Process- and Tool-centred Solutions for Ensuring Respect for Individuals' Fundamental Rights in Algorithmic Credit Scoring

Law and Technology in Europe Master's Thesis

Student: Eva Opsenica

Supervisor: Pieter Kalis

Second reader: Lisette ten Haaf

1 July 2022

Contents

I: Introduction

1.1 Background

Artificial intelligence (AI) is an umbrella term for a range of computational techniques and processes that improve the ability of machines to perform cognitive or perceptual functions that were previously carried out by human beings.[1] At the heart of these techniques and processes are algorithms, finite sequences of formal rules that enable a machine to obtain a result from an initial input of information.[2] One possible way to make use of this technology is by automating decision-making in the field of finance. More specifically, a type of AI algorithms known as machine learning (hereinafter: ML) algorithms can be used as a tool for predicting the likelihood of defaulting on credit repayments.[3] The estimated probability of default is expressed in a credit score; and although this process is always algorithmic, as the credit score is obtained by following a set of instructions on how to transform inputs such as past credit history into a numerical value,[4] 'algorithmic credit scoring' is used to describe automated decision-making (hereinafter: ADM) where ML algorithms assess individuals' creditworthiness.[5] Unlike traditional forms of credit scoring, which rely on credit data such as the amount of debt, repayment performance, and length of credit history, this new form of statistical analysis treats "all data as credit data",[6] including how quickly someone pays their phone bills.[7]

---

[1] Filippo Raso and others, 'Artificial Intelligence & Human Rights: Opportunities & Risks' (Berkman Klein Center Research Publication 2018) 10 <https://cyber.harvard.edu/publication/2018/artificial-intelligence-human-rights> accessed 21 March 2022; David Leslie and others, 'Artificial Intelligence, Human Rights, Democracy, and the Rule of Law: A Primer' (The Alan Turing Institute 2021) 13 <https://www.turing.ac.uk/research/publications/ai-human-rights-democracy-and-rule-law-primer-prepared-council-europe> accessed 21 March 2022.

[2] Council of Europe, European Commission for the Efficiency of Justice, 'European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment' (31st plenary meeting, Strasbourg, 3–4 December 2018) 69.

[3] Stanley Greenstein, 'Preserving the Rule of Law in the Era of Artificial Intelligence (AI)' (2021) Artificial Intelligence and Law (Online first articles) 1, 9 <https://link.springer.com/article/10.1007/s10506-021-09294-4> accessed 21 March 2022; World Bank Group, 'Credit Scoring Approaches Guidelines' (2 April 2020) 2–3.

[4] Matthew Adam Bruckner, 'The Promise and Perils of Algorithmic Lenders' Use of Big Data' (2018) 93(1) Chicago-Kent Law Review 3, 11.

[5] Nikita Aggarwal, 'The Norms of Algorithmic Credit Scoring' (2021) 80(1) Cambridge Law Journal 42, 43.

[6] Raso (n 1) 29.

[7] Nikita Aggarwal, 'Law and Autonomous Systems Series: Algorithmic Credit Scoring and the Regulation of Consumer Credit Markets' (*University of Oxford/Faculty of Law*, 1 November 2018) <https://www.law.ox.ac.uk/business-law-blog/blog/2018/11/law-and-autonomous-systems-series-algorithmic-credit-scoring-and> accessed 21 March 2022.; World Bank Group (n 3) 9–10.

1.2 Problem statement

Automating decision-making is a double-edged sword. On the one hand, machines are fast, powerful, and efficient.[8] On the other hand, they can be just as error-prone, arbitrary, and biased as humans.[9] For this thesis, three types of concerns warranting the regulation of ADM and the tools enabling it are relevant. First, it can be argued that allowing a machine to make decisions concerning humans poses a threat to individuals' identity and decision-making autonomy.[10] This is because a machine may fail to treat an individual as an individual by drawing a conclusion about them based on the characteristics they share with others, may misrepresent their behaviour and actions by making generalisations based on data about them, and may limit their freedom by limiting the opportunities available to them because of their 'profile'.[11] Second, replacing a human decision-maker with a machine can also mean eliminating their role of using cultural knowledge to consider additional information when reaching a particular conclusion, if this is necessary in order to make a socially or legally justifiable decision.[12] This may, in turn, lead to extreme errors and decisions without legitimate justifications, thus calling into question the decisional system's legitimacy.[13] Third, due to the biases of its programmers and biases embedded in datasets used for its training, a machine may generate outputs recreating social prejudice or historical discrimination, and even give rise to new grounds of unfavourable treatment.[14]

Allowing threats to individuals' identity and decision-making autonomy, the legitimacy of decisional systems, and individuals' right to non-discrimination is incompatible with any democratic society that upholds the values and objectives of the Rule of Law, such as fairness, human rights, and human flourishing.[15] The Rule of Law, as a principle of governance applicable to public and private entities,[16] enables individuals to flourish by ensuring a just society,[17] in which they can freely develop their identity, personality, social relations, and

---

[8] Ari Ezra Waldman, 'Power, Process, and Automated Decision-making' (2019) 88(2) Fordham Law Review 613, 614.

[9] ibid.

[10] Margot E. Kaminski, 'Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability' (2019) 92(6) Southern California Law Review 1529, 1541–1542.

[11] ibid 1541–1544.

[12] ibid 1546.

[13] ibid 1545; 1547.

[14] ibid 1540.

[15] Greenstein (n 3) 2–3; 26.

[16] ibid 6.

[17] ibid 7.

pursue their goals. Although access to credit is not a recognised human right, it plays a key role in this respect and can thus be considered a democratic sub-value.

The social and democratic importance of access to credit and financial inclusion is best pointed out by Brownlee and Stemplowska, who argue that ‘there is a non-contingent link between financial inclusion and social options, political options, education and training options, marriage and family options, security, access to continued employment, and the capacity to save and to plan for a future.’[18] Inequality in access to credit due to errors or algorithmic discrimination can thus exacerbate existing social inequalities and consequently undermine individuals’ personal freedom and development. For this reason, the use of ML algorithms to assess individuals’ creditworthiness cannot be seen solely as a matter of private law.

As algorithmic credit scoring involves the processing of personal data, plays a role in the conclusion of consumer credit agreements, and affects individuals’ rights to non-discrimination, privacy, and data protection, it triggers the application of various pieces of legislation. This thesis will thus consider the regulation of the practice in three contexts: consumer protection, data protection, and AI safety, all within the broader fundamental rights framework. One of the central questions that will be answered is how the mode of data preparation and ML operation, the computation of human identity,[19] and credit scoring using mathematical processes where a large amount of human labour is replaced with machine operation[20] threaten respect for the aforementioned fundamental rights. Based on those findings, this thesis will evaluate the extent to which the Consumer Credit Directive,[21] General Data Protection Regulation,[22] and recent EU legislative proposals – the Proposal for a Directive

---

[18] Kimberley Brownlee and Zofia Stemplowska, ‘Financial Inclusion, Education, and Human Rights’ in Tom Sorell and Luis Cabrera (eds), *Microfinance, Rights and Global Justice* (Cambridge University Press 2015) 55.

[19] Mireille Hildebrandt, ‘Privacy As Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning’ (2019) 20(1) Theoretical Inquiries in Law 83.

[20] Katsundo Hitomi, ‘Automation—Its Concept and a Short History’ (1994) 14(2) Technovation 121, 123.

[21] Directive 2008/48/EC of the European Parliament and of the Council of 23 April 2008 on credit agreements for consumers and repealing Council Directive 87/102/EEC [2008] OJ L133/66 [hereinafter: Consumer Credit Directive or ‘CCD’].

[22] Article 29 Data Protection Working Party, ‘Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679’ (No 17/EN, 6 February 2018) p 19; Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 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.

on consumer credits[23] and the Artificial Intelligence Act Proposal[24] – can mitigate the risks of algorithmic credit scoring.

In this respect, the Consumer Credit Directive and the General Data Protection Regulation appear hardly capable of ensuring respect for individuals' fundamental rights. The Consumer Credit Directive, in fact, is based on traditional credit scoring,[25] and the prohibition on automated individual decision-making set forth in Article 22 of the General Data Protection Regulation can be circumvented, with the Regulation failing to provide sufficiently strong countermeasures. However, the bigger problem is that the currently applicable legislation does not specifically address the risks stemming from the design and development of ML models. In other words, it does not take into account the role humans play during the construction, training, and deployment of algorithms[26] in ensuring respect for individuals' fundamental rights in algorithmic credit scoring.

Although the use of alternative data or oversight mechanisms for credit scoring could be subject to process-centred obligations, these would thus not address data scientists' decisions regarding the balancing of a data sample, the assignment of a 'fair' weight to input variables, or the interpretability of ML models. Since such decisions also affect the respect for fundamental rights, the aim of this thesis is to identify gaps in the legislation and therefrom suggest EU-level solutions, targeting both the process of and tool for algorithmic credit scoring.

1.3 Research question

All above considered, this thesis will answer the following research question: *What risks does algorithmic credit scoring pose to the respect for individuals' rights to non-discrimination, privacy, and data protection, and what process- and tool-centred solutions to mitigate them can be envisioned at the EU level?*

---

[23] Commission, 'Proposal for a Directive of the European Parliament and of the Council on consumer credits' COM (2021) 347 final [hereinafter: Proposal for a Directive on consumer credits].

[24] Commission, 'Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts' COM (2021) 206 final [hereinafter: AI Act Proposal].

[25] Proposal for a Directive on consumer credits, Explanatory Memorandum 1, stating that 'Since the adoption of the 2008 Directive, digitalisation has profoundly changed the decision-making process [...] Digitalisation has also brought new ways of disclosing information digitally and assessing the creditworthiness of consumers using automated decision-making systems and non-traditional data.'

[26] Emily Berman, 'A Government of Laws not Machines' (2018) 98(5) Boston University Law Review 1277, 1325.

Accordingly, these sub-questions will be answered first:

1. What technology or tool enables algorithmic credit scoring, and how does the use of that tool impact the interpretation of the rights to non-discrimination, privacy, and data protection as protected in the EU?
2. How does algorithmic credit scoring affect individuals' access to credit, their private life, and personal data, and why does that pose a risk to the respect for their fundamental rights?
3. What EU legislation regulates the process of, or the tool for, algorithmic credit scoring, and what gaps can be identified in the legislation in regard to ensuring respect for individuals' fundamental rights?
4. What process- and tool-centred solutions could be employed with a view to filling the gaps in the legislation and thus ensuring respect for fundamental rights in algorithmic credit scoring, and where could they be regulated?

1.4 Research approach and methods, and thesis structure

For the purposes of this thesis, I will take a desk approach to research; I will collect and analyse legislation and other sources, which I will access either in the library or online. To identify legal sources, I will use the doctrinal legal method that focuses on the letter of the law.[27] To find other relevant sources, I will use the documentary method, which considers documents like scholarly articles as source material.[28]

To describe the technology enabling algorithmic credit scoring in Chapter II, I will use the descriptive method. In this chapter, I will also conceptualise the rights to non-discrimination, privacy, and data protection and present their protection in the EU, for which I will use the legal doctrinal method. In Chapter III, I will use the descriptive method to present the impact of algorithmic credit scoring on individuals' access to credit, their private life, and personal data, and the evaluative method to analyse the risks this practice poses to the respect for their rights to non-discrimination, privacy, and data protection. The legal doctrinal method will allow me to describe the legislative framework for algorithmic credit scoring in Chapter IV. To identify gaps in the relevant legislation in regard to ensuring respect for fundamental rights, I will use the evaluative method. Finally, in Chapter V, I will use the normative method to

---

[27] Terry Hutchinson, 'The Doctrinal Method: Incorporating Interdisciplinary Methods in Reforming the Law' (2015) 8(3) Erasmus Law Review 130, 131.

[28] Oxford Reference, 'Documentary Research' (*Oxford Reference*) <https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095724431> accessed 21 March 2022.

suggest process- and tool-centred solutions for ensuring respect for individuals' fundamental rights in algorithmic credit scoring.

