 1.

68 See ibid., p. 1: "complementing (not replacing) traditional underwriting signals." This would presuppose that if nobody is worse off, the possibility that some are being denied the opportunity to be better off based on a protected characteristic is not actionable at all. This issue will not be discussed in detail here, but see 4.1 further below.

3.2 Intentional Discrimination

Considering whether there was intentional discrimination, the CFPB noted that "[m]achine learning algorithms that sift through vast amounts of data could unearth variables, or clusters of variables, that predict the consumer's likelihood of default [...] but are also highly correlated with race, ethnicity, sex, or some other basis protected by law."[69] For example, "a variable indicating subscription to a magazine exclusively devoted to coverage of women's health issues"[70] might serve as a proxy to gender.

There is overt discrimination if the scoring agency should *explicitly* use non-traditional data involving protected categories: "If the scorer/lender is aware of this correlation," using proxies like these allows "ill-meaning lenders to intentionally discriminate and hide it behind a curtain of programming code."[71] Hence, if Upstart had intentionally used educational data in order to screen out members of protected classes (i. e. race, color, religion, sex, marital status, age, or national origin),[72] this would have constituted a violation of the ECOA.

However, due to the intricacies of machine learning models, not every scorer or lender will be aware of correlations in their model that may serve as proxies for membership in protected classes. Thus, even well-intentioned lenders relying on complex or black-box algorithms may end up working with scores which disparately impact protected groups.

Even if a scoring agency is aware of the relevant correlation, as was the case for Upstart, it does not usually focus intentionally on race. Instead, Upstart insists on only looking at correlations produced by its machine-learning algorithm on the basis of "a mix of all the variables used in Upstart's underwriting model."[73] Against this background, the fact that graduates from historically black colleges and universities (HBCUs) or Hispanic-serving institutions (HSIs) pay considerably more for a similar loan if compared to an NYU graduate is "a result of the model" – a reflection of the world as it is, out of Upstart's

69 See Request for information (fn. 64), p. 19.

70 Ibid.

71 Ibid.

72 15 U.S.C. § 1691(a)(1).

73 Upstart letter p. 4: "the model only processes variables in concert; it does not process variables in isolation"; on the discussion on HUD's interpretation of the Fair Housing Act in 2019 see *Talia B. Gillis*, False Dreams of Algorithmic Fairness: The Case of Credit Pricing, p. 10 (available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3571266 [last accessed 30 January 2021]).

reach, as it were.[74] "Such correlations are not per se discriminatory,"[75] the Bureau found.[76]

This is where one of the obstacles to applying traditional anti-discrimination laws to new technologies becomes evident. While traditional antidiscrimination doctrine asks for intentional discrimination on a basis such as race or gender, a lender relying on algorithmic scoring can point to the math behind the model, arguing that it is "the machine" making the decision. The EU Proposal addresses this "tendency of automatically relying or over-relying on the output produced by a high-risk AI system" as "automation bias."[77] Establishing intent would then require showing that the scorer (or lender) deliberately picked the offensive variable to "mask"[78] its bias, which will rarely be the case. Duties to review and backtest the models employed, on which the EU Proposal largely rests, will not suffice to establish *intentional* discrimination but could perhaps provide the basis for a claim of indirect discrimination.

3.3 Indirect Discrimination/Disparate Impact

When intent to discriminate cannot be established, the usual next step is to move on to a claim of disparate impact. This doctrine does not require the claimant to show intent but focuses on discrimination by statistical differences in aggregate outcomes across groups. Disparate impact would capture a facially neutral model that affects members of a protected group differently than members of another group.[79] The ECJ has long accepted what it calls "indirect discrimination" claims and the relevant anti-discrimination Directives incorporate this doctrine. The U.S. Supreme Court, on the other hand, is much more hesitant to do so outside of housing and employment law.[80] In 1971, the Court adopted the theory

74 Ibid. p. 18.

75 Request for information (fn. 64), p. 19.

76 The CFPB went on to state that there "may be fair lending risks" but did not elaborate in detail.

77 Art. 15 para. 4 lit. b.

78 *Solon Barocas/Andrew D. Selbst*, "Big Data's Disparate Impact", California Law Review 2016, 671, 692 et seqq.

79 In the context of credit scoring see *Gillis* (fn. 73), p. 27 et seq.

80 See U.S. Supreme Court, 30 March 2005, *Smith v. City of Jackson*, 544 U.S. 228 (2005) (affirming disparate impact claim brought under Age Discrimination in Employment Act); U.S. Supreme Court, 25 June 2015, *Texas Department of Housing and Community Affairs v. The Inclusive Communities Project*, 576 U.S. 519 (2015) (holding that disparate housing claims were cognizable under the Fair Housing Act); ECJ, 23 March 2004, *Collins*, C-138/02, ECLI:EU:C:2004:172.

of disparate impact when interpreting the Civil Rights Act of 1964[81] and the doctrine was codified in the Civil Rights Act of 1991. However, the court has since limited the doctrine by requiring plaintiffs to show discriminatory intent for some claims[82] and allowing discrimination based on bona fide occupational qualifications.[83] Today, the extent to which U.S. courts and agencies are open to applying disparate impact principles has remained an open question. While the CFPB and some U.S. courts have been open to applying disparate impact theory in the context of the ECOA, no Supreme Court guidance is available yet. Those arguing for a more narrow approach insist on the proximity between intentional discrimination and disparate impact, understanding the latter only as "an evidentiary tool used to identify genuine, intentional discrimination – to 'smoke out,' as it were, disparate treatment."[84] Others frame the principle more broadly as concerning "the consequences of [...] practices, not simply the motivation."[85] However, even such a broader interpretation of the principle does not necessarily justify a disparate impact claim which requires (i) a difference in treatment and (ii) the absence of a reasonable business rationale.[86]

