16

Artificial Intelligence and Financial Services

Katja Langenbucher

16.1 INTRODUCTION

"Information processing," "decision-making," and "achievement of specific goals." These are among the key elements defining artificial intelligence (AI) in a JRC Technical Report of the European Commission.[1] Information processing is understood as "collecting and interpreting inputs (in the form of data)," decision-making as "taking actions, performance of tasks (…) with certain level of autonomy," and achievement of specific goals as "the ultimate reason of AI systems."[2] All these elements play a key role in financial services. Against this background, it is unsurprising that AI has started to fundamentally change many aspects of finance.

The actors that are active in the financial world process vast amounts of information, starting from customer data and account movements over market trading data to credit underwriting or money-laundering checks. As the earlier definition suggests, it is one thing to collect and store these data, and yet another challenge to interpret and make sense of them. Artificial intelligence helps with both, for example by checking databases or crawling the internet in search of relevant information, by sorting it according to predefined categories, or by finding its own sorting parameter. In this way, AI provides input to decision-making of financial institutions, of financial intermediaries such as a broker or investment adviser, and of regulatory agencies, monitoring financial institutions and markets such as the US SEC or the European Central Bank.

Today, decision-making based on AI preparatory work often involves human actors. However, the spectrum of tasks that can be wholly or partly performed by AI is growing. Some of these tasks are repetitive chores such as a chatbot used in customer communication or a robo-advisor suggesting how to best invest. Others

[1] Samoili Sofia et al., *AI Watch. Defining Artificial Intelligence. Towards an Operational Definition and Taxonomy of Artificial Intelligence* (JRC Publications Repository, 2020) 8, https://publications.jrc.ec.europa.eu/repository/handle/JRC118163, accessed July 22, 2024 (with further references on each element, a fourth element listed in the report (but less salient for financial services) is the perception of the environment).

[2] Ibid.

https://doi.org/10.1017/9781009367783.020 Published online by Cambridge University Press

require enormous speed, for instance, high-frequency algorithmic trading of financial instruments, reacting in split seconds to new market information.[3]

AI involves a goal or a "definition of success"[4] which it is trained to optimize. A regulatory agency tasked with monitoring insider trading might employ an AI system to search market data for suspicious trades. The agency predefines what it understands as suspicious, for instance, large sales right before bad news is released to the market, and supervises what the AI system finds, to make sure it gets it right. With more sophisticated AI, regulators train the AI system to learn what a suspicious trade is. The Italian Consob, together with a group of researchers, has explored unsupervised machine learning of this type, allowing it to "provide an indication on whether the trading behavior of an investor or a group of investors is anomalous or not, thus supporting the monitoring and surveillance processes by the competent Authority and the assessment of the conduct."[5]

There is a broad range of ethical issues when employing AI in financial services. Many of these are not entirely novel concerns, but AI might make the risks they entail more likely to happen. As the OECD has noted in a report on AI, a tactic called "tacit collusion" to the detriment of market competition might become easier.[6]

> In a tacitly collusive context, the non-competitive outcome is achieved by each participant deciding its own profit-maximizing strategy independently of its competitors. (…) The dynamic adaptive capacity of self-learning and deep learning AI models can therefore raise the risk that the model recognizes the mutual interdependencies and adapts to their behavior and actions of other market participants or other AI models, possibly reaching a collusive outcome without any human intervention and perhaps without even being aware of it.[7]

Cyber security threats count among these risks:

> While the deployment of AI does not open up possibilities of new cyber breaches, it could exacerbate pre-existing ones by, *inter alia*, linking falsified data and cyber breaches, creating new attacks which can alter the functioning of the algorithm through the introduction of falsified data into models or the alteration of existing ones.[8]

[3] Jasmina Arifovic, Xuezhong He, and Lijian Wei, "High frequency trading in FinTech age: AI with speed" (2019), https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2771153, accessed July 22, 2024.

[4] Cathy O' Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Penguin Books, 2016) 21.

[5] Consob et al., "A machine learning approach to support decision in insider trading detection" (December 5, 2022), www.consob.it/documents/1912911/1933915/FinTech_11.pdf/eebb010d-e5e8-9f75-9e77-b2a1407e418f, accessed July 22, 2024.

[6] OECD, "Artificial intelligence, machine learning and big data in finance, opportunities, challenges and implications for policy makers" (2021) 40, www.oecd-ilibrary.org/docserver/98e761e7-en.pdf?expires=1721652654&id=id&accname=guest&checksum=0F3E9BBBAD171D31FCB2E11F4BE3D33C, accessed July 22, 2024.

[7] Ibid.

[8] Ibid., 38.

https://doi.org/10.1017/9781009367783.020 Published online by Cambridge University Press

The same goes for data protection. This has long been identified as a core policy concern in the digitalized world.[9] Black box algorithms compound the problem when consumers are not only uncertain that data are collected but do not know what an AI system will make of the information it processes:

> These systems run historical data through an algorithm, which then comes up with a prediction or course of action. Yet often we don't know how such a system reaches its conclusion. It might work correctly, or it might have a technical error inside of it. It might even reproduce some form of bias, like racism, without the designers even realising it.[10]

For financial services, the complicated interplay between data protection, biases of AI models, and big data available to be processed at low cost is of particular concern. The EU has selected credit scoring and creditworthiness assessments of natural persons as a "high-risk AI system," facing strict compliance requirements.[11] Additionally, the (reformed) Consumer Credit Directive[12] engages with consumer rights whenever an AI system produces a score.

In what follows, this chapter takes AI scoring and creditworthiness assessments as an example of how AI is employed in financial services (Section 16.2), for the ethical challenges this raises (Section 16.3), and for the legal tools that attempt to adequately balance advantages and challenges of this technique (Section 16.4). It closes with a look at scoring beyond the credit situation (Section 16.5).