1.5 Academic relevance of research

My research will build on previous findings on the implications of ADM for individuals' fundamental rights and proposals on how to ensure respect for them, which I will extend into the context of algorithmic credit scoring. My thesis will contribute to this field in three ways. First, I will analyse the impact of the use of ML algorithms on the interpretation of the rights to non-discrimination, privacy, and data protection as protected in the EU. Second, in analysing the extent to which relevant EU legislation can mitigate the risks of algorithmic credit scoring, I will consider the regulation of the practice in three contexts: consumer protection, data protection, and AI safety. In doing so, I will also review the recent legislative proposals for a Directive on consumer credits and the Artificial Intelligence Act, which have not yet been extensively discussed. Lastly, taking into account the European Data Protection Supervisor's Opinion on the Proposal for a Directive on consumer credits[29] and both legal and non-legal materials, I will propose both process- (ADM) and tool-centred solutions (ML algorithms) for ensuring respect for fundamental rights in algorithmic credit scoring; I believe, in fact, that tool-centred solutions have not been sufficiently explored.

---

[29] European Data Protection Supervisor, 'Opinion 11/2021 on the Proposal for a Directive on consumer credits' (26 August 2021).

II: Algorithmic credit scoring and fundamental rights

2.1 Algorithmic credit scoring

ADM can be defined as a process of deciding where a machine reaches a conclusion without human intervention regarding the decision itself.[30] ADM systems are often powered by ML algorithms that reach conclusions based on their analysis of extremely large and complex data sets, collectively known as Big Data.[31] What sets these algorithms apart from data mining algorithms is their ability to 'learn' from the data and use this knowledge to predict future outcomes.[32] Even though they do not possess the human cognitive abilities associated with a learning process, they are 'learning' by adjusting their performance according to their experience.[33] In other words, they 'program themselves over time with the rules to accomplish a task, rather than being programmed manually with a series of predetermined rules'[34] based on their analysis of incoming data.[35]

Nowadays, ML algorithms are increasingly used before concluding consumer credit agreements as a tool to assess individuals' creditworthiness or how 'financially sound [they are] to justify the extension of credit', i.e. loans or other similar forms of financial accommodation.[36] This is referred to as algorithmic credit scoring.[37] Algorithmic credit scoring builds on traditional credit scoring by using ML algorithms to analyse a greater amount of more diverse data, including so-called 'alternative data'.[38] The term is used to refer to 'the massive volume of data that is generated by the increasing use of digital tools and information systems',[39] which can be financial but non-credit data, such as rental and mobile bill payment

---

[30] Merriam-Webster Dictionary, 'Decision-making' (*Merriam-Webster*) <https://www.merriam-webster.com/dictionary/decision-making> accessed 21 March 2022; Merriam-Webster Dictionary, 'Decision' (*Merriam-Webster*) <https://www.merriam-webster.com/dictionary/decision> accessed 21 March 2022; Information Commissioner's Office, 'Automated Decision-making and Profiling' (*ico*) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/> accessed 21 March 2022; Hitomi (n 20) 122.

[31] Waldman (n 8) 614; European Parliament, 'Big Data: Definition, Benefits, Challenges (infographics)' (*European Parliament*, 29 March 2021) <https://www.europarl.europa.eu/news/en/headlines/society/20210211STO97614/big-data-definition-benefits-challenges-infographics> accessed 21 March 2022.

[32] Berman (n 26) 1279.

[33] Harry Surden, 'Machine Learning and Law' (2014) 89(1) Washington Law Review 87, 89.

[34] ibid 94.

[35] ibid.

[36] CCD, art 8; art 3(c); Merriam-Webster Dictionary, 'Creditworthy' (*Merriam-Webster*) <https://www.merriam-webster.com/dictionary/creditworthy> accessed 21 March 2022.

[37] Aggarwal (n 5).

[38] ibid 46.

[39] World Bank Group (n 3) 9.

history, as well as non-financial, non-credit data.[40] Such non-financial data can consist of social media data, such as the number of posts a user makes and their frequency,[41] or psychometric data inferred from a social media user's network and connections.[42]

When algorithms are tasked with predicting the likelihood of defaulting on credit repayments, they are normally first fed a chosen training dataset where the data's features constituted by a set of attributes of data instances[43] and the desired output (the target variable) are known.[44] This serves the purpose of a process known as supervised learning.[45] The goal of supervised learning is to train an algorithm to recognise patterns in the training data, thus generating an internal model ('model') with the ability to produce the desired outcome by capturing the relationships between features in the data that the algorithm have not seen before and the target variable.[46] At this stage of the process, data scientists are tasked with monitoring the model's functioning and deciding whether it is correctly detecting relevant patterns, usually by using a separate validation dataset;[47] if so, positive feedback is given to the model, which allows it to improve its performance.[48] Once a model that can satisfactorily produce the desired output has been generated, the performance of which is ultimately evaluated using a testing dataset,[49] it can finally be deployed outside the controlled context of a data lab.[50]

Thus, although such models' ability to automate predictions implies – figuratively speaking – a certain degree of autonomy and intelligence,[51] their predictions are essentially pre-determined by the human decisions and assumptions involved in the construction, training, and oversight of the models' functioning.[52] This also includes calibrating the weight of the data's features, interpreting the model's outputs,[53] and monitoring the model's performance 'in the wild', finally concluding as to whether it continues to be sufficiently reliable.[54] The role of humans in ADM is significant even if ML algorithms adopt a deep learning (DL) architecture

---

[40] Aggarwal (n 7).

[41] ibid; World Bank Group (n 3) 11.

[42] World Bank Group (n 3) 12.

[43] European Telecommunications Standards Institute, 'Experiential Networked Intelligence (ENI); Definition of Data Processing Mechanisms' (ETSI GR ENI 009 V1.1.1, June 2021) 21 <https://www.etsi.org/committee/1423-eni> accessed 15 June 2022.

[44] Berman (n 26) 1286–1287.

[45] ibid.

[46] Surden (n 33) 93; Berman (n 26) 1287.

[47] European Telecommunications Standards Institute (n 43); AI Act Proposal, art 3(30).

[48] Janneke Gerards and Raphaële Xenidis, *Algorithmic Discrimination in Europe: Challenges and Opportunities for Gender Equality and Non-discrimination Law* (Publications Office of the European Union 2021) 35.

[49] European Telecommunications Standards Institute (n 43); AI Act Proposal, art 3(31).

[50] Gerards and Xenidis (n 48).

[51] Surden (n 33) 88; 90.

[52] Kaminski (n 10) 1539.

[53] Berman (n 26) 1325–1326.

[54] Gerards and Xenidis (n 48) 40.

that enables automated data extraction from raw data.[55] This is because, even in those instances, humans develop the model and monitor the continued quality and validity of its outputs.[56] From this perspective, algorithmic credit scoring as a form of ADM can be seen as an expression of human decision-making power over the design and development of ML models transferred into code,[57] and the regulation of the practice seen as a way of controlling it.

2.2 Fundamental rights in the EU

According to Article 6 of the Treaty on European Union (hereinafter: TEU),[58] the sources of EU fundamental rights are the general principles of EU law, the Charter of Fundamental Rights of the European Union (hereinafter: CFR),[59] and the European Convention on Human Rights (hereinafter: ECHR).[60] As recognised in Article 2 TEU and the Preamble of the Statute of the Council of Europe, fundamental rights, the Rule of Law, and democracy are interdependent.[61] Fundamental rights ensure that individuals as citizens can participate in the making of binding collective decisions that in turn influence their fundamental rights,[62] thereby enabling democracy as a system of 'popular sovereignty' or 'the rule of the people, either by the people themselves or through others that are elected, influenced, and controlled by the people'.[63] The protection of fundamental rights is also essential for the full realisation of the Rule of Law,[64] which, in its original sense, stands for "the empire of laws and not of men",[65] emphasising that laws are to serve the public good instead of public officials' interests.[66]

---

[55] ibid 41; Jason Brownlee, 'What Is Deep Learning?' (*Machine Learning Mastery*, 16 August 2019) <https://machinelearningmastery.com/what-is-deep-learning/> accessed 3 May 2022.

[56] Gerards and Xenidis (n 48) 41.

[57] Waldman (n 8) 615.

[58] Consolidated Version of the Treaty on European Union [2012] OJ C326/1 [hereinafter: TEU].

[59] Charter of Fundamental Rights of the European Union [2010] OJ C83/2 [hereinafter: CFR].

[60] Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) [hereinafter: ECHR]; Sybe A. de Vries, 'Balancing Fundamental Rights with Economic Freedoms According to the European Court of Justice' (2013) 9(1) Utrecht Law Review 169, 177 <https://www.utrechtlawreview.org/articles/abstract/10.18352/ulr.220/> accessed 3 May 2022.

[61] Advisory Council on International Affairs, 'The Will of the People? The Erosion of Democracy under the Rule of Law in Europe' (No 104, 2017) 10 <https://www.advisorycouncilinternationalaffairs.nl/documents/publications/2017/06/02/the-will-of-the-people> accessed 3 May 2022.

[62] Leslie and others (n 1).

[63] Frank Hendriks, *Vital Democracy: A Theory of Democracy in Action* (Oxford Scholarship Online 2010) ch 1, p 5–6.

[64] Venice Commission, 'Rule of Law Checklist' (Study No 711/2013, 18 March 2016) 9.

[65] Mortimer N.S. Sellers, 'What Is the Rule of Law and Why Is It So Important?' in James R. Silkenat, James E. Hickey Jr. and Peter D. Barenboim (eds), *The Legal Doctrines of the Rule of Law and the Legal State (Rechtsstaat)* (Springer 2014) 4.

[66] ibid 3–4.

Respect for fundamental rights is one of the values of the Rule of Law aimed at creating a society in which individuals can reach their potential of meeting set goals.[67] Although fundamental rights documents were designed to protect individuals from the abuse of government power,[68] private entities 'may pose just as strongly a risk to the effective enjoyment of fundamental rights as public bodies and government agents' when wielding significant power.[69] While the European Court of Justice (hereinafter: ECJ) has so far been reluctant to allow fundamental rights to be invoked in horizontal relations outside the area of non-discrimination law[70] and to impose substantive positive obligations on the States,[71] it has often referred to the case-law of the European Court of Human Rights (hereinafter: ECtHR). The ECtHR has developed a full-fledged doctrine of positive obligations,[72] following which the States, including all 27 EU countries, must secure the rights enshrined in the ECHR in horizontal relations through legislation, measures and actions, or via national courts.[73]

The collecting and processing of personal data for the purpose of creditworthiness assessment is an example of how a private entity can pose a significant risk to individuals' effective enjoyment of fundamental rights.[74] This is because there is a significant power imbalance in the relationship between creditors and consumers, and algorithmic credit scoring is widening this gap by allowing creditors to gain insights about individuals that go beyond the limits of human observation. As this power is exercised for corporate interests,[75] such as preventing losses due to non-performing loans,[76] there is thus a risk that creditors would set a private standard of protection of the fundamental rights[77] to non-discrimination, privacy, and data protection in order to achieve corporate goals. In such horizontal relations, states thus ought to secure individuals' effective enjoyment of fundamental rights. Credit scores, however, determine individuals' access to credit and thus their ability to fully participate in society or improve their standard of living,[78] so algorithmic credit scoring can also significantly affect the

---

[67] Greenstein (n 3) 7.

[68] Janneke Gerards, *General Principles of the European Convention on Human Rights* (manuscript, Cambridge University Press 2019) 110.

[69] ibid.

[70] Steven Greer, Janneke Gerards and Rose Slowe, *Human Rights in the Council of Europe and the European Union* (Cambridge University Press 2018) 310–311.

[71] ibid 320–321.

[72] ibid 320.

[73] Gerards (n 68) 105; 118.

[74] ibid 110.

[75] Waldman (n 8) 616.

[76] Aggarwal (n 7).

