Autonomous Systems and the Law

Bearbeitet von
Edited by Nikita Aggarwal, Horst Eidenmüller, Luca Enriques, Jennifer Payne, and Kristin Zwieten

larger volume and variety of data ('alternative' data); the second relates to the use of more sophisticated techniques to analyse the data. Regarding the first dimension, alternative data includes both *non-credit,* financial data (for example, direct data on rental and mobile phone bill payments),[11] as well as *non-credit, non-financial* data – for example, 'social' data captured from consumers' social media networks, and 'behavioural' data about consumers' habits and preferences.[12]

The second and more recent development embodied by algorithmic credit scoring is the use of ML techniques to analyse the data. This in turn impacts the first dimension: the types of data that can be used. Significantly, ML algorithms can parse very large volumes of data – especially, raw, unstructured, high-dimensional, and/or anonymized data – to find correlations that could be (more) relevant to predicting a borrower's creditworthiness. Notably, ML can more accurately capture non-linear relationships in the data, as well as reflect changes in the population and environment by 'learning' from new training data. A form of ML called 'deep learning', using multi-layer neural networks,[13] has shown particular promise in analyzing unstructured and high-dimensional data.[14]

c) Impact of Algorithmic Credit Scoring on Consumer Credit Markets

Expanding the number and types of measured variables, and employing ML techniques, thus allows for a more detailed, multi-dimensional observation of a borrower's characteristics that can be used to estimate their creditworthiness. This is particularly important for thin file and no-file borrowers, who may present an acceptable credit risk despite not having any conventional, financial credit data to support this assessment. As such, by enabling more accurate creditworthiness assessment, algorithmic credit scoring stands to enhance the efficiency of consumer credit markets.

Furthermore, by widening access to credit for thin-file and no-file borrowers, algorithmic credit scoring can help to redress extant distributional and fairness concerns in consumer credit markets, given that these borrowers are more likely to be from low-income, less educated and ethnic minority backgrounds.[15] Algorithmic credit scoring could also reduce the scope for

[11] See U.S. Bureau for Consumer Financial Protection 'Request for Information Regarding Use of Alternative Data and Modeling Techniques in the Credit Process' (2017) 82 FR 11183 https://bit.ly/244qbDP.

[12] See M Hurley and J Adebayo, 'Credit Scoring in the Era of Big Data' (2016) 18 Yale Journal of Law and Technology 148.

[13] Computer networks of nodes or units connected by links that simulate the neural circuits in human brains.

[14] See I Goodfellow, Y Bengio and A Courville *Deep Learning* (MIT Press 2016), Ch 1. Note that algorithmic credit scoring providers do not necessarily use all categories of alternative data e. g. Aire (aire.io) and ZestFinance (https://www.zestfinance.com/zaml) claim that they do not use 'social' data as they are less reliable.

[15] See CFPB, 'Becoming Credit Visible' (2017) https://www.consumerfinance.gov/about-us/newsroom/consumer-financial-protection-bureau-study-finds-consumers-lower-income-areas-are-more-likely-become-credit-visible-due-negative-records/; J Y Campbell et al, 'Consumer Financial Protection' (2011) 25(1) Journal of Economic Perspectives 91, 100.

unfairness due to 'statistical discrimination'. By increasing the observability of non-protected characteristics relevant to a borrower's creditworthiness, the incentive for lenders to rely on conventionally more observable yet protected characteristics, such as sex or race, as statistical proxies for creditworthiness, should reduce.[16]

Conversely, however, there is a risk that the opacity and complexity of certain ML approaches could make it more difficult to pre-empt or verify *ex post* whether the system has (inadvertently) facilitated unlawful discrimination, by relying on protected characteristics, or their proxies, in reaching a credit decision.[17] Relatedly, biases in the training, validation and/or test data used to build ML scoring models could perpetuate past discrimination in lending. For example, an ML model trained on data from a predominantly white population could result in bias against lending to non-white populations. Likewise, spurious correlations in these datasets can lead to inaccurate (and potentially unfair) predictions when the model is applied to new, 'out of sample' data.[18]