16.2 AN ILLUSTRATION: AI-BASED CREDITWORTHINESS EVALUATIONS AND CREDIT SCORING[13]

A financial institution that hands out a loan and prices interest rates must first conduct an assessment of the borrower's credit risk. This is an evident business rationale and is required by several laws. Some of these have the overall stability of the

[9] Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (Public Affairs, 2019).

[10] Tom Cassauwers, "Opening the 'black box' of artificial intelligence" *Horizon* (December 1, 2020), https://ec.europa.eu/research-and-innovation/en/horizon-magazine/opening-black-box-artificial-intelligence, accessed July 22, 2024.

[11] Regulation (EU) 2024/1689 of the European Parlament and of the Council laying down harmonised rules on artificial intelligence of June 13, 2024 (Artificial Intelligence Act, AI Act) and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828, Art. 6(2), Annex III No. 5(b).

[12] Directive (EU) 2023/2225 of the European Parliament and of the Council of October 18, 2023 on credit agreements for consumers and repealing Directive 2008/48/EC (Directive (EU) 2023/2225).

[13] In what follows, I draw on a working paper of mine: Katja Langenbucher, "Consumer Credit in The Age of AI – Beyond Anti-Discrimination Law" ECGI Law Working Paper N° 663/2022 (November 2022), www.ecgi.global/working-paper/consumer-credit-age-ai—beyond-anti-discrimination-law, accessed July 22, 2024. In the following text, I sometimes include literal quotes from my working paper. For better readability, I do not use quotation marks.

https://doi.org/10.1017/9781009367783.020 Published online by Cambridge University Press

financial system in mind. To reduce risk, they attempt to ensure that financial institutions have clearly established procedures to hand out credit and monitor credit portfolios.[14] Other laws focus on both, financial stability, and the creditor. Following the financial crisis of 2008, irresponsible lending practices on mortgage markets had been identified as a potential source of the crisis.[15] Reacting to this concern, EU legislators aimed at restoring consumer confidence.[16] After the pandemic, and fueled by concerns about increasing digitalization, the Proposal for a Consumer Credit Directive explicitly stresses that the assessment of the creditworthiness of the borrower will be done "in the interest of the consumer, to prevent irresponsible lending practices and overindebtedness."[17]

16.2.1 *Traditional Methods to Predict Credit Default Risk*

When going about an evaluation, the lender faces uncertainty about an applicant's credit default risk. In the parlance of economics, he must rely on observable variables to reconstruct hidden fundamental information.[18] Sociologists add the role of trust in social relations to explain the denial or success of a loan application.[19] The potential borrower will provide some information himself, for instance on existing obligations, income, or assets. To reduce uncertainty, the lender will often require additional input. Historically, a variable as qualitative and vague as "character" was "considered the foundation of consumer creditworthiness."[20] Starting in the 1930s, lenders profited from advances in statistics which allowed to correlate attributes of

[14] Directive 2013/36/EU of the European Parliament and of the Council of June 26, 2013, on access to the activity of credit institutions and the prudential supervision of credit institutions and investment firms, amending Directive 2002/87/EC and repealing Directives 2006/48/EC and 2006/49/EC [2013] OJ L176/38, Art. 79.

[15] Directive 2014/17/EU of the European Parliament and of the Council of February 4, 2014, on credit agreements for consumers relating to residential immovable property and amending Directives 2008/48/EC and Regulation (EU) No 1093/2010 [2014] OJ L60/34, Recital (4).

[16] Directive 2014/17/EU, Recitals (3), (5), Arts. 18, 20.

[17] Commission, COM(2021)347 final, Art. 18(1).

[18] Robert Bartlett et al., "Consumer-lending discrimination in the FinTech era" (2022) *Journal of Financial Economics*, 143: 30; Dagobert Brito and Peter Hartley, "Consumer rationality and credit cards" (1995) *Journal of Political Economy*, 103: 400; Christine Parlour and Uday Rajan, "Competition in loan contracts" (2001) *The American Economic Review*, 91: 1311; Joseph Stiglitz and Andrew Weiss, "Credit rationing in markets with imperfect information" (1981) *The American Economic Review*, 71: 393; see Alya Guseva and Akos Rona-Tas, "Uncertainty, risk, and trust: Russian and American credit card markets compared" (2001) *American Sociological Review*, 66: 623 on uncertainty and institutions which allow for reducing uncertainty to measurable risk.

[19] Akos Rona-Tas and Alya Guseva, "Consumer credit in comparative perspective" (2018) *Annual Review of Sociology*, 44: 55; Guseva and Rona-Tas (n 18).

[20] Josh Lauer, *Creditworthy: a history of consumer surveillance and financial identity in America* (Columbia University Press, 2017) 199ff on the five variables used by the mail-order firm Spiegel in the 1930s; tracing the historical development: Danielle Citron and Frank Pasquale, "The scored society: Due process for automated predictions" (2014) *Washington Law Review* 89(1): 8ff.

https://doi.org/10.1017/9781009367783.020 Published online by Cambridge University Press

individual loan applicants with high or low credit default risk.[21] Depending on the country, the relevant characteristics "can include a wide variety of socioeconomic, demographic, and other attributes or only those related to credit histories."[22] Being a white, middle-aged man with a stable job and living in wholly owned property usually predicted a lower credit-default risk than being a young man living in a shared flat, an unmarried woman, or a person of color without a stable job.