In Upstart's case, establishing a disparate impact claim would, *first,* require proof that students of HBCUs and HSIs have been treated differently by Upstart's model than students of non-minority serving institutions. The reason for this is some version of a "don't compare apples with oranges" argument. Discrimination presupposes that one group has been treated differently than another group that is otherwise equal in all relevant respects. But "the devil is in the detail" in this case, particularly with regard to the assessment of what we are prepared to treat as *equal in all relevant respects*. The disparate treatment may simply reflect existing inequality. Such inequality in financial capacity, Upstart may have claimed, will have to be taken into account by a lender because he is required to run a realistic risk assessment.[87] Hence, the entire exercise is far from a mathematical one. Normative issues arise in deciding on the characteristics of the baseline population against which to compare the allegedly discriminated-against group.[88] For instance, should the "control group" consist of any-

81 U.S Supreme Court, 8 March 1971, *Griggs v, Duke Power Co.*, 401 U.S. 424 (1971).
82 U.S Supreme Court, 7 June 1976, *Washington v, Davis*, 426 U.S. 229 (1976).
83 U.S Supreme Court, 27 June 1977, *Dothard v. Rawlinson*, 433 U.S. 321 (1977).
84 U.S. Supreme Court, 29 June 2009, *Ricci* v. *DeStefano* 557 U.S. 595 (2009) (Scalia J., concurring).
85 U.S Supreme Court, 8 March 1971, *Griggs v, Duke Power Co.*, 401 U.S. 424 (1971).
86 *Gillis* (fn. 73), p. 24 et seqq.
87 In more detail at *Langenbucher* (fn. 46), 552 et seq.
88 On this point: *Gillis* (fn. 73), p. 89.

one who could objectively be interested in a loan, anyone who actually applied for a loan, or anyone who is in the exact same position save for enrollment in a HBCU or HSI? Once a plaintiff has established the relevant groups, they must show disparate treatment. The more off-the-rack and standardized a credit contract appears, the more straightforward this exercise is. By contrast, the more individualized the pricing scheme, the more complex and normatively challenging it will be to establish disparate treatment. Lastly, one will have to settle on the level of outcome disparity one is willing to accept: is a small difference in borrowing conditions acceptable? If so, how small?

Second, even if disparate impact has successfully been established, it might still be justified. The U.S. Supreme Court looks for a "business necessity" and the need for "practical business choices" underlying the disparately impactful practice.[89] Similarly, the ECJ accepts practices that are "objectively justified by a legitimate aim [if] the means of achieving that aim are appropriate and necessary."[90] Both courts ask defendants to show that there is no alternative practice available that would produce less discriminatory results.[91] Establishing a business necessity defense will usually be a very straightforward exercise, as long as the scorer/lender can show that his model, diligently developed,[92] suggests a higher statistical probability of default for the relevant group.

4 Generalizing Upstart?

4.1 The Argument that "everyone is better off"

Upstart received the first no-action letter issued by the CFPB concerning a FinTech lender in the context of disparate impact prohibitions. While some of its reasoning has to do with the specifics of Upstart's business model, the focus of this last part is to understand the extent to which the decision has more far-reaching implications. One of the charms of Upstart's model is that it offers a second chance to borrowers who are ineligible under traditional scoring models without treating other borrowers differently. Compared to a world without Up-

89 U.S. Supreme Court, 25 June 2015, *Texas Department of Housing and Community Affairs v. The Inclusive Communities Project*, 576 U.S. 519, 531–532 (2015); *Gillis* (fn. 73), 27 fn. 76, 80, 213.
90 ECJ, 14 March 2017, *G4S Secure Solutions*, C-157/15, ECLI:EU:C:2017:203 (citing EU Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation).
91 In more detail at *Langenbucher* (fn. 46), p. 554 et seq.
92 See below at 4.2.

start, no one seems to be worse off than before. Against this background, it is tempting to understand the no-action letter as relying on this (specific) business model of Upstart.

However, the CFPB made very clear that an alternative scoring model would not necessarily run afoul of its rules even if it left some borrowers worse off. "It is important to note," states the Bureau's request for information, "that to the extent alternative data or modeling techniques could help a creditor identify consumers who are *more and less* likely to default than their current credit score suggests, alternative data could in fact *decrease or increase* a given consumer's likelihood of receiving credit, or could *raise or lower* the price that any individual is offered for that credit."[93] The CFPB seems unfazed: "Though this could be seen as a detriment to consumers who are less likely to receive credit (or whose prices increase), it could also be seen as an improvement in risk assessment, which may provide greater certainty and allow a lender to increase credit availability for those who qualify. Indeed, in the longer term consumers whose credit scores understate their true risk may be better served if they do not obtain additional credit that they cannot repay."[94]

Hence, while Upstart presents a specific case in that "everyone is better off," even if the extent to which this is true still varies across white, Black and Hispanic Americans, the CFPB did not stress this argument. Instead, it explicitly embraced denying credit on the basis of an AI model, arguing that AI models deliver better predictions on the probability of repayment.

4.2 Ensuring Quality

We have said further above[95] that a business necessity defense requires a carefully and diligently developed AI scoring model. This points towards the enormously complex question of how to assess the quality of the data and of the model. Not only courts dealing with discrimination lawsuits, but also banking supervisory authorities will have to address the choice of scoring methodology and input data.[96] Following Upstart's argument, we have so far assumed that the non-traditional scoring model succeeds in producing better quality results

93 Request for information (fn. 64), p. 14.
94 Ibid.
95 See section 3.3. above.
96 *Gillis* (fn. 73), p. 49 et seq.; on the latter *Langenbucher* (fn. 46), p. 561 et seq.

than the traditional one. Indeed, in a joint statement, the CFPB and other financial regulators accepted that alternative data may "improve the speed and accuracy of credit decisions" and may allow extension of credit to those underserved in the "mainstream credit system."[97] However, there are number of potential issues to keep in mind.