[77] Oreste Pollicino and Giovanni De Gregorio, 'Constitutional Law in the Algorithmic Society' in Hans-W. Micklitz and others (eds), *Constitutional Challenges in the Algorithmic Society* (Cambridge University Press 2021) 7.

[78] AI Act Proposal, rec 37.

objects protected by other fundamental rights and, in turn, have implications for individuals' ability to reach their potential. Despite access to credit not being a right, its complementary role to the effective enjoyment of other fundamental rights can thus be recognised, which is why any state under the Rule of Law ought to ensure respect for fundamental rights in algorithmic credit scoring.

Bearing in mind the above, I will now turn to how the use of the technology enabling algorithmic credit scoring affects the interpretation of three EU fundamental rights, whose objects of protection are directly affected by the practice: the right to non-discrimination, the right to privacy, and the right to data protection.

2.2.1 Right to non-discrimination

1. Concepts

Three interrelated concepts are relevant when discussing discrimination by ML models: 'bias', discrimination, and 'machine fairness'. Firstly, in a computational context, 'bias' refers to a '"systematic error" of any kind in the outcome of algorithmic operations.'[79] Such a systematic error can be unjust if 'the outputs of an algorithm benefit or disadvantage certain individuals or groups more than others without a justified reason for such unequal impacts.'[80]

The concept of bias is usually associated with the replication and reinforcement of existing societal biases against underprivileged and marginalised communities.[81] In this sense, it shares similarities with the concept of discrimination,[82] which EU law describes as the treatment of a person less favourably than another in a comparable situation because of a protected characteristic (direct discrimination),[83] or putting persons who share a protected characteristic at a particular disadvantage compared with other persons by means of a provision, criterion, or practice, which, at first glance, appears as neutral (indirect discrimination).[84]

---

[79] Gerards and Xenidis (n 48) 47.

[80] Nima Kordzadeh and Maryam Ghasemaghaei, 'Algorithmic Bias: Review, Synthesis, and Future Research Directions' (2021) 31(3) European Journal of Information Systems 1 <https://www.tandfonline.com/doi/full/10.1080/0960085X.2021.1927212> accessed 5 May 2022 (emphasis omitted).

[81] ibid 1; 3.

[82] Gerards and Xenidis (n 48) 47.

[83] Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin [2000] OJ L180/22 [hereinafter: Racial Equality Directive], art 2(2)(a) – almost identical definitions can be found in directives 2000/78/EC, 2004/113/EC and 2006/54/EC.

[84] Racial Equality Directive, art 2(2)(b) – almost identical definitions can be found in directives 2000/78/EC, 2004/113/EC and 2006/54/EC.

Lastly, 'machine fairness' can be best defined as an umbrella term for various computational techniques aimed at minimising algorithmic bias.[85] In the context of a given task, ML engineers and data scientists must thus limit themselves to applying a set of selected statistical fairness criteria.[86] This implies a value-laden decision that may not be the same as the decision that would be taken by other stakeholders, such as regulators and the public.[87]

2. Extent of EU-level protection

Discrimination is prohibited by the ECHR, Protocol No. 12 to the ECHR, and the CFR. The principle of non-discrimination, as enshrined in Articles 2 and 3(3) TEU, is also one of the general principles of EU law. According to Article 14 ECHR, it is prohibited to discriminate 'on any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status.'[88] This prohibition, however, refers only to discrimination in the enjoyment of the rights and freedoms set forth in the Convention.[89] Conversely, Article 1(1) of Protocol No. 12 prohibits discrimination in relation to 'any right set forth by law', which may also be granted under national law.[90] Similarly, Article 21 CFR contains a stand-alone prohibition against '[a]ny discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation'[91] and, nationality.[92] Although this prohibition applies insofar as the States are implementing Union law, the non-discrimination provisions in these laws (e.g. directives) apply to private entities if their actions fall within the material scope of their application.[93]

---

[85] Gerards and Xenidis (n 48) 48.

[86] Deirdre K. Mulligan and others, 'This Thing Called Fairness: Disciplinary Confusion Realizing a Value in Technology' (2019) 3 Proceedings of the ACM on Human-Computer Interaction 1, 15–16 <https://dl.acm.org/doi/abs/10.1145/3359221> accessed 5 May 2022.

[87] ibid 4.

[88] ECHR, art 14.

[89] Registry of the European Court of Human Rights, 'Guide on Article 14 of the European Convention on Human Rights and on Article 1 of Protocol No. 12 to the Convention—Prohibition of discrimination' (updated on 31 August 2021, Council of Europe/European Court of Human Rights 2021), para 3.

[90] ibid 9.

[91] CFR, art 21(1).

[92] CFR, art 21(2).

[93] Frederik Zuiderveen Borgesius and Janneke Gerards, 'Protected Grounds and the System of Non-discrimination Law in the Context of Algorithmic Decision-making and Artificial Intelligence' (working draft, 2021) 1, 23–24 <https://works.bepress.com/frederik-zuiderveenborgesius/65/> accessed 6 May 2022; CFR, art 51.

As the language of these provisions makes clear, non-discrimination law relies on the notion of protected characteristics such as sex, race, and nationality.[94] In the case of Article 14 ECHR, Article 1 Protocol No. 12, and Article 21 CFR, the list of protected grounds of discrimination is not closed, as the provisions contain the wording 'such as'. Zuiderveen Borgesius and Gerards refer to this type of systems as 'hybrid systems' with a semi-closed list of grounds and a fully open possibility of exemptions from the prohibition of discrimination.[95] Such systems have the advantage of allowing courts to add other grounds, thereby encouraging societal and political debates on whether certain grounds should be added.[96] Nevertheless, there are limits to the possible grounds, as they must be compatible with, and follow the logic of, the provided list.[97]

Conversely, Zuiderveen Borgesius and Gerards explain that the relevant provisions in non-discrimination directives such as the Racial Equality Directive[98] are either 'fully closed' systems (direct discrimination) or hybrid systems with a closed list of grounds and mostly a fully open possibility of exemptions (indirect discrimination).[99] Such systems have significant drawbacks. Firstly, these systems can be unsuccessful in addressing intersectional discrimination, where 'two or multiple grounds operate simultaneously and interact in an inseparable manner, producing distinct and specific forms of discrimination.'[100] Secondly, these systems may not be able to tackle direct discrimination hidden behind seemingly neutral grounds.[101] Finally and most importantly, as these systems contain an exhaustive list of grounds, they cannot address the less favourable treatment of individuals based on unenumerated characteristics.

3. Interpretation in the context of the use of ML algorithms

As algorithmic bias can involve the less favourable treatment of individuals based on non-protected characteristics,[102] the differences in the 'openness' of systems have important

---

[94] Frederik Zuiderveen Borgesius, 'Discrimination, Artificial intelligence, and Algorithmic Decision-making' (Council of Europe, Directorate General of Democracy 2018) 20 <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73> accessed 5 May 2022.

[95] Zuiderveen Borgesius and Gerards (n 93) 30; 34.

[96] ibid 31; 39.

[97] ibid 39.

[98] See n 83.

[99] Zuiderveen Borgesius and Gerards (n 93) 34.

[100] Council of Europe, 'Intersectionality and Multiple Discrimination' (*Council of Europe*) <https://www.coe.int/en/web/gender-matters/intersectionality-and-multiple-discrimination> accessed 6 May 2022; Zuiderveen Borgesius and Gerards (n 93) 45.

[101] Zuiderveen Borgesius and Gerards (n 93) 48.

[102] Gerards and Xenidis (n 48) 47.

implications for the recognition of less favourable treatment in ADM as discrimination. These non-protected characteristics can be 'proxy variables' for protected characteristics with an (in)sufficient degree of overlap with the corresponding protected ground to be regarded as the same thing.[103] ML models can also generate outputs that give rise to the less favourable treatment of individuals based on newly invented classes, such as the type of web browser they use.[104] And, finally, individuals can be put in a disadvantageous position based on a combination of protected and non-protected characteristics.[105] In all these cases, algorithmic bias may escape current non-discrimination laws.[106] Thus, the right to non-discrimination cannot always be exercised and can protect only against limited instances of unfavourable treatment by ML models.

In some cases of algorithmic bias where the link between the proxy variable and the protected characteristic would prove insufficiently direct so as to establish direct discrimination, algorithmic discrimination could perhaps be recognised by applying the concept of indirect discrimination.[107] There is, however, a chance that the definition of protected grounds would be too narrow to subsume the proxy variable.[108] The Court of Justice of the European Union (hereinafter: CJEU), in fact, held in *Jyske Finans* that 'a person's country of birth cannot, in itself, justify a general presumption that that person is a member of a given ethnic group'.[109] Considering this reasoning, it is questionable whether postcode and residency data, which algorithms have used to infer people's ethnicity, would be recognised as proxy variables in particular cases.[110] Moreover, difficulties in recognising algorithmic discrimination could also arise from the very concept of indirect discrimination, which leaves the possibility of exemptions fully open;[111] namely, that a practice may be objectively justified by a legitimate aim and the means of achieving it may be appropriate and necessary.[112]

Nevertheless, ML models' ability to give rise to new grounds of unfavourable treatment suggests that a new system of non-discrimination law with a wider conception of (algorithmic) discrimination is needed. As previously explained, even in systems with a semi-closed list of

---

[103] ibid 63–64.

[104] Zuiderveen Borgesius (n 94) 35–36.

[105] Gerards and Xenidis (n 48) 65.

[106] Zuiderveen Borgesius (n 94) 5.

[107] Gerards and Xenidis (n 48) 71.

[108] ibid 63.

[109] Case C-668/15 *Jyske Finans A/S v Ligebehandlingsnaevnet, acting on behalf of Ismar Huskic* [2017] EU:C:2017:278, para 20.

[110] Gerards and Xenidis (n 48) 71.

[111] ibid 73.

[112] Racial Equality Directive, art 2(2)(b) – identical language can be observed in directives 2000/78/EC, 2004/113/EC and 2006/54/EC.

grounds, there are limits to what can be considered grounds of discrimination. For instance, according to the ECtHR, discrimination must be based on personal status or personal characteristics for Article 14 ECHR to apply.[113] The less favourable treatment of individuals based on new grounds such as the type of web browser they use thus cannot fit the current conception of discrimination.

The need for a new system of non-discrimination law is also supported by the fact that the CJEU has been reluctant to recognise multiple discrimination in practice.[114] For instance, in *Parris* the Court held that 'while discrimination may indeed be based on several (…) grounds (…) no new category of discrimination resulting from the combination of more than one of those grounds (…) may be found to exist where discrimination on the basis of those grounds taken in isolation has not been established.'[115] Hence, a ML model's outputs could cumulatively disadvantage[116] certain individuals or groups more than others without a justified reason for such unequal impacts; however, this would not be considered discrimination if discrimination could not be proven in relation to each ground, which, as explained above, would be particularly difficult if the unfair differentiation were based on proxy variables.

In short, when discussing the right to non-discrimination in the context of the use of ML algorithms, it is necessary to differentiate between algorithmic bias and algorithmic discrimination. The next section takes a look at the rights to privacy and data protection as protected in the EU, and considers how the use of ML algorithms affects their interpretation.

2.2.2 Rights to privacy and data protection

1. Concepts and extent of EU-level protection

Article 7 CFR corresponds to Article 8 ECHR, which states that '[e]veryone has the right to respect for his private and family life, his home and his correspondence',[117] also referred to as the right to privacy. 'Privacy', entailing but not limited to values legally protected under the right to privacy,[118] is a complex and multi-faceted concept shaped by and changing together with societal norms.[119] The ECtHR defines the concept through a 'pragmatic, common-sense

---

[113] Zuiderveen Borgesius and Gerards (n 93) 31.

[114] Gerards and Xenidis (n 48) 65.

[115] Case C-443/15 *David L. Parris v Trinity College Dublin and Others* [2016] EU:C:2016:897, para 80.

[116] Council of Europe (n 100).

[117] ECHR, art 8(1).