The exercise requires two main ingredients: mathematical tools and access to data. The former serves to form statistical buckets of similarly situated persons and to correlate individual attributes of loan applicants with attributes that were in the past found to be relevant predictors of credit default risk. The latter was initially provided by lenders compiling their own data, for instance on credit history with the bank, loan amounts, or in- and outgoing payments. Later, credit registries spawned that are, until today, a key source of data.[23] The type of data a credit registry collects is highly standardized, but varies across countries. Examples are utilities data, late payments, number of credit cards used, collection procedures, or insolvency proceedings.[24] Using the available input data, the probability of credit default risk is often expressed in a standardized credit score.[25]

16.2.2 *AI-Based Methods to Predict Credit Default Risk*

Over the last decade, both ingredients,[26] mathematical tools and access to data, have changed radically. Digitization across many areas of life leaves digital footprints of consumers. Collecting and processing these "big data" allows us to refine statistical output that can now work with large amounts of information, far beyond what traditional credit registries hold. Artificial intelligence in the form of machine learning helps to correlate variables, interprets results, sometimes learns from these, and, in that way, finds ever more complex patterns across input data.[27]

"All data is credit data"[28] is an often-quoted remark, hinting at the potential to use unanticipated attributes and characteristics toward a prediction of credit

[21] Rona-Tas and Guseva (n 19) 61; Lauer (n 20) 210ff.

[22] Rona-Tas and Guseva (n 19) 62.

[23] Ibid., 61–62; see WorldBank, "Doing business project on the public credit registry coverage," https://data.worldbank.org/indicator/IC.CRD.PUBL.ZS, accessed July 22, 2024, showing an enormous increase of data collection.

[24] World Bank Group, "Credit scoring approaches guidelines" (2019) 10, https://thedocs.worldbank.org/en/doc/935891585869698451-0130022020/original/CREDITSCORINGAPPROACHESGUIDELINESFINALWEB.pdf, accessed July 22, 2024.

[25] Rona-Tas and Guseva (n 19) 62; as an overview, see the World Bank Group (n 24).

[26] See Section 16.2.1.

[27] World Bank Group (n 24) 16–17.

[28] See Quentin Hardy, "Just the facts. Yes, all of them." *The New York Times* (March 25, 2012), https://archive.nytimes.com/query.nytimes.com/gst/fullpage-9A0CE7DD153CF936A15750C0A9649D8B63.html, accessed July 22, 2024; discussion at Emily Rosamond, "'All data is credit data,' reputation, regulation and character in the entrepreneurial imaginary" (2016) *Paragrana*, 25: 112.

https://doi.org/10.1017/9781009367783.020 Published online by Cambridge University Press

default risk. This starts with an applicant's online payment history or performance on a lending platform but does not stop there.[29] Age or sex, job or college education, ZIP code, income, or ethnic background can all be relevant predictors. Depending on a jurisdiction's privacy laws, more variables can be scrutinized. Such "alternative data" include, for instance, preferred shopping places, social media friends, political party affiliation, number of typos in text messages, brand of smartphone, speed in clicking through a captcha exercise, daily work-out time, or performance in a psychometric assessment. All these are potential sources for correlations, the AI system will detect between individual data points, and the goal it is optimizing. An example of an optimization goal (which is also called "definition of success"[30]) is credit default. An AI system working with this definition of success is useful for a lender who wishes to price his loan according to the probability of default. The system will find various correlations between the big data input variables and the optimization goal. Some correlations might be unsurprising, such as a high steady income and low default. Others will be more unexpected. The German company Kreditech illustrates this. It learned that an important variable its AI system found was a specific type of font on an applicant's electronic devices.[31] A research study provides another illustration by concluding that:

> customers without a financial application installed on their phones are about 25% more likely to default than those with such an app installed. (…) In contrast, those with a mobile loan application are 26% more likely to default.[32]

16.2.3 *Inclusion Through AI-Based Methods*

A score that an AI model develops based on alternative data can provide access to finance for persons who have found this difficult in the past, due to their unusual profile. A recent immigrant will often not be able to provide a utility payment or a credit history in his new country of residence. However, this "thin-file applicant" might have relevant alternative data of the type described earlier to support his creditworthiness assessment.

[29] For the following examples, see Langenbucher (n 13) 3.

[30] O'Neil (n 4) 21.

[31] See Katja Langenbucher and Patrick Corcoran, "Responsible AI credit scoring – a lesson from Upstart.com" (2022) *European Company and Financial Law Review*, 5(141): 165ff on this example. The reason for this unanticipated correlation – as they later figured out – was that this unusual font was used by an online gambling site, see Sachverständigenrat für Verbraucherfragen, "Consumer-friendly scoring" (2018) 62, https://fragdenstaat.de/dokumente/238777-report-consumer-friendly-scoring/, accessed July 22, 2024.

[32] Sumit Agarwal et al., "Financial inclusion and alternate credit scoring: role of big data and machine learning in FinTech" (2021) 4, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3507827, accessed July 22, 2024.

https://doi.org/10.1017/9781009367783.020 Published online by Cambridge University Press

Several empirical studies have found that AI-based credit scoring broadens financial access for some thin-file applicants.[33] One important source of data are mobile phones. "The type of apps installed or information from call log patterns," so the authors of one study find, "outperforms a traditional model that relies on credit score."[34] Another study, working with an e-commerce company, produced good predictions using only ten digital footprint variables.[35] The authors find, for example, that the difference in default rates between customers using an Apple and customers using an Android device is equivalent to the difference in default rates between a median credit score and the 80th percentile of the credit score.[36] Yet another illustration is provided by the US online lending company Upstart.[37] Upstart claims to outperform traditional scoring as to all borrowers, specifically as to those with traditionally low credit scores. A study on this company shows that:

> more than 30% of borrowers with credit scores of less than 680 funded by Upstart over our sample period would have been rejected by the traditional model. We further find that this fraction declines as credit score increases, that is the mismatch between the traditional and the Upstart model is magnified among low-credit score borrower.[38]

A US regulatory agency, the Consumer Financial Protection Bureau, investigated Upstart's business model and confirmed that, in the aggregate, applicants with low credit scores were approved twice as frequently by Upstart if compared with a hypothetical lender.[39]

16.3 ETHICAL CONCERNS AROUND AI-BASED METHODS TO PREDICT CREDIT DEFAULT RISK

16.3.1 *Algorithmic Discrimination*

The US lender Upstart illustrates not only how AI can further inclusion. Additionally, it provides an example of AI models producing unequal output across groups of loan applicants. Among those who profit, persons facing historical or current

[33] Tetyana Balyuk, "FinTech lending and bank credit access for consumers" (2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2802220, accessed July 22, 2024; Bartlett et al. (n 18) 55.