4.2.1 Quality of the Data and "biased AI"

The CFPB has found that alternative data may raise "accuracy concerns because the data are inconsistent, incomplete, or otherwise inaccurate."[98] The Bureau is aware that traditional scoring models raise such concerns, too. However, because non-traditional data are not often sourced for the purpose of a credit rating, the CFPB worries that quality standards may be lower.[99]

Additionally, the regulatory framework applicable to traditional credit bureaus provides safeguards for borrowers who want to know which data has been used and/or correct mistakes.[100] The FCRA entitles credit applicants to the information in their report, and they may dispute the completeness or accuracy of that information with the CRA, which must notify furnishers of the dispute and update or delete the disputed information within 30 days.[101] In this way, inaccurate data which might have entered a scoring model can be rectified, allowing for the eventual score to more accurately reflect reality. Not all of these legal safeguards apply to non-traditional data. Even if the aforementioned rights to access the data and correct errors exist, consumers might not understand how and which data impacts their credit standing, and therefore not proceed with such claims.

The use of alternative data has the potential to complicate the FCRA regulatory scheme in a number of ways. Traditional FICO scores only require furnishers to send data relevant to the basic FICO factors, like a consumer's credit usage and repayment history, which are generally well-documented by furnishers and borrowers alike. Some alternative data points, on the other hand, are neither well documented nor well understood by consumers, making it difficult to know

97 Board of Governors of the Federal Reserve System et al., "Interagency Statement on the Use of Alternative Data in Credit Underwriting", 3 December 2019, https://files.consumerfinance.gov/f/documents/cfpb_interagency-statement_alternative-data.pdf (last accessed 22 January 2021).
98 Request for information (fn. 64), p. 16.
99 Request for information (fn. 64), p. 17.
100 Request for information (fn. 64), p. 17.
101 15 U.S.C. § 1681i(a)(1)(A).

when and how to exercise their FCRA/GDPR rights to access, challenge, and correct inaccurate information. Depending on the number of variables that factor into a model, the sheer number of data points could make the exercise of those rights impracticable. Furthermore, it may be difficult under the FCRA for furnishers to determine whether information they have provided is 'negative' and therefore requires notice to be sent to the consumer. This determination is straightforward for the traditional FICO factors, all of which have a binary set of outcomes: on-time payments are good and late payments are bad, lower credit usage is good and higher usage is bad, etc. Furnishers can easily understand these dichotomies, but may have no way of knowing whether, for example, a consumer's choice to frequent certain websites or live in a particular zip code would have a positive or negative effect on that person's creditworthiness. This also impacts the "explainability" of their credit decisions. The CFPB points out that traditional scoring agencies have been transparent about most of the input they use and about how consumers may work on behavioral changes in order to better their score.[102] This is why the ECOA expects lenders to explain why they reached an adverse credit decision in specific terms which will be comprehensible to the applicant.[103] Safeguards such as these do not work as well when dealing with alternative data. Explainability presupposes precise understanding of the model, which is not necessarily a given when black-box algorithms are used. Some FinTech lenders, such as, for instance, Underwrite.ai,[104] have started to address this issue by trying to ensure that their models' decisions come with explanations that correspond to the categories of explanation that have traditionally been given to denied applicants.[105]

If the accuracy concerns identified by the CFPB have discriminatory potential, they are often addressed as "biased AI."[106] A correlation the algorithm de-

102 Request for information (fn. 64), p. 17.

103 15 U.S.C. § 1691(d)(1).

104 Underwrite.ai also uses machine learning to generate scores, but relies only on data of past cured loans, Underwrite.ai, About Us, https://www.underwrite.ai/about (last accessed 22 January 2021).

105 Underwrite.ai, Frequently Asked Questions, https://www.underwrite.ai/faq (claiming that their model can explain exactly why it reached a lending decision in a way that its fully FCRA-compliant) (last accessed 22 January 2021).

106 *Karen Hao*, "This is how AI bias really happens—and why it's so hard to fix", MIT Technology Review (4 February 2019), https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/ (Explaining that AI is biased because models are programmed "for various business reasons other than fairness or discrimination," datasets are "unrepresentative of reality[...] or reflect[...] existing prejudices," and bias may be tects may be rooted in historical (discriminatory) data which no longer represents today's reality.[107] When this happens, the score the algorithm computes is based on an outdated legal restriction or threshold. If this restriction or threshold no longer reflects today's world, it rules out borrowers who may in fact have an attractive risk profile.

An example for quality concerns due to biased AI are gender discrimination claims. In many countries, the law required a husband's signature for his wife to take out a loan. If a woman was unmarried, even if her income was secured, this would have lowered her score. An AI trained on historical data would have "learned" that being married is "better" than being unmarried. Once the law changes, the AI not only discriminates against unmarried women, but also turns away potentially good customers, thus raising a further-reaching quality issue.

Against this background it is worth noting that in its Whitepaper on AI, the EU Commission envisages "obligations to use data sets that are sufficiently representative."[108] The Proposal on an AI Act includes more detailed provisions on data quality management.[109] However, while some such data quality issues may be easily recognized and fixed by re-training the AI, historically biased data of this type is often hidden and detected only by chance – or not at all.

4.2.2 Quality of the Model

A related but distinct problem arises when the underlying data is bias-free but the software itself suffers inadequacies that disproportionately affect certain groups. A much-cited example concerns researchers at MIT finding that Amazon's facial recognition software had more difficulty identifying the gender of female and darker-skinned faces.[110] Another example is provided by an algorithm used by a health insurance company. The model assigned risk scores on the

introduced "during the data preparation stage" when variables are selected.") (last accessed 22 January 2021).

107 *See, e.g., Alexander D'Amour et al.*, "Underspecification Presents Challenges for Credibility in Modern Machine Learning" (2020), at https://arxiv.org/abs/2011.03395 (last accessed 22 January 2021) (expressing concern that machine learning systems like national language processors rely on "shortcuts that reinforce societal biases around protected attributes such as gender.").