Algorithmic credit scoring could furthermore become a source of inefficiency and unfairness in consumer credit markets if it is used by lenders to more effectively exploit the cognitive and behavioural limitations of borrowers. Inter alia, a lender could use behavioural insights derived from algorithmic credit scoring to more precisely target a borrower, or profiled groups of borrowers, with unfavourable credit offers at moments of extreme vulnerability. This could increase the chance that a borrower reflexively agrees to an unfavourable contract, without carefully reviewing its terms or shopping around for a better offer.[19]

The question arises whether these same technologies deployed in the hands of borrowers could attenuate such risks. For example, ML and Big Data are already being used to build highly personalized web plugins and mobile apps that seek to counteract behavioural biases, and 'nudge' consumers into making better financial decisions.[20] However, whether these applications can be effective in this way depends on the extent to which they are adopted by consumers. Less financially literate consumers are less

---

[16] On statistical discrimination, see V D Rougeau and K N Hylton, 'Lending Discrimination: Economic Theory, Econometric Evidence, and the Community Reinvestment Act' (1996) Scholarly Works Paper 874 http://scholarship.law.nd.edu/law_faculty_scholarship/874. Credit providers are prohibited from discriminating against borrowers, directly or indirectly, on the basis of legally protected characteristics such as sex, race or religion (Chapter 1 and 2, UK Equality Act, 2010).

[17] See S Barocas and A Selbst, 'Big Data's Disparate Impact' (2016) 104 CALIF. L. REV. 671.

[18] See C O'Neil *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Random House 2016), ch 8 'Collateral Damage'.

[19] See further G Wagner and H Eidenmüller, 'Down by Algorithms? Siphoning Rents, Exploiting Biases and Shaping Preferences – The Dark Side of Personalized Transactions' (2019) University of Chicago Law Review, Forthcoming https://ssrn.com/abstract=3160276.

[20] See for example, *Clarity Money* (https://claritymoney.com) and *Cleo* (https://www.meetcleo.com/). See also FCA, 'Applying Behavioural Economics at the Financial Conduct Authority' (2013) Occasional Paper No. 1 https://www.fca.org.uk/publication/occasional-papers/occasional-paper-1.pdf.

likely to understand the value of these solutions, in order to avail of them in the first instance. Moreover, to the extent that these solutions do not fully replace financial decision-making by consumers, the latter, compromised by behavioural and cognitive weaknesses, could simply ignore the advice offered by the relevant app.

More importantly, it is questionable whether these applications will be able to fully overcome the informational and behavioural advantage that lenders have over borrowers, and which they use to exploit borrowers. In particular, lenders enjoy privileged access to aggregate financial transaction data (conventional 'credit data') and product use patterns gleaned from multiple transactions with borrowers over time, that could be difficult for third party consumer-helping platforms to substitute.[21]

3. Regulatory Challenges and Opportunities

Algorithmic credit scoring thus presents itself as a double-edged sword. On the one hand, it stands to benefit consumer credit markets, inter alia, by improving the accuracy of creditworthiness assessment and thereby widening access to credit from mainstream lenders. On the other hand, it could generate new sources of inefficiency and unfairness through the exploitation of consumers' cognitive and behavioural weaknesses, and unlawful discrimination. To the extent that the market, in the form of consumer-helping applications, is unable to offer a complete solution to these risks, consideration must be given to whether and if so how government-backed regulation should be strengthened.

As a general matter, the principles and conduct-based approach of the UK consumer credit regulatory regime[22] gives regulators flexibility to respond dynamically to the use of new and fast evolving technologies, such as algorithmic credit scoring, by market participants. In particular, the principles that firms must 'treat customers fairly', act with 'due care, skill and diligence', and ensure that product marketing is 'clear, fair and not misleading', provide a broad legal basis for regulators to respond to potential exploitation and discrimination against consumers through the use of algorithmic credit scoring, and for firms to design appropriate systems and controls in order to achieve the outcomes enshrined in these principles.[23]