[34] Agarwal et al. (n 32) 26.

[35] Tobias Berg et al., "On the rise of FinTechs – credit scoring using digital footprints" (2018) *NBER* Working Paper 24551, 2.

[36] Berg et al. (n 35) 3.

[37] Langenbucher and Corcoran (n 31); Marco Di Maggio, Dimuthu Ratnadiwakara, and Don Carmichael, "Invisible primes: FinTech lending with alternative data" (2022) 2, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3937438, accessed July 22, 2024.

[38] Di Maggio et al. (n 37) 4.

[39] Patrick Alexander Fickling and Paul Watkins, "An update on credit access and the Bureau's first no-action letter" (2019), www.consumerfinance.gov/about-us/blog/update-credit-access-and-no-action-letter/, accessed July 22, 2024.

https://doi.org/10.1017/9781009367783.020 Published online by Cambridge University Press

discrimination are underrepresented. A well-documented example concerns a report from a US NGO which ran a mystery shopping exercise with Upstart.[40] It involved three loan applicants that were identical to college degree, job, and yearly income, but had attended different colleges: New York University, Howard University, which is a historically Black college, and New Mexico State University, a Hispanic-serving institution. Holding all other inputs constant, the authors of the study found that a hypothetical applicant who attended Howard or New Mexico State University would pay significantly higher origination fees and interest rates over the life of their loans than an applicant who attended NYU.

One explanation for these findings points to the AI system predicting credit default in a world of (past and current) discrimination where Black and Hispanic applicants face thresholds that otherwise similarly situated borrowers do not. Along those lines, the unequal output reflects real differences in credit default risk.

Alternatively (or additionally), the AI system's result might be skewed by a variety of algorithmic biases.[41] In this case, its credit score or creditworthiness assessment paints a picture that does not correctly reflect actual differences. Historical bias provides one example. A machine-learning system is trained based on past data, involving borrowers, their characteristics, behavior, and payment history.[42] Based on such data, the AI system learns which individual attributes (or bundles of attributes) are good predictors of credit default. If certain groups, for example unmarried women, have in the past faced cultural or legal obstacles in certain countries, the AI system learns that being an unmarried woman is a negative signal. It will weigh this input variable accordingly. In this way, an attribute that would have lowered a credit score in the past will lower it today, even if the underlying discriminatory situation has been overcome (as is partly the case for unmarried women). Majority bias is another example.[43] The AI system builds its model by attributing weight to input variables.

[40] Student Borrower Protection Center, "Educational redlining" (2020), https://protectborrowers.org/wp-content/uploads/2020/02/Education-Redlining-Report.pdf, accessed July 22, 2024, methodology described at 16.

[41] For a critical discussion, see Laura Blattner and Scott Nelson, "How costly is noise? Data and disparities in consumer credit" (2021), www.researchgate.net/publication/351656623_How_Costly_is_Noise_Data_and_Disparities_in_Consumer_Credit, accessed July 22, 2024; Citron and Pasquale (n 20); Margot Kaminski, "Binary governance: Lessons from the GDPR's approach to algorithmic accountability" (2019) *Southern California Law Review*, 92: 1529, 1538; O'Neil (n 4); from the perspective of sociology: Jenna Burrell and Marion Fourcade, "The society of algorithms" (2021) *Annual Review of Sociology*, 47: 213, 224; Dan L. Burk, "Algorithmic legal metrics" (2021) *Notre Dame Law Review*, 96: 1147, 1163; Barbara Kiviat, "The art of deciding with data: evidence from how employers translate credit reports into hiring decisions" (2019) *Socio-Economic Review*, 17: 283; Ead, "The moral limits of predictive practices: The case of credit-based insurance scores" (2019) *Socio-Economic Review* 84: 1134; Pauline Kim, "AI and inequality," in Kristin Johnson and Carla Reyes (eds), *The Cambridge Handbook on Artificial Intelligence and the Law* (forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3938578, accessed July 22, 2024.

[42] See Section 16.2.2.

[43] Solon Barocas and Andrew Selbst, "Big data's disparate impact" (2016) *California Law Review*, 104: 671, 689; Talia Gillis, "The input fallacy" (2022) *Minnesota Law Review*, 106: 1175–1261; Jennifer Graham,

https://doi.org/10.1017/9781009367783.020 Published online by Cambridge University Press

What it finds for most candidates that were successful in the past will be accorded considerable weight, for instance, stability as to place of residence. Candidates who score badly in that respect because their job requires them to move often will face a risk premium. It is important to understand that the machine-learning system finds correlations only and does not attempt to establish causation. An individual might have very good reasons for moving houses often, he might even hold a high-paying job that requires him to do so. The AI system will still count this as a negative attribute if the majority of past candidates did not move often.