[118] Bert-Jaap Koops and others, 'A Typology of Privacy' (2017) 38(2) University of Pennsylvania Journal of International Law 483, 491–492.

[119] Judith DeCew, 'Privacy' (*Stanford Encyclopedia of Philosophy*, 18 January 2018) <https://plato.stanford.edu/entries/privacy/> accessed 16 June 2022, quoting Daniel Solove.

approach rather than a formalistic or purely legal one',[120] which has allowed the Court to move away from the classic interpretation of the right to privacy as a tool to protect individuals against unlawful state interference in their private sphere and now interpret it (inter alia) as a personality right serving to protect individuals' development of their identity and personality and thus their dignity.[121] Accordingly, protection is not afforded only in vertical relations between individuals and public authorities,[122] but, as the ECtHR held in *Bărbulescu v. Romania*, the States must ensure effective respect for the right to privacy in horizontal relations.[123]

In addition to the right to privacy, EU law also guarantees a traditionally different but related 'right to the protection of personal data'.[124] This right differs from the right to privacy in that the responsibilities in relation to the processing of personal data mainly derive from secondary legislation (the General Data Protection Regulation), the right cannot be invoked by legal persons,[125] and it has distinct elements, such as the requirements to process personal data fairly, for specified purposes, and to ensure the data subject's right of access to the data.[126] These elements justify the need for a stand-alone right and are important for protecting the aspect of privacy known as informational privacy.[127] Koops and others list informational privacy as one of the eight primary ideal types of privacy, together with bodily, spatial, communicational, proprietary, intellectual, decisional, associational, and behavioural privacy.[128] Associational, behavioural, and informational privacy are particularly relevant for this thesis, as the objects of protection of the rights to privacy and data protection which are directly affected by algorithmic credit scoring, protect these types or aspects of privacy.

Regarding associational privacy, Koops and others write that this aspect of privacy is characterised by 'individuals' interests in being free to choose who they want to interact

---

[120] *Botta v Italy* App no 21439/93 (ECtHR, 24 February 1998), para 27.

[121] Bart van der Sloot, 'Privacy as Personality Right: Why the ECtHR's focus on Ulterior Interests Might Prove Indispensable in the Age of "Big Data" ' (2015) 31(80) Utrecht Journal of International and European Law 25–26 <https://utrechtjournal.org/articles/10.5334/ujiel.cp/> accessed 9 May 2022; Registry of the European Court of Human Rights, 'Guide on Article 8 of the European Convention on Human Rights—Right to respect for private and family life, home and correspondence' (updated on 31 August 2021, Council of Europe/European Court of Human Rights 2021), para 5.

[122] Registry of the European Court of Human Rights (n 121).

[123] *Bărbulescu v Romania* App no 61496/08 (ECtHR, 5 September 2017), paras 108–111.

[124] CFR, art 8(1); van der Sloot (n 121) 39–40.

[125] Juliane Kokott and Christoph Sobotta, 'The Distinction Between Privacy and Data Protection in the Jurisprudence of the CJEU and the ECtHR' (2013) 3(4) International Data Privacy Law 222, 225.

[126] CFR, art 8(2).

[127] Yvonne McDermott, 'Conceptualising the Right to Data Protection in an Era of Big Data' (2017) 4(1) Big Data & Society 1, 2 <https://journals.sagepub.com/doi/10.1177/2053951716686994> accessed 10 May 2022.

[128] Koops and others (n 118) 566–569.

with'[129] or, in other words, to be able to associate with whomever they choose without being monitored.[130] This type of privacy takes place in the 'semi-private zone', characterised by actions and communications in semi- or quasi-public spaces such as offices, meeting places, or cafés.[131] In this respect, it is similar to behavioural privacy, which is characterised by individuals' interest to remain 'hidden' while carrying out publicly visible activities.[132] Thus, although behaviour in the public zone cannot be completely excluded from observation by others, individuals have an interest in others 'seeing [them] but not taking notice (or perhaps rather, demonstrating not to take notice)'.[133]

The individuals' interest in remaining 'inconspicuous among the masses'[134] has also been recognised by the ECtHR, which held that '[t]here is (…) a zone of interaction of a person with others, even in a public context, which may fall within the scope of "private life".'[135] The Court confirmed this in *Peck v. the United Kingdom* and *Von Hannover v. Germany*,[136] reiterating in the latter case that the concept of private life or privacy 'includes a person's physical and psychological integrity; the guarantee afforded by Article 8 of the Convention is primarily intended to ensure the development, without outside interference, of the personality of each individual in his relations with other human beings'.[137] Accordingly, it can be said that the objects of protection associated with associational and behavioural privacy are social relations (particularly in connection to associational privacy) and autonomy as free identity- and personality-building and free decision-making.[138]

Finally, Koops and others conceptualise informational privacy as an overarching aspect of each type of privacy, which can concern information relating to any of the four private zones, namely the private (solitude), intimate (small unit of social interactions), semi-private, and public zone.[139] According to Koops and others, this type of privacy is characterised by 'the interest in preventing information about one-self to be collected and in controlling information about one-self that others have legitimate access to.'[140] Informational privacy has also been

---

[129] ibid 568.
[130] ibid 503.
[131] ibid 551; 568.
[132] ibid 568.
[133] ibid 552; 568.
[134] ibid 568.
[135] *P.G. and J.H. v the United Kingdom* App no 44787/98 (ECtHR, 25 September 2001), para 56.
[136] *Peck v the United Kingdom* App no 44647/98 (ECtHR, 28 January 2003), para 57; *Von Hannover v. Germany* App no 59320/00 (ECtHR, 24 June 2004), para 50.
[137] *Von Hannover v. Germany* [50].
[138] Koops and others (n 118) 542; Registry of the European Court of Human Rights (n 121), para 74.
[139] Koops and others (n 118) 545–554; 568–569.
[140] ibid 568.

referred to by other scholars as ‘privacy of personal data’[141] and ‘privacy of data and image’,[142] and resonates with the notion of data protection.[143]

Although the right to privacy is narrower than the concept of (informational) privacy and it is therefore possible to envision a situation where information relating to an identified or identifiable natural person (‘personal data’)[144] would be excluded from the scope of private life,[145] if personal data are collected on a precise individual, the data operation will most likely fall within said scope.[146] As the ECtHR held in *Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland*, in such a case, the protection of informational privacy is ‘of fundamental importance to a person’s enjoyment of his or her right to respect for private and family life’,[147] therefore Article 8 ECHR ‘provides for the right to a form of informational self-determination, allowing individuals to rely on their right to privacy as regards data’.[148]

Thus, if informational privacy is considered in its overarching nature, the rights to privacy and data protection can be understood as rights that protect individuals’ control over how they present themselves to the world through information about them, thus allowing them to protect their Selves from the world and freely develop their identity and personality.[149]

2. Interpretation in the context of the use of ML algorithms

I believe that individuals’ control over how they present themselves to the world through information about them acquires a new meaning in the context of the use of ML algorithms, which affects the objects that the rights to privacy and data protection seek to protect so as to safeguard individuals’ informational privacy. The reason for this is that ‘self-presentation involves not only managing the expressions of the self that one *gives*, but also those one *gives off*’,[150] for instance, through his appearance and behaviour.[151] Koops explains that while individuals are generally aware of the expressions they give off while they are offline and how these may impact their future social relations, they give off expressions online through the data

---

[141] ibid 499
[142] ibid 502.
[143] ibid 499.
[144] GDPR, art 4(1).
[145] See e.g. Registry of the European Court of Human Rights, ‘Guide to the Case-Law of the European Court of Human Rights—Data protection’ (updated on 31 December, Council of Europe/European Court of Human Rights 2021), para 11.
[146] ibid para 12.
[147] *Satakunnan Markkinapörssi Oy and Satamedia Oy v Finland* App no 931/13 (ECtHR, 27 June 2017), para 137.
[148] ibid.
[149] van der Sloot (n 121) 26–27.
[150] Bert-Jaap Koops, ‘Privacy Spaces’ (2018) 121(2) West Virginia Law Review 611, 656 (emphasis in original).
[151] ibid.

inferred from their activities usually without realising it, and, even if they do, they are unable to adjust their behaviour because they neither understand the consequences of such expressions nor know how to avoid them.[152] Taking this in the context of the use of ML algorithms, the algorithms are not only able to infer data from seemingly neutral online behaviour (e.g. using a certain web browser) but also from actions in 'real life' (e.g. shopping at a particular chain of stores[153]), and individuals cannot control such inferences.

Moreover, the individualistic shaping of one's image is unattainable in the context of the use of ML algorithms. This is because, in a ML system, 'individuals also possess a profiling identity constructed from connections with groups of other data subjects based upon dimensions (e.g. behaviours, demographic attributes) deemed relevant'.[154] In other words, the features that define the groups into which the algorithm places them, which Mittelstadt refers to as 'behavioural identity tokens', affect their image.[155] As Mittelstadt explains, the actions of members of a particular group can change the tokens, which in turn affects other members, for example, by making them appear less creditworthy.[156] Yet, even though the characteristics that ML algorithms attribute based on group membership may be untrue and disputable, individuals cannot have control over them in advance. For this reason, I argue that in the context of the use of ML algorithms, informational privacy refers to one's ability to be aware of their algorithmic identity and have the power to contest it.

In sum, the rights to privacy and data protection also protect the development of individuals' identity and personality through the protection of their informational privacy as an overarching aspect of each type or aspect of privacy. Due to the ability of ML algorithms to infer data from (data about) seemingly neutral behaviour and actions and group membership, individuals cannot, in the as yet understood sense of it, control how they present themselves to the world through information about them, and thus, informational privacy takes on a new meaning in this context. Against this background, I now turn to how algorithmic credit scoring affects individuals' access to credit, their private life, and personal data, and why that poses a risk to the respect for their rights to non-discrimination, privacy, and data protection.

---

[152] ibid 657.

[153] White & Case LLP, 'Algorithms and Bias: What Lenders Need To Know' (*White & Case*, 20 January 2017) <https://www.whitecase.com/publications/insight/algorithms-and-bias-what-lenders-need-know> accessed 10 May 2022.

[154] Brent Mittelstadt, 'From Individual to Group Privacy in Big Data Analytics' (2017) 30(4) Philosophy & Technology 475, 478.

[155] ibid.

[156] ibid 478–479.

III: Risks of algorithmic credit scoring

3.1 Risks to individuals' right to non-discrimination

Algorithmic credit scoring can allow 'thin-file' applicants who lack credit history and would thus likely not qualify for a loan based on traditional credit scoring to access credit by being assessed based on alternative data.[157] As access to credit affects individuals' ability to improve their standard of living, enabling them to qualify for credit despite having a thin file supports their right to an adequate standard of living. In addition, because of historical discrimination, members of marginalised groups could lack credit history, so algorithmic credit scoring could also promote equality.[158] Finally, assessing applicants' creditworthiness based on data rather than human judgment could also reduce bias in the credit-granting process, thereby potentially safeguarding individuals' right to non-discrimination. Yet, individuals can nonetheless be discriminated against by ML models and thus denied access to credit.

In this respect, there are two challenges in terms of enabling individuals' access to credit. The first concerns preventing discrimination in algorithmic credit scoring, whereas the second concerns detecting it after it has occurred. Credit scores that give rise to discrimination can be, in fact, difficult to detect and contest on this basis due to the opacity of ML models' functioning. This can be attributable to the intrinsic opacity of a ML model, which stems from the fact that humans reason differently from machines and therefore cannot fully understand how complex systems like artificial neural networks work.[159] For this reason, such systems are often described as 'black boxes', as their inputs and outputs can be observed, but not the in-between process.[160] Humans thus cannot always pinpoint the input variables that determined a ML model's inference or understand why the model predicted an event the way it did[161] – nor can the model shed light on its operation by providing reasons for its findings.[162] However, the functioning of ML models is not necessarily opaque because of their intrinsic opacity, as not

---

[157] Aggarwal (n 7).