To complement their dialogue with firms under the principles-based approach, regulators could themselves make greater use of ML and Big Data techniques to more directly detect, understand and remedy undesirable behaviour by market participants. This includes, for example, empirically

[21] Banks share financial account data on a voluntary, reciprocal basis (see 'Principles of Reciprocity' http://www.scoronline.co.uk/sites/default/files/por_version_36.pdf.). The new 'data portability' rules under the EU General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679 [2016] OJ L 119/1) and 'access-to-account' rule under the Second EU Payment Services Directive (PSD2) (Directive (EU) 2015/2366 [2015] OJ L 337/35) could, however, improve access to financial data for non-bank financial services providers.

[22] See J Armour et al (n 7), Ch 24.4.2.

[23] FCA Handbook, PRIN 2.1.

assessing how consumers respond to particular forms of product marketing, in order to ascertain whether it is 'clear, fair and not misleading'. These findings can be used to inform regulatory changes, for example, mandating greater personalisation of information disclosure by firms,[24] or requiring firms to adjust the consumer choice architecture in a more targeted way (for example, changing the default settings on their website or app to mitigate common consumer mistakes).[25]

Likewise, firms could be required to put in place more robust governance and oversight arrangements specifically relating to their ML systems and processes, including algorithmic credit scoring systems.[26] Inter alia, this could encompass procedures for data quality verification, as well as continuous model feedback testing, cross-validation and auditing[27] to mitigate data overfitting and algorithmic bias risks.[28] These procedures should build on the data protection auditing, certification, impact assessment and data protection 'by design and default' provisions under the GDPR, the new data protection regime in the EU.[29]

Indeed, cross-sectoral data protection regulation provides an important additional mechanism for mitigating potential discriminatory and unfair treatment of credit consumers due to the processing of their personal data through algorithmic credit scoring. Inter alia, the overarching principles that guide the GDPR could, if interpreted strictly, significantly restrict the potential for firms to abuse consumers' personal data. These include, in particular, the principles of 'purpose limitation' – requiring personal data to be collected only for 'specified, explicit, and legitimate purposes and not further processed in a manner that is incompatible with those purposes' – and 'data minimisation', requiring personal data to be 'adequate, relevant and limited to what is necessary in relation to the purpose for which they are processed'.[30]

The GDPR furthermore expands the rights of data subjects to control the use of their data, including a potentially broader right to receive 'meaningful information about the logic involved' in automated decision-making (the so-

---

24 See L Strahilevitz & A Porat, 'Personalizing Default Rules and Disclosure with Big Data,' 112 Michigan Law Review 1417 (2014); C Busch, 'Implementing Personalized Law: Personalized Disclosures in Consumer Law and Privacy Law' (2019), University of Chicago Law Review, Forthcoming https://ssrn.com/abstract=3181913.

25 See FCA (n 20).

26 J Armour et al (n 7), Ch 12.3.3.

27 See C Sandvig et al, 'Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms', in 'Data and Discrimination: Converting Critical Concerns into Productive Inquiry', preconference at the 64th Annual Meeting of the International Communication Association (Seattle, WA May 22 2014) https://bit.ly/1tGotry.

28 See C O'Neil (n 18); J Kroll et al, 'Accountable Algorithms' (2017) 165 U. Pa. L. Rev. 633.

29 See n 21. The FCA's Innovation Hub and Regulatory Sandbox offer important experimental fora for testing both algorithmic credit scoring systems, as well as regulatory solutions such as algorithmic auditing (FCA, 'Regulatory Sandbox' (11 May 2015) https://www.fca.org.uk/firms/regulatory-sandbox).