108 White Paper (fn. 30), p. 19, also on record-keeping.

109 See Art. 10.

110 *James Vincent*, "Gender and racial bias found in Amazon's facial recognition technology (again)", The Verge (25 January 2019), https://www.theverge.com/2019/1/25/18197137/amazon-rekognition-facial-recognition-bias-race-gender (last accessed 22 January 2021).

basis of total health-care costs per year, not taking into account that – statistically – black people went to see a doctor later than white people. This resulted in them having to be sicker before being referred to additional help. The "faulty" variable was found by mere chance, when the health insurance let a university use its data for research purposes.[111] Hence, to fully address bias, lenders utilizing AI should be aware that discriminatory decision-making may arise from biased data,[112] biased software, and the interaction between the two.

Models which rely on variables which have to do with the behavior of a borrower raise further concerns. Traditional scores address behavior which is subject to change, such as the number of credit cards used or the paying back of a loan in time. By contrast, some of the scoring agencies working with alternative data focus on behavioral clues to the borrower's personality. These agencies may often lack any interest in consumers changing the relevant behavior. If these companies rely heavily on non-traditional data such as friends on social networks,[113] fonts used in text messages[114] or performance in fitness tracking apps,[115] they may prefer that the correlations discovered between those alternative data points and credit risk retain their predictive power.

The German FinTech Kreditech provides an illustration. The company had found a strong correlation between a specific font found on electronic devices of applicants for a loan and probability of repayment. Borrowers with the specific font on their device presented a high-risk group. Kreditech has speculated that the reason for this statistical correlation is that online gambling sites use the same font.[116] Finding the font in text messages is a statistical clue that this person may engage in online gambling, which lowers their statistical probability of repaying a loan on time. An obvious data quality issue emerges: not everybody using the font will be an online gambler, and not every online gambler presents

111 See *Ziad Obermeyer/Brian Powers/Christine Vogeli/Sendhil Mullainathan*, "Dissecting racial bias in an algorithm used to manage the health of populations", Science 2019, 447.
112 See, e. g., *Philipp Hacker*, "A Legal Framework for AI Training Data", 13 Law, Innovation & Technology (forthcoming 2021) (discussing various ways bias can distort the data used to train AI models, and therefore the models themselves).
113 Report Sachverständigenrat für Verbraucherfragen (SVRV), Consumer-friendly scoring, p. 52, available at: https://www.svr-verbraucherfragen.de/wp-content/uploads/Report.pdf (last accessed 22 January 2021).
114 Ibid., p. 62.
115 Ibid., p. 101 et seq.
116 Id., p. 62; referencing *Karsten Seibel*, "Gegen Kreditech ist die Schufa ein Schuljunge", WELT, https://www.welt.de/finanzen/verbraucher/article139671014/Gegen-Kreditech-ist-die-Schufa-ein-Schuljunge.html (last accessed 22 January 2021).

a bad credit risk.[117] Let us further assume the consumer learns that his credit assessment is based (among other things) on the font he uses. He now ceases use of this font on his devices, while his online gambling habits remain the same. Because the algorithm has lost the statistical indicator, it will become less precise. Companies using non-traditional data may therefore have an incentive to not be transparent about such indicators and their relation to consumer behavioral traits. Revealing this information would open up their models to the challenge of "gaming the system."[118] The consumer changes the font but goes ahead with his gambling habit. The statistical clue is then open to manipulation, raising yet another quality issue. The same story could not be told for traditional scoring bureaus, which rely on financial indicators that can only be 'manipulated' by actually improving one's capacity for repayment. These traditional scoring models are not interested in withholding information on how consumers may better their FICO score because traditional data are not used as statistical clues, pointing towards more hidden behavioral traits, in the way the text message font pointed towards an online gambling habit.

Looking beyond the world of heavily regulated and supervised financial institutions, AI scoring also raises issues in the context of other, more predatory lending models. We have so far assumed that the AI models used are trained to "assess creditworthiness." However, this is an oversimplification. Models are used by a scorer/lender with a specific business model in mind. The lender might train his model to detect a borrower with a high likelihood of paying back a long-term loan with market interest rates. However, he might also train his model to detect borrowers who seem likely to default in the long run but show a high probability of performance over the first couple of months—perhaps at very high interest rates. In the words of the CFPB, these consumers are "more likely" to default, but this does not rule out a business model under which they may be attractive customers. Payday loan companies, for example, issue small

117 Perhaps unsurprisingly, Kreditech is insolvent, as well as its successor "monedo": *Caspar Tobias Schlenk*, Kreditech-Nachfolger: Die Hintergründe der Monedo-Insolvenz, Finance Forward, 8 September 2020, https://financefwd.com/de/monedo-insolvenz/ (last accessed 22 January 2021).

118 Strandburg and Cofone assert that disclosing the methodology of a decision-making algorithm such that its subjects can game the system is socially desirable "when the potential for socially undesirable gaming is low," and that algorithm creators may act strategically in deciding whether and what to disclose to consumers, *Ignacio Cofone/Catherine Strandburg*, "Strategic Games and Algorithmic Transparency" (Working Paper) p. 3, available at https://www.law.nyu.edu/sites/default/files/Strategic%20Games%20and%20Algorithmic%20Transparency.pdf (last accessed 30 January 2021).

loans (usually $500 or less) to be repaid in a single payment on the borrower's next payday. Payday lenders often do not consider an applicant's ability to repay, but charge fees of $10 to $30 per $100 borrowed (for reference, a $15 fee per $100 borrowed is the equivalent of a 400 % annual interest rate).[119] Though traditional FICO scores don't serve the payday lending model well, such lenders could employ AI scoring models to identify those likely to make the single lump-sum repayment on payday.

5 Conclusion

Upstart's business model and the CFPB's no-action letter have served as a useful illustration of problems in applying the traditional regulatory frameworks for credit scoring and data privacy to AI-based scoring. Despite the promise to offer more attractive credit options to traditionally underserved borrowers, alternative scoring models give rise to important risks. Some of these seem somewhat technical, but are no less salient. Such issues concern the quality of data and models used in algorithm-based credit scoring as well as the applicability of procedural safeguards such as access to data, the right to rectification of errors and to contradict the use of data, and the efficient enforcement of rights. The complex question of how to apply anti-discrimination laws shows the pitfalls of alternative scoring that aspires to create more fair lending.