[158] Raso and others (n 1) 26.

[159] Monika Zalnieriute, Lyria Bennett Moses and George Williams, 'Automating Government Decision-making: Implications for the Rule of Law' in Siddharth Peter de Souza and Maximilian Spohr (eds), *Technology, Innovation and Access to Justice: Dialogues on the Future of Law* (Edinburgh University Press 2021) 98; Madalina Busuioc, 'Accountable Artificial Intelligence: Holding Algorithms to Account' (2020) 81(5) Public Administration Review 825, 829–830.

[160] Busuioc (n 159); Dallas Card, 'The "Black Box" Metaphor in Machine Learning' (*Medium*, 5 July 2017) <https://dallascard.medium.com/the-black-box-metaphor-in-machine-learning-4e57a3a1d2b0> accessed 16 June 2022.

[161] ibid 829.

[162] Berman (n 26) 1352.

every ML model is a black box;[163] the opacity of a ML model's functioning can be intentional, as its logic may be protected as a trade secret.[164] In any event, the opacity of ML models' functioning presents a barrier to individuals understanding how their credit score has been calculated and determining whether they have been discriminated against.

Discrimination in algorithmic credit scoring can be blamed on different actors and can be caused by various factors[165] that play a role during the preparation for and subsequent training of an algorithm, selection of a model, and in the decision-making stage of an ADM process. Starting with the preparation for and subsequent training of an algorithm, data scientists must first define the target variable (the desired output, e.g. creditworthiness) and its associated class labels. The value of the target variable is to be predicted based on other features (independent variables) in a dataset, i.e. a collection of data points or observation records consisting of features.[166] The possible values of this feature are divided into mutually exclusive classes or categories by labels; hence they are referred to as class labels.[167] In order to be able to train an algorithm, data scientists must also prepare the training dataset and carry out a process known as 'feature selection'. This is the process of deciding on the features or variables to be used for ML with the aim of selecting a set of features from the (training) dataset used solely for training the algorithm, which will allow the model to make optimal predictions.[168] Factors associated with all of these tasks can result in credit scores that give rise to discrimination.[169]

Starting with the first task, data scientists must define the target variable by translating the problem into computer-readable language,[170] which becomes challenging when they need to create new classes, as is the case for creditworthiness.[171] This is because 'creditworthiness' is not a binary problem with a well-understood definition such as 'it is, or it is not spam'.[172] Instead, data scientists need to rely on criteria developed by the credit industry or the creditor that are used to decide whether to approve or reject a credit application (e.g. not missing credit repayments).[173] Choices made at this stage can be the cause of algorithmic discrimination

---

[163] Busuioc (n 159) 829.

[164] Jenna Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3(1) Big Data & Society 1, 3 <https://journals.sagepub.com/doi/full/10.1177/2053951715622512> accessed 14 May 2022.

[165] Gerards and Xenidis (n 48) 37.

[166] Raghav Bali and Dipanjan Sarkar, *R Machine Learning by Example* (Packt 2016) 181; 191.

[167] Solon Barocas and Andrew D. Selbst, 'Big Data's Disparate Impact' (2016) 104(3) California Law Review 671, 678; Bali and Sarkar (n 166) 181.

[168] Bali and Sarkar (n 166) 188; 191.

[169] Barocas and Selbst (n 167) 677–690.

[170] Barocas and Selbst (n 167) 678; Gerards and Xenidis (n 48) 39.

[171] Barocas and Selbst (n 167) 679; regarding the definition of creditworthiness, See n 36.

[172] Barocas and Selbst (n 167) 678–679.

[173] ibid 679.

because if certain behaviour is systematically higher among members of groups defined by a protected characteristic, a class label correlating to such behaviour puts these groups in a disadvantageous position.[174]

Next, algorithmic discrimination can result from biased training data,[175] which reflects a concept known as 'GIGO' ('garbage in, garbage out').[176] Firstly, such biased data can be the consequence of choices concerning data labels.[177] This is because a labelled dataset, i.e. a dataset in which the raw data's features have been identified by humans who have chosen and assigned labels to describe those features accordingly, serves as a standard or 'ground truth' against which data scientists assess a ML model's performance.[178] Since the labelled dataset is considered the ground truth, the accuracy of the model ultimately depends on the accuracy of the labels.[179] Therefore, if the labels reflect bias, the model will generate biased outputs that may amount to discrimination.[180]

Secondly, raw data can also be biased[181] – the records relied upon by data scientists to build a training dataset can, in fact, contain past discriminatory decisions, which, for example, speak of systemic rejection of credit applications of members of groups defined by a protected characteristic.[182] Moreover, due to past selective observation/reporting, such as overreporting defaults by members of certain groups, the records can under- or overrepresent groups defined by a protected characteristic.[183] Such imbalanced data inevitably cause a ML algorithm to draw skewed lessons.[184] For instance, if the training dataset overstated the relative incidence of default by members of a group defined by a protected characteristic, the algorithm would learn that they are more likely to default, which could lead to a low credit score for an otherwise creditworthy member of that group.[185] In short, if the inputs are 'garbage' or biased, a ML model will generate 'garbage' or biased outputs that may give rise to discrimination.[186]

---

[174] ibid 680.
[175] ibid.
[176] ibid 683.
[177] ibid 681.
[178] Amazon Web Services, 'What Is Data Labeling for Machine Learning?' (*aws*) <https://aws.amazon.com/sagemaker/data-labeling/what-is-data-labeling/> accessed 15 May 2022; CloudFactory, 'The Ultimate Guide to Data Labeling for Machine Learning' (*cloudfactory)* <https://www.cloudfactory.com/data-labeling-guide> accessed 15 May 2022.
[179] Amazon Web Services (n 178).
[180] Barocas and Selbst (n 167) 683–684.
[181] ibid 682.
[182] ibid.
[183] ibid 684; 687.
[184] ibid 687.
[185] ibid.
[186] ibid 683–684.

Yet another factor that can have severe implications for the treatment of groups defined by a protected characteristic is the choice of features or variables and their assigned weight or importance.[187] This is because the selected features can reflect inaccurate generalisations and thus not reflect statistical differences between members of such groups.[188] Such features may also appear neutral despite being, in fact, proxy variables for protected characteristics.[189] For example, individuals can be assigned a lower credit score for shopping at low-end retailers.[190] If such stores were disproportionately located in minority communities, the ML model could incorrectly and systematically deem individuals who shop there as less creditworthy for behaviour associated with a group defined by a protected characteristic,[191] which Hurley and Adebayo have termed 'creditworthiness by association'.[192]

Nevertheless, proxy discrimination can occur even if the features are sufficiently granular.[193] In other words, features that hold 'demonstrable and justifiable relevance to the decision at hand'[194] can lead to systematically less favourable determinations for members of groups defined by a protected characteristic, for example, because they share these features to a lesser extent.[195] Barocas and Selbst attribute this form of algorithmic discrimination to redundant encoding, i.e. cases where a particular piece of data is highly correlated with group membership.[196] For example, although sex can be omitted as an input variable, it can be inferred from the fact that someone is a single parent.[197] As Barocas and Selbst explain, withholding such variables can prevent discriminatory outcomes but at the expense of the overall accuracy of a ML model's determinations.[198] Yet, even if the aim were to remove all correlated features, this could prove a particularly difficult task, as there is no straightforward answer to the question of how correlated with a protected characteristic a feature needs to be in order to become a concern for discrimination.[199] And it is precisely this indeterminacy, together with

---

[187] ibid 687; Mikella Hurley and Julius Adebayo, 'Credit Scoring in the Era of Big Data' (2016) 18(1) Yale Journal of Law and Technology 148, 180.

[188] Barocas and Selbst (n 167) 688–689.

[189] ibid 691.

[190] Hurley and Adebayo (n 187) 183.

[191] White & Case LLP (n 153); Barocas and Selbst (n 167) 689.

[192] Hurley and Adebayo (n 187) 149.

[193] Barocas and Selbst (n 167) 691.

[194] ibid 720.

[195] ibid 691.

[196] ibid 691–692.

[197] Richard B. Cohen, 'What Is Redundant Encoding And Should I Care?' (*Lexology*, 29 November 2016) <https://www.lexology.com/library/detail.aspx?g=d4f54bc5-704d-4ad3-9047-500493cdc41d> accessed 16 May 2022.

[198] Barocas and Selbst (n 167) 721–722.

[199] ibid 720.

proxy discrimination, that can be exploited,[200] as a creditor wishing to mask direct discrimination can do so by using proxy variables.[201]

Finally, factors that play a role during the selection of a ML model or in the decision-making stage of an ADM process can contribute to discrimination in algorithmic credit scoring. Firstly, a model consisting of assumptions, formulae, and parameter values is selected from several iterations of models based on criteria such as maximising prediction accuracy, minimising the error rate, etc.[202] This requires trade-offs, which can augment the likelihood of algorithmic discrimination. This is because a model that, for instance, promises to maximise profits and minimise losses for the creditor could be chosen over a model that strives for machine fairness[203] – the extent to which such a trade-off would have a different impact on different groups could become apparent only after the model's deployment.[204]

Secondly, a specific factor can play a role during the decision-making stage of algorithmic credit scoring, which, I believe, has not been addressed enough. The factors mentioned so far are mainly tied to data scientists' decisions, but creditors play a role in these decisions, namely how to define creditworthiness, what features to use, what to prioritize when choosing a model, and their records can also be used as raw data. In contrast, the following factor mainly concerns creditors' actions in the decision-making stage; this factor is a heuristic that Alon-Barkat and Busuioc refer to as 'selective adherence'.[205]

In contrast to automatic adherence to algorithmic advice ('automation bias') or 'undue deference to automated systems by human actors that disregard contradictory information from other sources or do not (thoroughly) search for additional information',[206] Alon-Barkat and Busuioc describe selective adherence as the selective acceptance of the outputs of a system, i.e. when they correspond to the decision-maker's pre-existing biases.[207] While the authors write about selective adherence in the context of algorithmic decision-support systems used in the

---

[200] ibid 692–693

[201] ibid.

[202] Bali and Sarkar (n 166) 192.

[203] ibid 236, stating that '[i]f we incorrectly predict a customer with a good credit rating as bad, we incorrectly deny him the credit loan but there is neither any profit or any loss', thus seemingly implying that unbiased outputs may be traded off for lack of loss.

[204] Michael Veale and Reuben Binns, 'Fairer Machine Learning in the Real World: Mitigating Discrimination Without Collecting Sensitive Data' (2017) 4(2) Big Data & Society 1, 3 <https://journals.sagepub.com/doi/abs/10.1177/2053951717743530> accessed 16 May 2022.

[205] Saar Alon-Barkat and Madalina Busuioc, 'Human-AI Interactions in Public Sector Decision-making: 'Automation Bias' and 'Selective Adherence' to Algorithmic Advice' (preprint, arXiv:2103.02381, 28 January 2022) 1 <https://arxiv.org/abs/2103.02381> accessed 16 May 2022.

[206] ibid 7.

[207] ibid 9.

public sector,[208] it is possible to imagine a case where a human reviewer would review automatically generated credit scores before the approval or denial of credit applications, and it could be the case that they would only challenge the credit scores of those who are not members of groups against which they are biased. Indeed, human oversight can have different manifestations,[209] including both oversight where the ML model's outputs do not become effective unless validated by a human and oversight in the form of ex post human intervention.[210] In any event, as Alon-Barkat and Busuioc point out, selective adherence has implications for 'the extent to which human decision-makers [in-the-loop] can actually function as effective decisional mediators and safeguard against [the risk for algorithmic bias].'[211] These words lead me to question whether the right to obtain human intervention under Article 22(3) of the General Data Protection Regulation (hereinafter: GDPR) can effectively safeguard individuals' right to non-discrimination, to which I will return in Chapter IV.

3.1.1 Focal point of the following legislative analysis

In light of the findings of this section, what must be answered is what mechanisms in legislation would work toward preventing or at least alleviating discrimination and what would work toward detecting it, and should therefore be considered when evaluating the extent to which relevant EU legislation can mitigate the risks to the respect for individuals' right to non-discrimination.