30 Art 5(1) GDPR. See further T Zarsky, 'Incompatible: The GDPR in the Age of Big Data' (2017) Vol 47 No. 4(2) Seton Hall Law Review.

called ‘right to explanation’).[31] An expansive interpretation of this right (and corresponding duty) by credit providers and/or regulators – for example, requiring credit providers to provide an *ex post* explanation to individual borrowers of the specific reasons underlying each credit decision – could better support borrowers in challenging discriminatory and unfair credit decisions.[32]

On the other hand, an expansive interpretation of data protection principles, rights and duties risks undermining the potential efficiency and fairness gains from algorithmic credit scoring. With respect to the principle of purpose limitation, as this contribution has highlighted, algorithmic credit scoring largely relies on repurposing data to uncover hidden insights about a borrower’s creditworthiness. Likewise, a more onerous ‘right to explanation’ could be undesirable if it restricts firms to using statistical techniques that are simpler and more ‘explainable’, yet less effective in assessing creditworthiness.[33]

4. Conclusion

Algorithmic credit scoring, and the Big Data and ML technologies underlying it, present both benefits and risks for consumer credit markets. This contribution has argued that the broadly principles and conduct-based approach of UK consumer credit regulation provides the flexibility necessary for regulators and market participants to respond dynamically to these new technological risks. This approach could be enhanced through the introduction of more robust product oversight and governance requirements for firms in relation to their use of ML systems and processes. Supervisory authorities could also themselves make greater use of ML and Big Data techniques in order to strengthen their supervision of consumer credit firms. Finally, cross-sectoral data protection regulation, recently updated in the EU under the GDPR, offers an important avenue to mitigate risks to consumers arising from the use of their personal data. However, the interpretation of this regime in the consumer finance context needs to be carefully calibrated, so as not to also inhibit the potential benefits of new technological applications such as algorithmic credit scoring, and Big Data and ML more generally.

[31] See B Flaxman and S Goodman, ‘EU Regulations on Algorithmic Decision-Making and a “Right to Explanation”’ (2016) arXiv:1606.08813 [stat.ML].

[32] See further S Wachter, B Mittelstadt and L Floridi, ‘Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation’ (2017) Volume 7(2) International Data Privacy Law 76.

[33] See Accenture, ‘Model Behaviour. Nothing Artificial – Emerging Trends in the Validation of Machine Learning and Artificial Intelligence Models’ (2018) https://accntu.re/2HQcFzi (suggesting that banks are avoiding ML techniques and certain types of alternative data for risk modelling out of fear that the increased complexity and opacity of their decision-making processes may not meet model validation, data integrity and audit requirements for regulatory capital purposes). On interpretability in machine learning, see A Selbst and S Barocas, ‘The Intuitive Appeal of Explainable Machines’ (2018) Fordham Law Review, Forthcoming https://ssrn.com/abstract=3126971.

VIII. Regulating Robotic Conduct: On ESMA's New Guidelines and Beyond

Florian Möslein

FinTech is rapidly transforming the financial services sector. Based on a broad range of new technologies and innovations, it increasingly attracts the interest of national, global and European regulators.[1] In the UK, for instance, HM Treasury published a Regulatory Innovation Plan which covers a number of actions that financial services regulators are taking to 'create a more supportive and agile regulatory and enforcement framework' for new business models and disruptive technologies, while breaking down barriers to entry and boosting productivity in financial services.[2] More recently, the European Commission released its long-awaited FinTech Action Plan,[3] and at the global level, the Financial Stability Board (FSB) issued a report on the financial stability implications of FinTech.[4]

So-called 'robo-advice' forms a more specific, important part of that FinTech sector: automated financial product advisors are emerging all across the financial services industry, helping clients to choose investments, banking products, and insurance policies.[5] While such advisors may have the potential to lower the cost and increase the quality of financial advice, they also pose significant challenges for regulators. The FSB has also highlighted the use of artificial intelligence and machine learning in financial services.[6] The respective regulatory challenges are of particular interest here because robo-advisors are a prime example of autonomous systems, processing great volumes of financial data on the basis of algorithmic decision making, including machine learning technologies. What we can currently observe with respect to robo-advice is indeed a law of autonomous systems in the making. This process

---

[1] For an extensive account of the legal and economic background to FinTech, see the contributions in Florian Möslein and Sebastian Omlor (eds), *FinTech-Handbuch – Digitalisierung, Recht, Finanzen* (Beck 2018 forthcoming).