Another empirical study suggests that there can be yet further factors at play. Researchers explored data on the US Government Sponsored Enterprises Fannie Mae and Freddie Mac.[44] These offer mortgage-backed lenders a guarantee against credit risk, charging a fee that depends only on the borrower's credit score and the loan-to-value ratio. Against that background, one would assume that for candidates with identical scores and LTV ratios, interest rates must be identical. This was not what the study found. Hispanic and Black borrowers faced markups of 7.9 basis points for purchase mortgages and 3.6 basis points for refinance mortgages. The "morally neutral" AI system might have understood this as a promising strategy to meet the goal of maximizing profit for the lender. Instead, the lender might have himself formulated the AI's definition of success as identifying applicants who were open to predatory pricing, for instance because they urgently needed a loan or were financially less literate than other borrowers.[45]

16.3.2 *Opaque Surveillance*

Let us revisit the lender who faced uncertainty as to potential borrowers. Today, he uses credit scores and creditworthiness assessments that rely on a limited list of input variables. Usually, their relevance for credit default risk is obvious, and advanced statistics allows for good predictions. For the applicant, this entails the upside that he will typically know which attributes are important to be considered a good credit risk.[46] Against that background, one would expect that an unobservable

"Risk of discrimination in AI systems, evaluating the effectiveness of current legal safeguards in tackling algorithmic discrimination" in Alison Lui and Nicholas Ryder (eds), *FinTech, Artificial Intelligence and the Law* (Routledge, 2021), p. 211; Katja Langenbucher, "Responsible A.I. credit scoring – a legal Framework" (2020) *European Business Law Review* 31: 527; Burk (n 41); Antje von Ungern-Sternberg, "Diskriminierungsschutz bei algorithmischen Entscheidungen" in Anna Katharina Mangold and Mehrdad Payandeh (eds), *Handbuch Antidiskriminierungsrecht* (Mohr Siebeck, 2022) nt 15ff.

[44] Bartlett et al. (n 18) 31.

[45] Ibid., 32: "The fact that the relation between the rate differential and either credit score or realized default is minor suggests the income and LTV results may instead reflect something else, such as the correlation between income, financial sophistication, and a propensity to shop for rates"; similarly Gillis (n 43) 1188 ("personalized pricing"); Christophe Hurlin, Christophe Pérignon, and Sébastien Saurin, "The Fairness of credit scoring models" (2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3785882, accessed July 22, 2024: "lack of fairness."

[46] But see Rona-Tas and Guseva (n 19) 62: "they act as disciplining devices."

https://doi.org/10.1017/9781009367783.020 Published online by Cambridge University Press

and hardly quantifiable variable such as "character"[47] loses significance. With AI-based scoring, this might change, if the seemingly objective machine lends new credibility to such concepts. The researchers who used mobile phone data for credit scoring interpreted their findings as a strategy to access "aspects of individuals behavior" that "has implications for the likelihood of default."[48] The authors of the e-commerce study[49] explicitly suggest that the variables they investigated provide a "proxy for income, character and reputation."[50]

A borrower whose credit application is assessed by an algorithm might feel compelled to give access to his personal data unless he is prepared to accept a lower credit rating. At the same time, he does not necessarily know which elements of the data he hands over will be relevant. Credit applicants today often know what is important for obtaining credit and have legal rights to be informed about a denial.[51] Under an AI black box model, not even the lender is necessarily aware of what drives the credit score his AI system produces.[52] If he does, his incentives to inform the applicant are often small. This is especially likely if the lender feels he found a variable that is a powerful predictor but at the same can be manipulated by the applicant. Consider, for instance, a finance or a dating app on his phone, one helping, the other hurting his credit score, while both apps are easily installed/uninstalled.

AI scoring of this type places consumers in a difficult spot. They are likely to worry about a "world of conformity"[53] where they "fear to express their individual personality online" and "constantly consider their digital footprints."[54] They might feel that they are exposed to arbitrary decisions that they do not understand and that are sometimes unexplainable even to the person using the algorithm.[55] Economists have predicted that consumers might try to randomly change their online behavior in the hope for a better score.[56] Manipulation along those lines will work better for some variables (such as regularly charging a mobile device) than for others (such as changing mobile phone brand or refraining from impulse shopping).[57] One strategy is to mimic the profile of an attractive borrower. This suggests side effects to the overall usefulness of AI scoring. If it is costless to mimic an attractive borrower, an

[47] See Section 16.2.1.
[48] Agarwal et al. (n 32) 3.
[49] See Section 16.2.3.
[50] Berg et al. (n 35) 3.
[51] See, for example, Section 202.9(a)(2)(i), (b)(2) of US Regulation B, implementing the Equal Credit Opportunity Act.
[52] Fickling and Watkins (n 39) 8; Burrell and Fourcade (n 41) 226.
[53] Berg et al. (n 35) 26.
[54] Ibid., 6; Burk (n 41).
[55] Fickling and Watkins (n 39) 18; Burrell and Fourcade (n 41) 226.
[56] Berg et al. (n 35) 25 referencing the Lucas Critique, see Robert Lucas, "Econometric policy evaluation: A critique" (1976) *Carnegie-Rochester Conference Series on Public Policy*, 1: 19.
[57] Berg et al. (n 35) 25–26.

https://doi.org/10.1017/9781009367783.020 Published online by Cambridge University Press

uninformative pooling equilibrium evolves: All senders choose the same signals.[58] Firm behavior might adapt as well.[59] A firm whose products signal low creditworthiness could try to conceal its products' digital footprint. Commercial services may develop, offering such services or making consumers' digital footprint look better. Along similar lines, the US CFPB fears that the chances to change credit standing through behavior may become a random exercise.[60] While the applicant today receives meaningful information about (many of) the variables which are relevant for his score, with opaque AI modelling, this is not guaranteed any more.

16.4 LEGAL TOOLS REGULATING DISCRIMINATION AND SURVEILLANCE FOR AI-BASED CREDIT SCORING AND CREDITWORTHINESS EVALUATION

Using AI for scoring and underwriting decisions does not raise entirely novel concerns. However, it compounds some of the well-known risks or shines an unanticipated light on existing strategies to deal with these challenges.

16.4.1 *Anti-Discrimination Laws Faced with AI*

Is unequal output across groups of applicants, for instance as to sex or race, necessarily a cause for concern? Arguably, the answer depends on the context and the goal pursued by the lender.