Other risks are less technical. They have to do with the fairness of scoring as such.[120] What makes a scoring model "fair" is the subject of ongoing debate, and traditional scoring models also implicate fairness concerns. That said, the "unfair" label would certainly apply to models that violate the ECOA's antidiscrimination provisions, and because AI models may create more overlap between variables that predict likelihood to repay and variables correlated with membership in protected classes, those models may well raise questions of fairness. The CFPB's Request for Information on alternative credit scoring touches upon the matter very briefly when it claims that "using some alternative data, especially data about a trait or attribute that is beyond a consumer's control to change, even if not illegal to use, could harden barriers to economic and social mobility, particularly for those currently out of the financial mainstream."[121] Let us be re-

119 Consumer Financial Protection Bureau, What is a payday loan?, (2 June 2017), https://www.consumerfinance.gov/ask-cfpb/what-is-a-payday-loan-en-1567/ (last accessed 22 January 2021).
120 On the concept of "fair" scoring and on remedies and sanctions see in more detail *Langenbucher* (fn. 46), p. 527 et seq, p. 565 et seq.
121 Request for information (fn. 64), p. 18.

minded, again, that traditional credit scoring exercises have held the same potential. The underserved borrowers, those with a "thin file" who may be ineligible for traditional scores, have faced barriers to economic and social mobility for a long time. Despite its promise to serve the unbanked, AI-based scoring may well deepen this problem.

One reason for this is the opaqueness of behaviorally oriented models such as the gambling site font example set forth above. While traditional scoring models rely on variables which are open to behavioral change, such as, for instance, reducing late payment on bills, AI models that use the correlation between probability of repayment and a certain behavior may provide fewer opportunities for such change because borrowers may remain unaware of which variables influenced their credit decision, and to what degree. Lenders/scorers are not interested in disclosing the use of these variables because they wish to disallow "gaming the system." In many instances, depending on the complexity of the AI or the efforts of the lender/scorer, the scorer might not even be aware of the impact of such variables in their model.

Another reason AI scoring might deepen disparities in access to credit is the seductive allure of AI modeling which the EU has referred to as "automation bias." Many have praised machine learning for its potential to detect previously unanticipated correlations and to replace human bias when making a loan decision, relying on the "objectivity" of machines.[122] Even if just as many others have pointed to flaws in that reasoning,[123] psychological research teaches us that it can be very appealing to outsource responsibility for decision-making. In other words, when responsibility for a decision can be shared with or transferred to another person (or to a decision-making computer program), the individual sharing responsibility is less likely to work to remedy the issue than if they bore full responsibility.[124] Linking this to the quality problems of data and models, we risk overstating what an algorithm can deliver.

122 *Cass R. Sunstein*, "Algorithms, Correcting Biases", 86 Social Research 2019, 499, 504 (Finding that, "for purposes of law and policy, some of the most important empirical research finds" that algorithms are unbiased, and that "well-designed algorithms should be able to avoid cognitive biases of many kinds.").
123 *Barocas/Selbst* (fn. 78), 678; *Talia B. Gillis/Jann L. Spiess*, "Big Data and Discrimination", 86 University of Chicago Law Review 2019, pp. 459, 475.
124 *Frederike Beyer/Nura Sidarus/Sofia Bonicalzi/Patrick Haggard*, "Beyond self-serving bias: diffusion of responsibility reduces sense of agency and outcome monitoring", 12 Social Cognitive Affective Neuroscience 2017, pp. 138, 144 (concluding that the presence of other actors reduces one's sense of responsibility in remedying a problem).

This is also where the Bureau's assessment, that "in the longer term consumers whose credit scores understate their true risk may be better served if they do not obtain additional credit that they cannot repay," risks missing the point. "Understating risk" requires the modeler to define risk and to determine what the model will understand as "success."[125] It depends, as we have seen, on the quality of the data and the model, both in a narrow sense of the care with which the data/model have been sourced/developed, but also in a broader sense of biases inherent to the data or model. On the one hand, these circumstances hold a real risk of creating a new group of underserved borrowers, again ossifying existing inequalities. On the other hand, overly liberal expansion of credit to different groups could result in a crisis of indebtedness at a social scale. Effective regulation may play an important role in steering FinTech lenders clear of these extremes.

The most intricate problem linked to the fairness of scoring arises even if we assume an ideal world in which all data could be de-biased and lenders could efficiently screen all models for discriminatory effects. Linking credit outcomes to behavioral traits increases the risk that a model will reproduce and even worsen deeply embedded social biases and inequalities. Not only one's ZIP code or payment history, which are – at least in theory – subject to change, but also one's hobbies or friends, taste in restaurants or shopping habits, efficiency in filling out a web form, model of smartphone or amount of spelling mistakes, age or health might be considered predictive of success. Graduating from a HBCU could hurt an applicant's chances if compared to graduating from a non-minority-serving institution, as could a preference for budget supermarkets as opposed to more expensive organic grocers, using an Apple instead of an Android smartphone,[126] going online during the day or at night, using price comparison websites or not,[127] and the list goes on. A lack of disclosure and explainability undermine the applicant's opportunities to learn from a credit decision and adapt their

125 *Cathy O'Neill*, Weapons of Math Destruction, 2016, p. 21 et seqq.
126 See *Marianne Bertrand/Emir Kamenica*, "Coming Apart? Cultural Distances in the United States Over Time", National Bureau of Economic Research, Working Paper No. 24771, 2018 (showing research relating brand of phone owned with income).
127 See *Tobias Berg/Valentin Burg/Ana Gombović/Manju Puri*, "On the Rise of FinTechs—Credit Scoring Using Digital Footprints", Michael J. Brennan Irish Finance Working Paper Series, Paper No. 18–12, 2019 ("For example, customers coming from a price comparison website are almost half as likely to default as customers being directed to the website by search engine ads", p. 3).