[2] HM Treasury, 'Regulatory Innovation Plan' (4 April 2017) www.gov.uk/government/publications/hm-treasury-regulatory-innovation-plan, accessed 10 July 2018.

[3] Communication from the Commission to the European Parliament, the Council, the European Central Bank, the European Economic and Social Committee and the Committee of the Regions, 'FinTech Action plan: For a more competitive and innovative European financial sector', COM (2018) 109.

[4] Financial Stability Board, 'Financial Stability Implications from FinTech Supervisory and Regulatory Issues that Merit Authorities' Attention' (27 June 2017) www.fsb.org/wp-content/uploads/R270617.pdf, accessed 10 July 2018.

[5] More extensively, see Tom Baker and Benedict Dellaert, 'Regulating Robo Advice Across the Financial Services Industry' (2018) 103 Iowa L Rev 713.

[6] Financial Stability Board, 'Artificial Intelligence and Machine Learning in Financial Services – Market Developments and Financial Stability Implications' (1 November 2017) www.fsb.org/wp-content/uploads/P011117.pdf, accessed 10 July 2018.

can teach us important lessons both for future law-making in the field of autonomous systems and for regulating robotic conduct in general.

The 'Guidelines on certain aspects of the MiFID II suitability requirements', which the European Securities and Markets Authority (ESMA) has recently published, are a first step in this specific rule-making process.[7] In fact, ESMA expressly aims to 'consider recent technological developments of the advisory market, i.e. the increasing use of automated or semi-automated systems for the provision of investment advice or portfolio management (so-called 'robo-advice')'.[8] It builds on a report on automation in financial advice, published by the Joint Committee of the European Supervisory Authorities,[9] and is based on the Commission Delegated Regulation regarding organisational requirements and operating conditions for investment firms.[10] In its consultation paper, ESMA identified three main areas where specific needs for protection may arise, namely: (1) the information that should be provided to clients on the financial advice when it is provided through an automated tool; (2) the assessment of the suitability of financial products for the client, with particular attention to the use of online questionnaires with limited or without human interaction; and (3) the organisational arrangements that firms should implement when providing robo-advice.[11] Both client information and the arrangements necessary for them to understand investment products, i.e. the first two areas, do not specifically regulate robotic conduct, but simply focus on the electronic communication between advisors and their clients. In other words, these provisions focus on humans interacting with machines rather than on the machines themselves (and similar provisions might even be applied if the investment advice was elaborated by humans but was only delivered via electronic means).

Of more specific interest is therefore the third area – the draft guidelines concerning the organisational arrangements that firms should implement when providing robo-advice. These rules are designed to apply even if the interaction with clients does not occur through automated systems and only

---

[7] European Securities and Markets Authority (ESMA), 'Final Report: Guidelines on certain aspects of the MiFID II suitability requirements (ESMA 35-43-869)' (28 May 2018) www.esma.europa.eu/press-news/esma-news/esma-publishes-final-guidelines-mifid-ii-suitability-requirements, accessed 10 July 2018 (hereinafter: 'ESMA Final Report'). See also the related Consultation Paper (ESMA 35-43-748) (13 July 2017) www.esma.europa.eu/sites/default/files/library/2017-esma35-43-748_-_cp_on_draft_guidelines_on_suitability.pdf, accessed 10 July 2018 (hereinafter: 'ESMA Consultation Paper').

[8] ESMA Final Report (n 7) 5.

[9] Joint Committee of the three European Supervisory Authorities (ESAs) – EBA, EIOPA and ESMA, 'Report on Automation in Financial Advice' (16 December 2016) https://esas-joint-committee.europa.eu/Publications/Reports/EBA%20BS%202016%20422%20(JC%20SC%20CPFI%20Final%20Report%20on%20automated%20advice%20tools).pdf, accessed 10 July 2018.

[10] Commission Delegated Regulation (EU) 2017/565 of 25 April 2016 supplementing Directive 2014/65/EU of the European Parliament and of the Council as regards organisational requirements and operating conditions for investment firms and defined terms for the purposes of that Directive [2016] OJ L87/1.

[11] ESMA Consultation Paper (n 7) 13.