Under the assumption that the AI model presents an unbiased picture of reality, an economist would find nothing wrong with unequal output if it tracked features that are relevant to the lender's business strategy. Any creditor must distinguish between applicants for a loan,[61] and score and rank them, usually according to credit default risk. This produces statistical discrimination that reflects the state of the world. Prohibiting unequal output entirely would force the lender to underwrite credit default risk he does not wish to take on.[62]

By contrast, if the assumption of the AI system reflecting an unbiased picture of the world does not hold, the unequal output can be a signal for potential inefficiencies.

[58] Ibid., 26. A higher cost for mimicking, those authors explain, results in a separating equilibrium with a highly informative digital footprint. They illustrate this with the example of Pentaquark, who rejects loans from applicants who "write a lot about their souls on Facebook, as these persons are usually too concerned about what will happen in thirty years, but not the fine print of today's life."

[59] Ibid., 26–27.

[60] Fickling and Watkins (n 39) 17; see on further concerns, such as "gaming the system"; Burk (n 41) 1187ff; Citron and Pasquale (n 20) 29ff; Langenbucher (n 43).

[61] See Section 16.2.1.

[62] Further side effects can hurt the borrower (if he ends up unable to repay and gets into financial difficulties), other borrowers (if the lender raises interest rates to finance the credit risk, he is forced to accept) and the economy as a whole (if unsustainable loans lead to a credit bubble and to market instability).

https://doi.org/10.1017/9781009367783.020 Published online by Cambridge University Press

An AI model's flawed output, for example due to historical or majority bias,[63] leads to opportunity cost if it produces too many false negatives: Candidates that would be a good credit risk but are flagged as a bad credit risk. In this case, the lender should have granted a loan, but he refused, based on the flawed AI system. If, by contrast, the model triggers too many false positives, it can skew the lender's portfolio risk. The lender has underwritten more contracts than his business model would have suggested to.

A lawyer has a more complicated answer to the question mentioned earlier.[64] Depending on the situation, unequal output can violate anti-discrimination laws. A clear case is presented by a lender who denies a loan *because of* a specific attribute – sex, race, religion, ethnicity, or similar protected characteristics. If this is the case (and the plaintiff can prove it), the lender is liable for damages.[65] The test prong *because of* is met, if the protected characteristic was one reason toward the decision.

Direct discrimination[66] in this form is not a risk that is unique to AI-based decision-making. Quite to the contrary, many point out that a well-trained, *objective* AI system will overcome human biases and discriminatory intentions.[67] However, if a lender pursues discriminatory motives or intentionally seeks members of protected communities because they are more vulnerable to predatory pricing, AI systems compound the risks plaintiffs face. This has to do with the AI system's potential to help the lender "mask" his true preferences.[68] Masking behavior can be successful because anti-discrimination law needs a hook, as it were, in the lender's decision-making process. The plaintiff must establish that he was discriminated against because of a protected characteristic, such as race or sex. A discriminatory lender can try to circumvent this legal rule by training its AI system to find variables that correlate narrowly with a protected characteristic. Eventually, circumventing the law will not be a valid strategy for the lender. However, the applicant might find it very hard to prove that this was the lender's motive, especially in jurisdictions that do not offer pre-trial discovery.[69]

[63] See Section 16.3.1.

[64] See in detail Langenbucher (n 13).

[65] Michael Heese, "Offene Preisdiskriminierung und zivilrechtliches Benachteiligungsverbot, Eine Zwischenbilanz" (2012) *Neue Juristische Wochenschrift* 572, 575–576. On punitive damages see, for example, US 15 U.S.C. § 1691e(b).

[66] The US terminology is: "disparate treatment."

[67] Cass Sunstein, "Algorithms, correcting biases" (2019) *Social Research: An International Quarterly*, 86: 499.

[68] Barocas and Selbst (n 43) 699, use the term "masking"; for German law, see Florian Rödl and Andreas Leidinger, "Diskriminierungsschutz im Zivilrechtsverkehr" In Anna Katharina Mangold and Mehrdad Payandeh (eds), *Handbuch Antidiskriminierungsrecht* (Mohr Siebeck, 2022), nt 57 ("cover-up").

[69] Such as the EU, see Geoffrey C. Hazard Jr., "Civil procedure rules for European courts" (2016) *Judicature*, 100: 58.

https://doi.org/10.1017/9781009367783.020 Published online by Cambridge University Press

The disproportionate output of an AI system does not always go back to business strategies that imply direct discrimination.[70] One of the most characteristic features of algorithm-based credit risk assessments is to find unanticipated correlations between big data variables and the optimization goal.[71] What happens if it turns out that a neutral attribute, for instance the installation of a finance app on a smartphone, triggers disproportionate results across sex? Unless a lender intentionally circumvents the rule to not discriminate against women, this is not a case of direct discrimination.[72] Instead, we potentially face indirect discrimination.[73] Under this doctrine, a facially neutral attribute that consistently leads to less favorable output for protected communities becomes "suspicious," as it were. Plaintiffs must establish a correlation between the suspicious attribute and the unequal output. Defendants will be asked to provide justificatory reasons for using the attribute, in spite of the troubling correlation.[74] In a credit underwriting context, the business rationale of the lender is a paradigm justificatory reason. The plaintiff might counter that there are equally powerful predictors with less or no discriminatory potential. However, the plaintiff will often fail to even establish the very first test prong, namely to identify the suspicious attribute. The more sophisticated the AI system and the larger the big data pool, the more likely it is that the AI system will deliver the same result without access to the suspicious variable. The reason for this is *redundant encoding*: The information encoded in the suspicious variable can be found in many other variables.[75]

16.4.2 *AI Biases and Quality Control in the EU Artificial Intelligence Act*

The EU AI Act[76] starts from the assumption that AI credit underwriting can lead to discriminatory output:

> In addition, AI systems used to evaluate the credit score or creditworthiness of natural persons (…) may lead to discrimination of persons or groups and perpetuate historical patterns of discrimination, such as that based on racial or ethnic origins, gender, disabilities, age or sexual orientation, or create new forms of discriminatory impacts.[77]

[70] US: "disparate treatment."
[71] See Section 16.2.2.
[72] US: "disparate treatment."
[73] US: "disparate impact"; in more detail: Langenbucher (n 13).
[74] See Deborah Hellman, "Measuring algorithmic fairness" (2020) *Virginia Law Review*, 106: 811, 852 on the US Supreme Court distinguishing between a defendant relying on a neutral variable "because of" or "in spite of" the foreseeable consequences.
[75] Von Ungern-Sternberg (n 43) nt 27.
[76] See Chapter 12.
[77] AI Act, Recital (58).

https://doi.org/10.1017/9781009367783.020 Published online by Cambridge University Press

However, the remedy the AI Act proposes is not to adjust anti-discrimination law. Instead, it proposes a strategy of product regulation. Artificial intelligence systems employed in the loan context "should be classified as high-risk AI systems, since they determine those persons' access to financial resources or essential services such as housing, electricity, and telecommunication services."[78]

Employing a high-risk AI system entails mandatory compliance checks as to monitoring, testing, and documentation with an eye on both software and data.[79] Article 9 of the Act has the software part in mind. It requires identifying risks to fundamental rights and developing appropriate risk management measures. Ideally, the design or development of the high-risk AI system eliminates or reduces these risks. If they cannot be eliminated, they must be mitigated and controlled.[80] Article 10 addresses training data which might lead to the biases mentioned earlier.[81] According to Article 14 of the proposal, high-risk AI systems must be "designed and developed in such a way (…) that they can be effectively overseen by natural persons."[82] Supervisory agencies are in charge of enforcing compliance with the AI Act. For credit institutions, the competent banking regulator is entrusted with this task. Nonbank entities, for instance credit scoring agencies, will be supervised by a different body in charge of AI.[83] Both supervisors must deal with the challenge of quantifying a fundamental rights violation (which includes balancing borrower's rights against competing rights) to produce a workable benchmark for risk management.

16.4.3 *Privacy and Retail-Borrower Protection Laws Faced with AI*

Private enforcement via litigation is not included in the AI Act.[84] Against this background, the following section looks at privacy and retail-borrower protection laws that provide such tools. Privacy law aims at keeping personal data private and subjecting its use by third parties to certain requirements. Retail-borrower protection laws cover many aspects of a transaction between borrower and lender. These include rights that are especially useful in the context of algorithmic credit evaluations, for example a right to be informed about a denial of credit.

[78] Ibid., Recital (58), Annex III No 5(b) lists credit-scoring and underwriting algorithms if they concern natural persons.

[79] Overview at Katja Langenbucher, "AI credit scoring and evaluation of creditworthiness – a test case for the EU proposal for an AI Act" (2022) *ECB Legal Conference* 362, 367, www.ecb.europa.eu/pub/pdf/other/ecb.ecblegalconferenceproceedings202204~c2e5739756.en.pdf, accessed July 22, 2024.

[80] AI Act, Art. 9(5).

[81] Section 16.3.1, see AI Act, Art. 10(2) lit g on examining the system "in view of possible biases."

[82] AI Act, Art. 14(1), on human oversight, see Langenbucher (n 79) 372.

[83] Critical on this distinction Langenbucher (n 79) 375ff.

[84] Critical ibid., 381ff.

https://doi.org/10.1017/9781009367783.020 Published online by Cambridge University Press

16.4.3.1 Credit Reporting and Data Privacy

Big data access is a key ingredient of algorithm-based credit underwriting.[85] At the same time, data collected from online sources are often unreliable and prone to misunderstandings. If an AI model is trained on flawed or misleading data, its output will likely not fully reflect actual credit default risk. However, this might not be immediately visible to the lender who uses the model. Even worse, automation bias, the tendency to over-rely on what was produced by automated models and disregard conflicting human judgments,[86] might induce the lender to go ahead with the decision prepared by the algorithm.

In many countries, credit reporting bureaus have traditionally filled the role of collecting data, some run by private companies, some by the government.[87] Those bureaus have their proprietary procedures to verify information. Additionally, there are legal rights for borrowers to correct false entries in credit registries. Illustrative for such rights is the US Fair Credit Reporting Act. It entitles credit applicants to access information a credit reporting agency holds on them and provides rights to rectify incorrect information.[88] However, one requirement is that the entity collecting big data qualifies as a credit reporting agency under the Act.[89] While some big data aggregators have stepped forward to embrace this responsibility, others claim they are mere "conduits," performing mechanical tasks when sending the data to FinTech platforms.[90]

EU law does not face this doctrinal difficulty. The prime EU data protection law is the General Data Protection Regulation (GDPR) that covers any processing of personal data under Article 2 of GDPR. For processing to be lawful, it must qualify for a justificatory reason under Article 6 GDPR. The data controller must provide information, *inter alia* on the purpose of data collection and processing pursuant to Article 13 GDPR. If sensitive data are concerned, additional requirements follow Article 9. However, in practice, GDPR requirements are often met by a standard tick-the-box-exercise whenever data are collected. Arguably, this entails rationally apathetic, rather than well-informed consumers.[91] When a data aggregator furnishes

[85] See Section 16.2.1.

[86] Definition (in a clinical context) by Kate Goddard, Abdul Roudsari, and Jeremy C. Wyatt, "Automation bias: A systematic review of frequency, effect mediators, and mitigators" (2012) *Journal of the American Medical Informatics Association*, 19: 121.

[87] International overview at Rona-Tas and Guseva, (n 19) 61ff.