behavior accordingly. ECOA's central tenet, to offer equal credit opportunities, seems severely compromised in such circumstances.[128]

6 Post Scriptum: AI Credit Scoring under the EU Proposal for an AI Act

The European Commission has on 21st April 2021 published a Proposal for a Regulation to lay down harmonized rules on artificial intelligence. Once passed, a Regulation is binding law in every Member State. In contrast to a Directive, it is directly applicable without the need to be transposed, Art. 288 para. 2 TFEU. However, this pan-European scope is not the only reason for a closer look at the Proposal. Its rules will apply to providers of AI systems within the EU as well as in third countries such as the United States. It will cover users of AI systems in third countries if the output they produce is used in the EU. Most importantly, the Proposal aims to shape global norms and standards. Given a trend called the "Brussels effect,"[129] an observation on how multinational companies have progressively adopted European standards on, for instance, data privacy, consumer safety, and antitrust, the claim to contribute once again in this fashion might not be entirely without merit and companies may well follow suit on AI regulation as well.

The Proposal should be understood against the context of a number of previous studies and official documents, such as the report of the High-Level Expert Group on AI on "Ethics Guidelines for Trustworthy AI" which we mentioned above. Additionally, in February 2020 the EU Commission published the White Paper on AI, accompanied by a Report on safety and liability implications. The European Parliament adopted resolutions on civil liability for AI, on an ethical framework for AI, and on related issues of intellectual property. Next steps for the Proposal to become law include first and second readings by the co-legislating bodies, the European Parliament and the Council, internal debates in Member States' national parliaments, and European and global lobbying efforts.

The Proposal starts from the assumption that several fundamental principles and standards apply horizontally across all AI use cases. Among these use cases, the Proposal singles out unacceptable and high-risk applications. Unacceptable

128 Once again, thoughts on potential solutions and preferable approaches to regulatory governance in this area are reserved for future papers. See also *Langenbucher* (fn. 46), 527 et seq, 565 et seq.

129 *Anu Bradford*, The Brussels Effect: How the European Union Rules the World, 2020.

use cases will be prohibited. For high-risk applications, the Proposal prescribes a variety of requirements of *ex ante* testing, certification, technical documentation, and monitoring, as well as *ex post* controls. By contrast, the Proposal explicitly encourages AI applications which qualify as neither unacceptable nor high-risk, posing only low or minimal risk. For these uses, the Proposal seeks to ensure an attractive environment for investment by combining legal certainty with effective enforcement and allowing for regulatory sandboxes while preventing market fragmentation.

6.1 A Risk-Based Approach

The drafters of the Proposal chose what they call a "risk-based approach." They frame this approach as the best answer to the tension between "promoting the uptake of AI and of addressing the risks associated with certain uses of such technology." According to the Proposal, risks are unacceptable, if they are "manipulative, exploitative and social control practices." Such risks "contradict Union values" and will be prohibited. Against this background, Art. 5 para. 1 of the Proposal lists AI practices which qualify as unacceptable. Among these we find the use of biometric identification systems in publicly accessible spaces for the purpose of law enforcement. The same goes for some cases of social scoring by public authorities, as set forth in Art. 5 para. 1 lit. c of the Proposal.

While few AI use cases are considered unacceptable and prohibited, the list of high-risk applications is longer. One Annex to the Proposal enumerates products for which Union legislation as to safety precautions is already in place, covering products as diverse as toys, explosives, medical devices, and civil aviation. AI systems which are used as safety components in such products will be considered high-risk. A second Annex to the Proposal lists areas of use, rather than products. These include biometric identification, operation of critical infrastructure, employment, access to essential private services, law enforcement, migration and administration of justice.

AI systems that "evaluate the creditworthiness of natural persons or establish their credit score" are listed as one instance of access to "essential private services." Interestingly, neither the High Level Expert Report nor the White Paper had taken up decisions about creditworthiness. The policy reasons for including AI scoring in the Proposal surface in Recital (37). Starting from the fundamental role of access to financial resources, the Proposal stresses the much-debated risks of AI scoring. These include "discrimination of persons or groups," dangers that these applications may "perpetuate historical patterns of discrimination, for example based on racial or ethnic origins, disabilities, age, sexual orientation," and the creation of "new forms of discriminatory impacts." Of course, the EU, like the United States, possesses a large body of law which prohibits discrimination – and in both jurisdictions, courts and scholars have been grappling with the challenges raised by the "new forms of discriminatory impacts" which the Proposal cites. Still, the document does not explicitly take up questions of algorithmic fairness, historic bias or discrimination as such. Instead, its approach brings product design to mind: certification procedures, data and model quality checks, technical documentation and *ex post* monitoring duties abound. Public authorities supervise, but private enforcement instruments are not included. This fundamental tension between the anti-discriminatory policy goal and the product-oriented, formalistic regulatory design shapes the Proposal.

6.2 Applying the EU Proposal to Algorithmic Credit Scoring

6.2.1 How to Distinguish High-Risk Credit Scoring from Prohibited Social Scoring

Algorithmic scoring has raised enormous concerns globally insofar as it is used for surveillance of private citizens, a practice usually addressed as "social scoring." The Proposal defines specific forms of social scoring which would be prohibited in the EU. These are:

"AI systems (placed on the market put into service or used) by public authorities or on their behalf for the evaluation or classification of the trustworthiness of natural persons over a certain period of time based on their social behaviour or known or predicted personal or personality characteristics, with the social score leading to either or both of the following:

(i) detrimental or unfavourable treatment of certain natural persons or whole groups thereof in social contexts which are unrelated to the contexts in which the data was originally generated or collected;
(ii) Detrimental or unfavourable treatment of certain natural persons or whole groups thereof that is unjustified or disproportionate to their social behaviour or its gravity;[130]