Starting with the first-mentioned type of mechanisms, given that the definition of the target variable and its associated class labels, construction of the training dataset, selection of features, and choice of a ML model can lead or contribute to algorithmic discrimination, providers of ML models would need to be required to prepare documentation alongside justifications or explanations and envisioned risks for discrimination detailing the process behind these tasks. This tool-centred mechanism would promote a more responsible approach to the performance of said tasks and, as providers can be external private entities, it would allow creditors to consider the envisioned risks and implement measures to mitigate them.

Next, since the imbalance of data in the training dataset poses a significant risk, there would also need to be a requirement for providers to ensure that the training data are sufficiently

---

[208] ibid 4.

[209] Commission, 'White Paper on Artificial Intelligence: A European approach to excellence and trust' (White Paper) COM (2020) 65 final, 21.

[210] ibid.

[211] Alon-Barkat and Busuioc (n 205) 10.

balanced. And lastly, given that human oversight during the decision-making stage also comes with a risk to the respect for individuals' right to non-discrimination, as it is possible that the human-in-the-loop is unaware of the possibility of selective adherence or automation bias and does not know how to react appropriately, creditors would need to be required to ensure that the human reviewer is aware of these heuristics.

Moving on to the second-mentioned type of mechanisms, since occurred algorithmic discrimination can be difficult to identify due to the black-box nature of many ML models, creditors would need to be required to rely on interpretable models. This implies a requirement for providers to build ML models that are sufficiently interpretable, and providers would also need to provide instructions on measures to facilitate human oversight. However, these tool-centred mechanisms would not work toward tackling intentional opacity, so the individual whose credit score has been calculated would also need to have a right to an explanation of how the ML model generated the credit score based on their algorithmic identity. Since this would serve no real purpose without the right to contest a discriminatory credit score, there would need to be another process-centred mechanism – a right to contest the credit score.

In analysing the EU legislative framework for algorithmic credit scoring, I will thus look for these mechanisms, which are set out in the following tables for the sake of readability. I now turn to the next section, where I will describe how algorithmic credit scoring affects individuals' private life and personal data and evaluate how that poses a risk to the respect for their rights to privacy and data protection.

Table 1: **Mechanisms for identifying algorithmic discrimination**

Actor	**Process-centred Mechanisms**	**Tool-centred Mechanisms**
Model Provider		Obligation to provide instructions on measures for facilitating human oversight; Model interpretability
Creditor		Obligation to rely on an interpretable model
Individual	Right to an explanation of how the model generated the credit score based on one's algorithmic identity	

Table 2: **Mechanisms for preventing/alleviating algorithmic discrimination**

Actor	**Process-centred Mechanisms**	**Tool-centred Mechanisms**
Model Provider		Obligation to prepare documentation alongside justifications/explanations and envisioned risks for algorithmic discrimination detailing the process of defining the target variable, class labels, data collection, data labelling, feature selection, and model selection; Balanced data
Creditor	Obligation to ensure that the human-in-the-loop is aware of the possibility of automation bias and selective adherence (in connection to human oversight)	
Individual	Right to contest the credit score	

3.2 Risks to individuals' rights to privacy and data protection

Algorithmic credit scoring can lead to intensive surveillance of individuals' behaviour, as it involves the use of large quantities of data about them, including financial but non-credit data and possibly non-financial, non-credit data,[212] giving creditors insight that goes beyond the limits of human observation. The algorithm first learns to link the input variables to the target variable based on the analysis of such data, all for discovering correlations that could more accurately determine an individual's creditworthiness in a particular case.[213] The result of this training is a model that the creditor can use to estimate the probability of default based on new data.[214] However, credit scores do not always accurately reflect applicants' ability to make credit repayments regularly and in full.[215] This is because the data about individuals that determine the credit scores, namely ML models' inferences based on the input variables, 'is not "made up," but, nor is it entirely real, in the sense of representing any reality—it is an extrapolation from fact and reality.'[216]

The first problem lies in the fact that ML algorithms operate on correlation and not causation. Although the two can correspond, this is not always the case, which undermines the legitimacy of credit scores based on such data.[217] Indeed, while the use of large amounts of diverse data could lead to a more accurate determination of creditworthiness, this increases the incidence of spurious or accidental correlations between creditworthiness and an input variable,[218] which occur when a third input variable enters and confounds the relationship between two other variables.[219] As was demonstrated in an incident where Dutch insurance companies charged extra for car insurance on the basis of the numbers indicating the apartments in which customers lived – presumably because the algorithm found a correlation between living in apartments with such numbers and the likelihood of being in a car accident,[220]

---

[212] Aggarwal (n 7).
[213] ibid.
[214] World Bank Group (n 3) 3.
[215] ibid.
[216] Mark Fenwick and Paulius Jurcys, 'From Cyborgs to Quantified Selves: Augmenting Privacy Rights with User-centric Technology and Design' (2022) 13(1) Journal of Intellectual Property, Information Technology and E-Commerce Law 20, 29.
[217] Emre Bayamlıoğlu and Ronald E. Leenes, 'The "Rule of Law" Implications of Data-driven Decision-making: A Techno-Regulatory Perspective' (2018) 10(2) Law, Innovation and Technology 295, 302; Kaminski (n 10) 1545; 1547.
[218] Hurley and Adebayo (n 187) 177; Jianqing Fan, Fang Han and Han Liu, 'Challenges of Big Data Analysis' (2014) 1(2) National Science Review 293, 298.
[219] Brian D. Haig, 'Spurious Correlation' in Neil J. Salkind (ed), *Encyclopedia of Measurement and Statistics* (online version, Sage Publications 2007) 2 <https://www.researchgate.net/publication/315829671_Spurious_correlation> accessed 20 May 2022.
[220] Zuiderveen Borgesius and Gerards (n 93) 16–17.

a significant amount of unrelated variables will inevitably be correlated when the dataset is sufficiently large.[221] Hence, less can be more when it comes to data.[222] In the words of Verhulst, "more data collection doesn't mean more knowledge. It actually means much more confusion, false positives and so on."[223]

The incident with the Dutch insurance companies also appears to reveal a deeper problem, lying in the fact that inaccurate credit scores can be the result of correlations with other people, who are (allegedly) similar, and by which individuals are clustered into groups (or classes) by a ML model.[224] As Mittelstadt notes, '[t]hese groups are imperfect reflections of the individuals contained within, constrained by the types of question they are designed to answer, and the flaws of the observables (i.e. the data) from which they are constructed.'[225] How such grouping of individuals or algorithmic classification[226] can lead to algorithmic discrimination was illustrated in the previous section. In the context of the right to privacy, algorithmic classification can have additional implications, as ML models' inferences made on the basis of group membership re-define individuals' identity as contained in the input variables.[227]

Despite this, individuals are normally not given access to these internal inferences, primarily due to trade secret protection related to ML models. As credit-scoring ML models' inferences do not generate and return data in the form of, for instance, personalised advertisements, the risk of denying access to algorithmic inferences here is arguably not that individuals would be unable to construct their internal identity freely; rather, this poses a risk to their control over their external identity or reputation, which, as the ECtHR held in *Pfeifer v. Austria*, 'forms part of [an individual's] personal identity and psychological integrity and therefore also falls within the scope of his or her "private life."'[228] Hence, denying individuals access to a ML model's inferences that re-define their algorithmic identity poses a risk to their autonomy as free identity-building and, in this way, the respect for their right to privacy.

Additionally, the data based on which ML models group individuals can be non-financial non-credit data. This potentially includes data about their 'network of individuals (such as friends, acquaintances, and coworkers) connected by interpersonal relationships',[229] which

---

[221] Fan, Han and Liu (n 218) 298–299.

[222] David Bollier, *The Promise and Peril of Big Data* (The Aspen Institute 2010) 14.

[223] ibid, quoting Stefaan Verhulst.

[224] Mittelstadt (n 154) 476; 478.

[225] ibid 479.

[226] ibid 477.

[227] ibid 483.

[228] *Pfeifer v Austria* App no 12556/03 (ECtHR, 15 November 2007), para 35.

[229] Merriam-Webster Dictionary, 'Social Network' (*Merriam-Webster*) <https://www.merriam-webster.com/dictionary/social%20network> accessed 20 May 2022.

creditors are motivated to use by the assumption that 'attractive prospects or customers are more likely to be connected to one another than to unattractive ones, and vice versa'.[230] Algorithmic classification based on such data has implications for associational privacy, which, as previously explained, is typified by individuals' interest in associating with whomever they choose without being monitored. If individuals' social network can affect their access to credit, it can be expected, in fact, that some will fear the influence of their social network over their credit score, which could dissuade them from associating with those they would not consider creditworthy. The risk of selectively forming social ties in order to improve credit scores is 'social fragmentation into sub-networks where people connect only to others who are very similar to them.'[231] Since this would most likely undermine individuals' free development of (their personality in their) relations with others, the use of data about individuals' social network in algorithmic credit scoring can be seen as an additional threat to the respect for the right to privacy.[232]

But the use of data about individuals' social network is not the only way algorithmic credit scoring can have a chilling effect, namely dissuading individuals from exercising their right to personal development.[233] This is because ML models can reveal even more about them than they even know about themselves,[234] be it through inferences made on the basis of the groups they have been clustered in within the system or (data about) their behaviour and actions. This has implications for their behavioural privacy, which, as previously explained, is characterised by their interest in remaining inconspicuous among the masses. If individuals' behaviour and habits, such as where they go shopping, what they post on their social media, and so on, can determine their creditworthiness, it can, in fact, be expected that this will dissuade some from engaging in any activity that could negatively affect their credit score. Indeed, as the inferences that could be drawn from alternative data are not as intuitive as those based on credit data (e.g. bad repayment performance suggests un-creditworthiness), individuals can struggle to ascertain what behaviour could lower their credit score[235] and adjust their behaviour without compromising their autonomy. Algorithmic credit scoring can thus be said to threaten

---

[230] Yanhao Wei and others, 'Credit Scoring with Social Network Data' (preprint, 2015) 35(2) Marketing Science 1, 3 <https://repository.upenn.edu/marketing_papers/383/> accessed 20 May 2022.

[231] ibid 32.

[232] See n 137.

[233] Registry of the European Court of Human Rights (n 121), para 74, stating that Article 8 ECHR protects 'the right to personal development'.

[234] Aggarwal (n 5) 59.

[235] ibid.

individuals' right to personal development by undermining their autonomy as free decision-making and, for this reason, pose a risk to the respect for their right to privacy.[236]

Finally, algorithmic credit scoring also poses a risk to the respect for individuals' right to data protection, not least due to ML models' inferences re-defining their algorithmic identity. This can best be illustrated with the following case from 2020, where the non-profit organisation 'NOYB' lodged a complaint under Articles 80(1) and 77(1) GDPR against the credit bureau 'CRIF' with the Austrian Data Protection Authority.[237] Credit bureaus are third-party public or private entities that collect (non)financial data on individuals and provide these data to creditors in an organised manner, usually by granting them access to their centralised database(s).[238] While a public centralised database (public credit registry) consists only of credit data reported by financial institutions under the supervision of a national central bank or supervisory authority,[239] a private credit bureau's central database(s) can consist of various data, including non-financial data.[240]

In the case at hand, NOYB lodged a complaint against the private credit bureau on behalf of an individual, whom I shall call 'Mr. X'. Mr. X was denied an energy supply contract due to his low credit score, which had been calculated by CRIF.[241] Confused by the low credit score, as he had not provided any data to CRIF nor was he aware that his credit score had been calculated, Mr. X decided to invoke his right of access and requested that CRIF provide him with information under Article 15 GDPR.[242] Upon his request, the credit rating agency informed him they had not kept any data on him.[243] As it thus appeared that his creditworthiness had been created out of nothing, Mr. X forwarded the received response to the energy supplier, who informed him that his credit score had been calculated based on his "person, name and date of birth".[244] In his correspondence with CRIF, Mr. X asked the credit rating agency to provide him with information on the basis of which he had been assigned a credit score of 446 out of 700 points, to which CRIF replied that no further information could be provided under

---

[236] Registry of the European Court of Human Rights (n 121), para 74.