[88] Langenbucher and Corcoran (n 31) 162.

[89] For a second concern see Section 16.4.3.2.

[90] Federal Trade Commission, "40 years of experience with the Fair Credit Reporting Act" (Report) (2011) 29, www.ftc.gov/sites/default/files/documents/reports/40-years-experience-fair-credit-reporting-act-ftc-staff-report-summary-interpretations/110720fcrareport.pdf, accessed July 22, 2024; see for a narrow reading of the LexisNexis product "Accurint" which was not considered delivering "credit reports": Pauline Kim and Erika Hanson, "People Analytics and the Regulation of Information under the Fair Credit Reporting Act" (2016) 61 *Saint Louis University Law Journal* 17, 28–29.

[91] Langenbucher (n 43) 535–536.

https://doi.org/10.1017/9781009367783.020 Published online by Cambridge University Press

data he lawfully collected to the lender or to a scoring bureau, there is no additional notice required.[92] This foregoes the potential to incentivize consumers to react in the face of a particularly salient use of their data.

16.4.3.2 Credit Scoring, Creditworthiness Evaluation, and Retail Borrower Rights

If collecting big data is the first important element of AI-based credit underwriting,[93] the way in which an algorithm assesses the applicant is the second cornerstone. The uneasy feeling of facing unknown variables, which drive scores and evaluations in opaque ways, might be mitigated if applicants receive meaningful explanations about which data were used and how the algorithm arrived at the output it generated.

US law provides two legal tools to that end. One rule was mentioned in the previous section.[94] It requires the lender to disclose that he used a credit report. In that way, it allows the applicant to verify the information in the credit report. In the old world of traditional credit reporting and scoring, based on a short list of input variables, this is an appropriate tool. It remains to be seen how this right to access information will perform if the input data are collected across a vast amount of big data sources. The second rule gives consumers a right to a statement of specific reasons for adverse action on a credit application.[95] The underlying rationale is to enable the applicant to make sure no discriminatory reasons underlie the denial of credit. In the current environment, US regulators have already struggled to incentivize lenders to provide more than highly standardized information.[96] With algorithmic scoring, this information is even harder to provide if the algorithm moves from a simple machine-learning device to more sophisticated black box or neural network models.

The EU GDPR includes no rule to specifically target the scoring or underwriting situation. General rules concern "decisions based solely on automated processing" and vest the consumer with a right to get "meaningful information about the logic involved" and "to obtain human intervention, to express his or her point of view and to contest the decision."[97] At the time of writing, a case was pending before the European Court of Justice to assess what these rules entail as to credit scoring.[98]

[92] See Regulation (EU) 2016/679 of the European Parliament and of the Council of April 27, 2016, on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1, Art. 13(1), (2) referencing the "time when personal data are obtained." According to Art. 13(3), this is different if "the controller intends to further process the personal data for a purpose other than that for which the personal data were collected."

[93] Section 16.4.3.1.

[94] Section 16.4.3.1.

[95] Section 202.9(a)(2)(i), (b)(2) of Regulation B, implementing the Equal Credit Opportunities Act.

[96] For details on the official interpretation of that rule, see Consumer Financial Protection Bureau, www.consumerfinance.gov/rules-policy/regulations/1002/9/#9-b-2-Interp-2, accessed July 22, 2024.

[97] See Regulation (EU) 2016/679, Art. 13(2) lit f, Art. 15(1) lit h, Art. 22.

[98] The case was decided on 7 December 2023, see ECLI:EU:C:2023:957.

https://doi.org/10.1017/9781009367783.020 Published online by Cambridge University Press

In contrast with the GDPR, the EU Consumer Credit Directive directly engages with algorithmic decision-making in the underwriting context.[99] It includes a right to inform the consumer whose credit application is denied if the "application is rejected on the basis of a consultation of a database."[100] However, it refrains from requiring specific reasons for the lender's decision. Art. 18 includes more detailed access, namely a right to "request a clear and comprehensible explanation of the assessment of creditworthiness, including on the logic and risks involved in the automated processing of personal data as well as its significance and effects on the decision."[101] What such information would look like in practice, and whether it could be produced for more sophisticated algorithms, remains to be seen. The same concern applies to a different strategy proposed by the Directive. It entirely prohibits the use of alternative data gathered from social networks and certain types of sensitive data.[102] Arguably, redundant encoding[103] can make this a toothless rule if the same information is stored in a variety of different variables.

16.5 LOOKING AHEAD – FROM CREDIT SCORING TO SOCIAL SCORING

Scoring consumers to assess their creditworthiness is an enormously important use of the novel combination that big data and AI bring about. Chances are that scoring will not stop there but extend to more areas of social life, involving novel forms of social control.[104] When considering high-risk areas, the EU AI Act not only has "access to essential private and public services and benefits"[105] in mind. The lawmakers have set their eyes on social scoring as well, which they understand as an "evaluation or classification of natural persons or groups of persons over a certain period of time based on their social behavior or known, inferred or predicted personal or personality characteristics."[106] Social scoring of this type is prohibited if it occurs out of the context in which the data were collected or leads to unjustified treatment.[107] However, the success of a substantive rule depends on efficient means of enforcement. Faced with the velocity of digital innovation, it is doubtful that either public or private enforcement tools can keep pace.

[99] Directive (EU) 2023/2225 (n 12).
[100] Directive (EU) 2023/2225, Art. 19(6).
[101] Directive (EU) 2023/2225, Art. 18(8) lit a.
[102] Directive (EU) 2023/2225, Art. 19(5).
[103] See Section 16.4.1.
[104] AI Act, Recital (28).
[105] AI Act, Recital (58).
[106] AI Act, Art. 5(1) lit c.
[107] AI Act, Art. 5(1) lit c.

https://doi.org/10.1017/9781009367783.020 Published online by Cambridge University Press