130 Art. 5.

Under this definition, AI credit scoring done by private entities does not qualify, unless performed on the behalf of a public authority. To the extent that public authorities engage in any form of credit scoring, application of the provision hinges on what "trustworthiness" entails. The Proposal does not offer a definition nor explains how "trustworthiness" differs from "creditworthiness." Recital (17), which sets out the policy goal for the prohibition, talks about "social scoring of natural persons for general purpose" and of "detrimental or unfavourable treatment of natural persons or whole groups thereof in social contexts." While it seems intuitive to understand creditworthiness as a sub-category of the more general term "trustworthiness," the Proposal seems to have a different, namely a "social" context in mind. Arguably, future work on the Proposal would profit from a brighter line between trustworthiness and creditworthiness, and between extending credit and "treatment (…) in social contexts" as listed above under (ii). Should these provisions apply to AI credit scoring, the use of alternative data like that posited above to generate credit scores could be found to constitute the use of data in contexts "unrelated to the contexts in which the data was originally generated or collected."

6.2.2 How to Ensure Compliance with the Proposal

Leaving public authorities (or work done on their behalf) aside, AI models intended for creditworthiness assessments and credit scoring qualify as high-risk, Art. 6 para. 2, Nr. 5 b Annex III. AI use cases which qualify as high-risk have to comply with the Proposal's risk and quality management framework. The Proposal follows an omnibus approach across all areas of AI applications, including medical, law enforcement, machinery and credit scoring.

This horizontal, omnibus approach differs markedly from the U.S. regulatory framework we have outlined above, which works with application area-focused legal rules such as, for instance, the ECOA, the FCRA, and the HUD. While the Proposal's approach offers legal security across different use cases, its requirements must be tailored to a variety of AI applications. The Proposal somewhat vaguely suggests doing so with the "intended purpose of the high-risk AI system and the risk management system" in mind.[131]

131 Art. 8 para. 2.

6.2.2.1 Risk Management and Quality Management Systems

An adequate risk management system is one of the core pillars of the Proposal. Such a system "shall consist of a continuous iterative process run throughout the entire lifecycle of a high-risk AI system, requiring regular systematic updating." Risks have to be identified and analyzed, estimated and evaluated.[132] Risk management concerns "known and foreseeable risks" as well as "risks that may emerge when the high-risk AI system is used in accordance with its intended purpose and under conditions of reasonably foreseeable misuse."[133] A post-marketing system is added,[134] and residual risks have to be "judged acceptable" and "shall be communicated to the user."

The need to adapt these general, omnibus requirements to the specifics of AI scoring systems surfaces clearly. "Risk" will come in very different shapes and forms across different AI use cases. As to AI scoring and creditworthiness assessments, the Proposal seems to understand "risk" as related to fundamental rights, and, more specifically, to discriminatory outcomes. However, "risk" is as vague a term as "fundamental rights" or "non-discrimination." How to apply non-discrimination doctrine to AI scoring is as hotly debated in the EU as in the U.S. Some of the relevant concerns that this ambiguity creates have surfaced in our discussion of the Upstart case. The decision to go ahead with this approach illustrates the built-in tension and the ambitiousness of the decision to use a formal, product-oriented regulatory design in order to realize substantive goals such as non-discrimination.

In addition to a risk management system, the Proposal requires "providers," the developers of AI systems, to ensure that compliance and quality management systems are in place,[135] Art. 16, 17, and that conformity assessments are undergone. Written documentation of the quality management system is expected, including, for instance, a strategy for regulatory compliance, test and validation procedures, procedures for data management, for post-market monitoring and for communication with national competent supervisory authorities, as well as an accountability framework setting out the responsibilities of management and staff. If the provider is a credit institution regulated by Directive 2013/36/EU ("CRD IV"), the obligation to put a quality management system in place is deemed to be fulfilled by complying with Art. 74 of CRD IV. Post-market monitoring is required of any provider under Art. 61, and is thus not limited to high-risk

132 Art. 9 para. 2 (a), (b).
133 Ibid. para. 2 (b).
134 Art. 9 para. 2 (d), Art. 61.
135 Art. 16, 17.

systems. However, Art. 62 provides that only high-risk providers have an obligation to report malfunctions to market surveillance authorities.

6.2.2.2 Data and Data Governance

Data and data governance are a core ingredient of AI scoring applications. Art. 10 of the Proposal lays down quality criteria for training, validating and testing with data sets. These concern design choices, data collection and preparation, the formulation of assumptions, examination of biases, the identification of gaps and more.

Data sets must be "relevant, representative, free of errors and complete" and have "appropriate statistical properties."[136] Again, the terms used are vague and need further interpretation. A data set will probably never be "free of errors" nor "complete." The relevance of data is often in the eye of the beholder, and it would be useful to further specify what may count as "representative."

A conscious choice has been made as to bias monitoring. While the extent to which one may use protected categories of data such a "race" in order to uncover bias is subject to debate under the U.S. framework. By contrast, the Proposal allows processing of such data if it "is strictly necessary for the purpose of ensuring bias monitoring, detection and correction."

6.2.2.3 Technical Documentation and Record-keeping, Accuracy, Robustness, and Cybersecurity

A number of requirements concern technical documentation, record keeping and conformity assessments. Technical documentation must be drawn up *ex ante*, and kept up to date. Logs for the automated recording of events have to be installed,[137] and kept by the providers.[138] Additionally, high-risk systems have to achieve an appropriate level of accuracy, robustness, and cybersecurity.[139] Machine learning applications, a standard feature of many AI scoring systems, have to address feedback loops which the Proposal defines as "possibly biased outputs due to outputs used as input for future operations."[140]

136 Art. 10 para. 3.
137 Art. 11, 12.
138 Art. 20.
139 Art. 15.
140 Art. 15 para. 3.

Art. 16, under the heading of "obligations of providers," lists these as well as other requirements specified in the Proposal. The Proposal does not address whether this is to be understood as an obligation giving rise to a private right of action, leaving the matter to the national law of the Member States. If the provider is established outside the Union and an importer cannot be established, the third country provider must establish an authorized representative.[141]

6.2.2.4 Transparency and Information

Seen from a U.S. perspective, informing retail borrowers about the data used for scoring, explaining basic workings of the scoring model and allowing for rectification constitute core elements of credit scoring regulation. By contrast, in the EU not only credit scoring regulation as such but also the institutional set-up of relevant scoring institutions differs between Member States. Taking this together with the EU GDPR providing for a reasonable degree of data protection (including rectification and some explainability), it is maybe unsurprising that there are no provisions in the Proposal on how to inform a borrower.