[237] European Center for Digital Rights, 'Credit Scoring: Negative Credit Rating Generated Without Data' (*noyb*, 4 August 2020) <https://noyb.eu/en/credit-scoring-negative-credit-rating-generated-without-data> accessed 19 May 2022.

[238] Federico Ferretti, 'The Legal Framework of Consumer Credit Bureaus and Credit Scoring in the European Union: Pitfalls and Challenges—Overindebtedness, Responsible Lending, Market Integration, and Fundamental Rights' (2013) 46(3) Suffolk University Law Review 791, 798.

[239] ibid 800.

[240] ibid 799.

[241] Complaint from the European Center for Digital Rights (NOYB) (31 July 2020) 2 <https://noyb.eu/en/credit-scoring-negative-credit-rating-generated-without-data> accessed 21 May 2022.

[242] ibid 2–3.

[243] ibid 2.

[244] ibid 3.

Article 15 GDPR because "[the] parameters [on whose basis] the transmitted credit rating value is calculated in detail and which algorithm in the background evaluates and links the data available on your person logically and mathematically is to be qualified as a trade secret."[245] Mr. X, in spite of how his creditworthiness was deemed by CRIF, was at the time of the complaint debt-free, with a fixed salary, and had never encountered any financial difficulties.[246] As stated by NOYB, 'CRIF has thereby created data that is incorrect in content and does not correspond to the reality of life.'[247] The Austrian DPA later criticised CRIF for insufficiently disclosing that, in most cases, credit scores were calculated merely based on demographic data, such as sex, age, and residency data, and held that the credit rating agency must disclose to the concerned individuals the logic behind the credit scores – but not the concrete computer logic.[248]

This case shows how algorithmic credit scoring can threaten respect for individuals' right to data protection. Firstly, ML models' ability to reveal more about individuals than they know about themselves can conflict with the fair processing of their personal data,[249] as according to the principle of fairness, personal data cannot be processed in an unexpected or unjustifiably detrimental manner.[250] This means that the processing of personal data must correspond to individuals' expectations,[251] and it would be hard to claim that receiving a low credit score solely on the basis of one's person, name, and date of birth is in line with their expectations, which can also be said for algorithmic inferences based on neutral behaviour, such as shopping at particular stores. The use of alternative data to assess individuals' creditworthiness, especially data about their social network, can also be considered unjustifiably detrimental, as it can negatively affect their access to credit, even though these data may not be in any way related to their ability to make credit repayments regularly.[252]

Secondly, the right to data protection is also underpinned by principles enshrined in the GDPR that are not highlighted in Article 8 CFR, including the principle of data minimisation.

---

[245] ibid 4; 5.

[246] ibid 9.

[247] ibid.

[248] European Center for Digital Rights, 'Data Voodoo: Credit Ranking Agency CRIF Creates Credit Rating Out of Thin Air' (*noyb*, 4 August 2021) <https://noyb.eu/en/data-voodoo-credit-ranking-agency-crif-creates-credit-rating-out-thin-air> accessed 21 May 2022.

[249] CFR, art 8(2).

[250] European Data Protection Board, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (version 2.0, 20 October 2020), para 69.

[251] ibid.

[252] Information Commissioner's Office, 'Principle (a): Lawfulness, Fairness and Transparency' (*ico*) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/lawfulness-fairness-and-transparency/> accessed 24 May 2022, addressing controllers and stating that they must consider whether the use of personal data will negatively affect individuals in an unfair way.

In line with the data minimisation principle, only personal data that are 'adequate, relevant and limited to what is necessary'[253] for the purpose of processing can be processed.[254] This means that personal data processed for the purpose of creditworthiness assessment must not only be clearly related to the ability to repay a financial accommodation[255] but can only be processed if this purpose of processing cannot be achieved in any other way.[256] Data about individuals' social network can hardly be considered relevant or necessary for the assessment of their ability to reimburse credit.[257] Other types of alternative data can be considered relevant, but their necessity is questionable when applicants do not have a thin file. Nevertheless, such data can be irrelevant for the assessment of creditworthiness, as demonstrated in the above case where the data about Mr. X's 'person' – suggesting alternative data – were not related to his ability to reimburse credit.

Finally, the right to data protection also protects how personal data are handled in terms of ensuring their continued accuracy,[258] which is why individuals have the right of access to data that have been collected concerning them and the right to have that data rectified.[259] Access to personal data allows them to verify the accuracy thereof and rectify the data if they are inaccurate,[260] which seeks to prevent decisions from being made on the wrong basis.[261] As the CJEU acknowledged in *Nowak*, otherwise accurate collected personal data can lead to inaccurate inferences,[262] which justifies individuals' access to those inferences.[263] As demonstrated in the above case, the data that formed the basis for Mr. X's credit score were inaccurate, as they did not reflect his true creditworthiness. Although it is unclear whether the data were accurate at the time they were obtained, it is possible that they were but led to the ML model's inaccurate inferences. And yet, Mr. X was not given access to those inferences, which would have enabled him to understand what led to his low credit score. Due to the opacity of ML models' functioning, individuals thus may not be able to access algorithmic inferences and verify their accuracy, which could lead to credit-granting decisions on the wrong basis.

---

[253] GDPR, art 5(1)(c).
[254] ibid.
[255] European Data Protection Supervisor (n 29), para 15.
[256] European Data Protection Board (250), para 71.
[257] European Data Protection Supervisor (n 29), para 16.
[258] European Data Protection Board (250), para 79.
[259] CFR, art 8(2).
[260] European Data Protection Board (250), para 79.
[261] ibid 78.
[262] Case C-434/16 *Peter Nowak v Data Protection Commissioner* [2017] EU:C:2017:994, para 54.
[263] ibid [56].

3.2.1 Focal point of the following legislative analysis

Considering all the above, the question to be answered is what mechanisms in the relevant legislation would work toward tackling the risks of algorithmic credit scoring. Starting with the risk to individuals' control over their external identity, in order to be able to verify the accuracy of their algorithmic identity and prevent credit-granting decisions on the wrong basis, individuals would need to have a right of access to algorithmic inferences. This process-centred mechanism also implies the existence of two tool-centred mechanisms – providers would need to be required to develop a ML model enabling for the automatic recording of algorithmic inferences, and creditors would need to keep them for a period of time.

As to the risk to individuals' free personality-building and social relations stemming from the use of data about their social network, their use would need to be prohibited. The use of these data, however, concerns only one type of alternative data whose processing can contrast with the principles of fairness and data minimisation. Thus, creditors would also need to be required to justify the use of alternative data for the assessment of applicants with credit history, which would promote respect for individuals' right to data protection.

Finally, these mechanisms would also work toward reducing the intensity of surveillance of individuals' behaviour, which threatens their autonomy as free decision-making and thus their ability to reach their potential, and thus work toward preventing a private digital Panopticon, where individuals would feel as though they are constantly being monitored and fear how their every action could affect data concerning them.[264]

These mechanisms are set out in the following table, and I will now look for them as I will be evaluating the extent to which relevant EU legislation can mitigate the risks of algorithmic credit scoring.

---

[264] Bert-Jaap Koops and Ronald E. Leenes, '"Code" and the Slow Erosion of Privacy' (2005) 12(1) Michigan Telecommunications and Technology Law Review 115, 117–118.

Table 3: **Privacy and data protection mechanisms**

Actor	Process-centred Mechanisms	Tool-centred Mechanisms
Model Provider		Obligation to develop a model enabling for automatic recording of algorithmic inferences
Creditor	Prohibition on the use of data about individuals' social network; Obligation to justify the use of alternative data for the assessment of applicants with credit history	Time-limited obligation to keep automatically generated recordings of algorithmic inferences
Individual	Right of access to algorithmic inferences about oneself	

IV: EU legislative framework

The previous chapter illustrated how algorithmic credit scoring affects individuals' access to credit, their private life, and personal data, and set out the mechanisms that would work toward tackling the identified risks to respect for the rights to non-discrimination, privacy, and data protection. As algorithmic credit scoring plays a role in the conclusion of consumer credit agreements and involves the processing of personal data using ML algorithms, it triggers the application of various pieces of legislation. This practice is thus regulated in three contexts: consumer protection, data protection, and AI safety. More specifically, algorithmic credit scoring is currently regulated by the Consumer Credit Directive (hereinafter: 'CCD'),[265] which will be replaced by the Proposal for a Directive on consumer credits,[266] the GDPR, and will also be regulated by the proposed Artificial Intelligence Act (hereinafter: AI Act Proposal).[267] The following sections thus analyse said EU legislation by considering the set-out mechanisms and identify gaps in regard to ensuring respect for individuals' fundamental rights in algorithmic credit scoring.

---

[265] See n 21.
[266] See n 23.
[267] See n 24.

4.1 Consumer protection

While each EU consumer law defines the notion of 'consumer' for its purposes, those definitions converge[268] and normally refer to an individual not engaged in commercial or trade activities.[269] The EU consumer policy aims to protect individuals as consumers from serious risks that they cannot tackle alone and empower them to make choices based on accurate, clear, and consistent information.[270] The policy's specific objectives are defined in Article 169(1) of the Treaty on the Functioning of the European Union, according to which 'the Union shall contribute to protecting the health, safety and economic interests of consumers, as well as to promoting their right to information, education and to organise themselves in order to safeguard their interests.'[271] Consumer protection is recognised as an important EU goal in Article 38 CFR, according to which 'Union policies shall ensure a high level of consumer protection.'[272]

Financial services are one of the key areas of EU consumer policy, as the complexity of financial products poses particular risks to consumers' economic interests.[273] Among the legislation adopted to tackle the risks of financial services is the CCD,[274] which applies to agreements concerning consumer credit in the form of a deferred payment, loan, or other similar financial accommodation (exclusive of certain credit agreements[275]).[276] Article 8(1) CCD imposes an obligation on the creditor to assess the consumer's creditworthiness before the conclusion of a credit agreement, thus regulating credit scoring, which seeks to prevent irresponsible lending practices and over-indebtedness.[277] However, although responsible lending can lead to the denial of credit applications, this would not justify a disproportionate exclusion of consumers from the credit market.

This is because access to credit and, more broadly, financial inclusion is an important social objective enabling consumers to be active market players,[278] thus contributing to the competitiveness of the EU single market.[279] At the same time, access to credit and financial

---

[268] Jana Valant, *Consumer Protection in the EU* (Publications Office of the European Union 2015) 4.
[269] Case C–105/17 *Komisia za zashtita na potrebitelite v Evelina Kamenova and Okrazhna prokuratura – Varna* [2018] EU:C:2018:808, para 33.
[270] Valant (n 268) 3.
[271] Consolidated Version of the Treaty on the Functioning of the European Union [2016] OJ C202/1, art 169(1).
[272] CFR, art 38.
[273] Valant (n 268) 9.
[274] ibid 9–10.
[275] CCD, art 2(2).
[276] CCD, art 2(1); art 3(c).
[277] CCD, art 8(1); Proposal for a Directive on consumer credits, art 18(1).
[278] Iain Ramsay, 'Changing Policy Paradigms of EU Consumer Credit and Debt Regulation' in Dorota Leczykiewicz and Stephen Weatherill (eds), *The Images of the Consumer in EU Law* (Hart Publishing 2016) 162–163.
[279] ibid 180.

inclusion also play a role in helping consumers meet their current consumption needs and improve their social position,[280] although this is often seen as of secondary importance. As Davies points out,

> The [Court of Justice] implies, with its exclusion of the social, that [the consumer] should seek to separate her consumption from her membership of society and should see herself, when she is in the market as just a consumer. Her citizenship, indeed her humanity are matters for the agora or the home, but not for the market.[281]

While individuals thus act as consumers in consumer credit agreements, they are first and foremost human beings, and algorithmic credit scoring affects not only their credit score and access to credit but also their ability to fully participate in society or improve their standard of living. The social importance of credit is also why I have so far referred to consumers as individuals who are entitled to the rights to non-discrimination, privacy, and data protection, and it is with respect for these that the following analysis of consumer law is concerned.