Importantly, transparency and the provision of information to "users," which Art. 13 requires, is not about informing borrowers. "Users," as defined in Art. 3 para. 4, means "any natural or legal person, public authority, agency or other body using an AI system under its authority." The end consumer (or "end borrower") is not herself "using" the AI system. She is rather, as it were, its object. Obligations towards this group of end consumers are limited to a number of specific instances such as, for example, emotion recognition, biometric categorization, or systems creating deep fakes.[142]

The "users" toward whom the AI system must be transparent are those who employ the system in their own business. This could be a lender, who uses an AI system for its own rating of borrowers. It could also be a scoring agency using AI systems as part of its scoring process. These users of AI systems are the beneficiaries of the duties of disclosure which the Proposal imposes on providers. Providers must furnish information on, among other things, the intended purpose of the AI system, the level of accuracy, potential risks for fundamental rights, the expected lifetime of the system and human oversight measures.

141 Art. 25.
142 Art. 52.

Users of high-risk systems have to comply with the instructions of use which includes certain monitoring instructions. Logs must be kept, if these are under the control of the user (not the provider).[143]

6.2.2.5 Human Oversight

The Proposal requires that high-risk AI systems feature an element of human oversight. Art. 14 explicitly mentions risks to fundamental rights, which have caused growing concern in AI scoring systems, and assumes that human oversight can prevent or minimize these risks.[144] Human oversight is to serve a monitoring function, allowing for detection of "dysfunctions and unexpected performance."[145] This section also addresses "automation bias,"[146] and requires that AI-based decision making systems leave open the possibility of foregoing use of the AI application in a particular case.[147]

6.2.3 Enforcement

The Proposal relies heavily on public enforcement of its regulations. Member States have to designate a notifying authority to carry out the conformity assessments required by Art. 30 of the Proposal, and to issue certificates of compliance under Art. 44. The Proposal stresses that these bodies must be competent, independent, objective and impartial.[148] Art. 48 requires that providers draw up declarations of conformity for AI systems they put on the market. The product-design framework of the Proposal is especially obvious in Art. 48 and 49, which address an EU declaration and a CE marking of conformity.[149] Art. 71 sets forth a framework for administrative sanctions including fines and penalties for non-compliance.

143 Art. 29 para. 5.
144 Art. 14 para. 2.
145 Art. 14 para. 4 (a).
146 Art. 14 para. 4 (b).
147 Art. 14 para. 4 (d).
148 Art. 33 para. 5.
149 The "CE" mark indicates that products traded in the European Economic Area meet the EU's safety, health, and environmental protection standards. European Commission, "CE marking" https://ec.europa.eu/growth/single-market/ce-marking_en (last accessed 5 August 2021).

National supervisory authorities are in charge of market surveillance.[150] An exception from the omnibus approach is made for financial institutions. If the AI system is used or placed on the market by a financial institution, the relevant financial supervisory authority is competent to regulate the system's use, mirroring the jurisdiction of specialized bodies such as the CFTC and the FTC in the United States.

The intent to allow for and support innovation is behind the regulatory sandbox regime established in Art. 53. Competent authorities are encouraged to establish controlled environments which facilitate the development and testing of new AI systems. Art. 54 grants exemptions from the GDPR's prohibition of data processing in these cases. Small-scale providers and start-ups get priority access to sandboxes under Art. 55.

6.3 Conclusion

By issuing Upstart its no-action letter, the CFPB acknowledged that the commercial lending industry, like many other fields, may be imminently and fundamentally changed by the introduction of artificial intelligence and machine learning technologies. These technologies hold legitimate promise for extending credit opportunities to those excluded by traditional credit scoring methodologies, but their complex, data-driven nature necessarily creates difficulties in the application of regulations attuned to more traditional methods of credit scoring. The data and algorithms used by FinTech lenders may replicate discriminatory outcomes. The complexity of AI models may limit the modeler's ability to anticipate and account for unintended disparate outcomes as well as the applicant's capacity to understand adverse decisions. A combination of updated regulations and careful use of AI by these lenders may go far in addressing these issues. However, CFPB's ready acceptance of Upstart's model—which relies heavily on education level, a variable with particularly great potential to introduce bias—raises questions of whether the regulatory environment is prepared for AI credit scoring to eventually permeate consumer lending practices. Time will tell whether more ambitious AI scoring methods, such as those relying on novel cell phone and search history data, gain sufficient prominence to warrant targeted changes to the regulatory frameworks in the United States.

The EU Proposal for an AI Act has ventured a first step in providing a regulatory framework specific to AI applications. The Proposal highlights the perils of

150 Art. 63.

bias and discrimination, and its risk-based approach takes up core quality and risk management issues. Needless to say, these will have to be adapted to the specifics of each AI application. Due to the omnibus approach underlying the Proposal, there is almost no guidance as to how different use cases (ranging from civil aviation to medical devices and credit scoring) would be treated. More importantly, there is a fundamental tension between the Proposal's policy goal to protect fundamental human rights and its risk-based philosophy. For financial institutions, the possibility to measure and evaluate risk with an eye on capital adequacy requirements is crucial. Translating the relative weights of conflicting human rights principles into computable variables of risk management is a daunting task. It remains to be seen whether, in its final form, the Proposal will include more concrete rules on risk and quality management.