4.1.1 Consumer Credit Directive

1. Extent of consumer protection

The CCD aims for a sufficient degree of consumer protection to ensure consumer confidence with a view to supporting the free movement of credit offers.[282] Consumer protection is ensured, inter alia, through the establishment of the creditor's obligation to provide the consumer with pre-contractual information[283] and the obligation to assess the consumer's creditworthiness.[284] Pre-contractual information covers the conditions, cost of the credit, and the consumer's obligations,[285] and includes information on the consumer's right to be informed immediately of the result of a database consultation carried out for the purposes of creditworthiness assessment.[286] This right concerns the consultation of a public database, which is only mandatory in the Member States whose own legislation requires creditors to

---

[280] Ramsay (n 278) 160; 162.

[281] Gareth Davies, 'The Consumer, the Citizen, and the Human Being' in Dorota Leczykiewicz and Stephen Weatherill (eds), *The Images of the Consumer in EU Law* (Hart Publishing 2016) 338.

[282] CCD, rec 8.

[283] CCD, art 5; art 6.

[284] CCD, art 8.

[285] CCD, rec 19.

[286] CCD, art 5(1)(q).

assess consumers' creditworthiness based on a consultation of the public credit registry.[287] Thus, the creditor must assess the consumer's creditworthiness based on 'sufficient information, where appropriate obtained from the consumer and, *where necessary*, on the basis of a consultation of the relevant database.'[288] Private databases are consulted on a voluntary basis,[289] and such consultations fall outside the scope of this provision.

2. Gaps in the protection of fundamental rights

As can be observed from the language of the above provision, the CCD does not address credit scoring using ML algorithms.[290] For this reason, the Directive can be said to regulate only the non-automated aspects of the process of algorithmic credit scoring, namely the creditor's obligations to assess the consumer's creditworthiness based on 'sufficient information' and consult the central credit register where necessary.

Therefore, the CCD does not contain mechanisms for mitigating the risks of algorithmic credit scoring. Given the Directive's subject matter, it can be said that there is firstly a lack of a prohibition on the use of data about individuals' social network and of an obligation to justify the use of alternative data for the assessment of applicants with credit history. The lack of these mechanisms threatens respect for consumers' rights to privacy and data protection in algorithmic credit scoring, as it leaves the risks to their autonomy and social relations and the processing of their personal data in line with the principles of fairness and data minimisation unaddressed. The CCD also lacks a consumer's right to an explanation of how the ML model generated the credit score based on their algorithmic identity, which poses a risk to the respect for consumers' right to non-discrimination, as they will hardly be able to identify a discriminatory credit score. Consequently, this undermines their ability to effectively contest the discriminatory credit score – however, the right to do so is also not provided for in the Directive.

All in all, it can be said that the current CCD does not mitigate the risks of algorithmic credit scoring that could be addressed through mechanisms within the subject matter of the Directive. With this in mind, I now turn to the analysis of the Proposal for a new CCD.

---

[287] Ferretti (n 238) 804.

[288] CCD, art 8(1) (emphasis added).

[289] Ferretti (n 238) 799.

[290] Proposal for a Directive on consumer credits, Explanatory Memorandum, 1, mentioning 'automated decision-making systems' and 'non-traditional data'.

4.1.2 Proposal for a Directive on consumer credits

1. Extent of consumer protection

On 30 June 2021, the European Commission adopted a Proposal for a Directive on consumer credits (hereinafter: 'Proposal'), which aims to adapt the rules on consumer credit to the changes brought about by the rapid technological developments that have taken place in the last few years.[291] The Proposal claims to uphold the respect for fundamental rights as enshrined in the CFR, particularly the rights to data protection, non-discrimination, and consumer protection.[292] With a view to ensuring a high and equivalent level of consumer protection in the Union and creating a well-functioning internal market, the Proposal introduces several measures.[293] These include the extension of the types of credit agreements regulated by the CCD,[294] the addition of a provision prohibiting discrimination on any ground referred to in Article 21 CFR with respect to the conditions to be fulfilled for being granted a credit when requesting, concluding, or holding a credit agreement,[295] and a modernised regime for assessing creditworthiness.[296]

This regime is set forth by Article 18 of the Proposal. According to Article 18(2) of the Proposal, creditworthiness shall be assessed based on

> *relevant* and accurate information on the consumer's income and expenses and other financial and economic circumstances which is *necessary* and proportionate such as evidence of income or other sources of repayment, information on financial assets and liabilities, or information on other financial commitments.[297]

Article 18(6) of the Proposal further states that where the creditworthiness assessment involves the use of automated processing of personal data, the consumer shall have the right to request and obtain human intervention on the part of the creditor to review the decision, request and obtain a clear explanation of the creditworthiness assessment, including on the logic and risks involved, and express their point of view and contest the creditworthiness assessment and the decision.

---

[291] Proposal for a Directive on consumer credits, Explanatory Memorandum, 3; rec 4.
[292] Proposal for a Directive on consumer credits, rec 25.
[293] Proposal for a Directive on consumer credits, rec 13.
[294] Proposal for a Directive on consumer credits, art 2.
[295] Proposal for a Directive on consumer credits, art 6.
[296] Proposal for a Directive on consumer credits, art 18.
[297] Proposal for a Directive on consumer credits, art 18(2) (emphasis added).

Finally, Article 18(7) provides that 'where the credit application is rejected the creditor (…) is required to inform the consumer without delay (…) of the fact that the assessment of creditworthiness is based on automated processing of data.' Hence, the Proposal regulates the process of algorithmic credit scoring.

2. Gaps in the protection of fundamental rights

Except for providing examples of information that could be relevant, necessary, and proportionate, Article 18(2) of the Proposal does not specify the types of data that can be used when conducting a creditworthiness assessment. In this respect, recital 47 to the Proposal does provide more clarity by stating that social media data or health data should not be used and by referring to the European Banking Authority's Guidelines on loan origination and monitoring (EBA/GL/2020/06) as a guide that could assist creditors in determining the types of data they should avoid using. But as the European Data Protection Supervisor points out, the fact that the operative part of the Proposal does not set any limitations on the types of data that can be used 'entails significant risks of excessive and unfair data processing.'[298] Nor can an overview of the usable data be found in the recitals, which, while mentioning two types of data that should not be used, refer to an external source as potential guidance regarding the use of other data.

A reference to social media data in the recitals instead of a prohibition on their use in the operative part of the Proposal does not mitigate the risks that the use of data about individuals' social network poses to the respect for their rights to privacy and data protection. Moreover, while Article 18(2) of the Proposal refers to the principle of data minimisation, the Proposal does not impose an obligation on the creditor to justify the use of alternative data for the assessment of applicants with credit history. The risk to the respect for consumers' right to data protection thus seemingly remains, as creditors could use certain types of data for the assessment of the creditworthiness of any applicant, even though the data that can be considered relevant and necessary can vary depending on an applicant's credit history or lack thereof.[299]

Next, while the Proposal strives to enhance consumer protection by introducing the consumer's rights under Article 18(6), it simultaneously appears to hamper their effectiveness.

---

[298] European Data Protection Supervisor (n 29), para 7.

[299] Information Commissioner's Office, 'Principle (c): Data Minimisation' (*ico*) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/data-minimisation/> accessed 30 May 2022.

This is because the right to an explanation is presented as a right to a ‘meaningful explanation of the assessment made and of the functioning of the automated processing used, including among others the main variables, the logic and risks involved’.[300] This suggests that the right is actually the ‘right to explanation’ as contained in the GDPR, which, as I will argue in the next section, does not entail a right to an explanation of how the ML model generated the credit score based on one’s algorithmic identity. This is also reflected in how the Proposal refers to the ‘main variables’ instead of referring to the ML model’s inferences.

The scope of the right to explanation also has implications for consumers’ ability to effectively contest their credit score, as a consumer might, for example, accept how their transaction history, used as an input variable, influenced their credit score as a correct reflection of their creditworthiness when the credit score would actually reflect creditworthiness by association and not their own. Moreover, it is also questionable to what extent the right to obtain human intervention, as presented in the Proposal, can serve as an effective safeguard against the risk for discrimination, as human oversight could lead to the human-in-the-loop rubber-stamping the credit scores (automation bias)[301] or selectively accepting them (selective adherence), which is left unaddressed.

Finally, according to Articles 13(2)(f) and 14(2)(g) GDPR, the creditor must provide the information under Article 18(7) of the Proposal irrespective of the outcome of the credit application and is also required to provide the consumer with meaningful information about the logic involved and the significance and envisioned consequences of such processing at the time when the consumer provides their personal data or, if personal data have not been obtained from the consumer, within the time limits set forth in Article 14(3) GDPR.[302] The Proposal thus introduces a degree of unclarity, which contributes to legal uncertainty regarding creditors’ obligations.[303] Given the power asymmetry between creditors and consumers, the creditor’s obligations in this respect should be clear, which would help to ensure that the consumer is provided with meaningful information.[304]

All this considered, it can be concluded that the protection against the risks of algorithmic credit scoring offered by the Proposal does not sufficiently ensure respect for individuals’ rights to non-discrimination, privacy, and data protection.

---

[300] Proposal for a Directive on consumer credits, rec 48.

[301] European Data Protection Supervisor (n 29), para 28.

[302] GDPR, art 13(2)(f); art 14(2)(g); (3).

[303] European Data Protection Supervisor (n 29), para 32.

[304] ibid para 10.

4.2 General Data Protection Regulation

1. Extent of data protection

The processing of consumers' personal data by automated means or the processing by other means of personal data forming part of a filing system or intended to form part of such a system is subject to the rules of the GDPR,[305] including the already mentioned principles relating to the processing of personal data and the data subject's right of access to personal data. While the GDPR as a whole thus regulates the processing of consumers' personal data for the purpose of creditworthiness assessment, Article 22 GDPR establishes a special regime for the process of algorithmic credit scoring, starting by setting forth a general prohibition on automated individual decision-making[306] under three cumulative conditions.[307] First, there is no human involvement in the decision process, which refers to the absence of human oversight of the decision by someone with the authority and competence to change the decision.[308] Second, the decision concerning the data subject produces legal effects or similarly significant effects, meaning it affects the rights or legal status of the data subject, or the effects are 'sufficiently great or important to be worthy of attention'.[309] Lastly, none of the exceptions listed in Article 22(2) GDPR applies.[310]

Since decisions affecting consumers' financial circumstances, such as their eligibility for credit, are considered sufficiently significant to meet the threshold of Article 22(1) GDPR,[311] fully automated credit scoring is, at least in theory, prohibited. However, the absence of a determination of the necessary level of human involvement[312] and the broad exception contained in Article 22(2)(a) GDPR mean in practice that creditors can circumvent the prohibition. Such decision-making, in fact, can be considered necessary for entering into a consumer credit agreement if human involvement would be impractical due to the amount of data being processed.[313] In such cases, the creditor as the data controller 'shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her

---

[305] GDPR, art 2(1).

[306] Article 29 Data Protection Working Party (n 22).

[307] GDPR, art 22(1)–(2).

[308] Article 29 Data Protection Working Party (n 22) 20–21.

[309] GDPR, art 22(1); Article 29 Data Protection Working Party (n 22) 21.

[310] GDPR, art 22(2).

[311] Article 29 Data Protection Working Party (n 22) 22.

[312] Reuben Binns and Michael Veale, 'Is That Your Final Decision? Multi-stage Profiling, Selective Effects, and Article 22 of the GDPR' (2021) 11(4) International Data Privacy Law 319, 332.

[313] Article 29 Data Protection Working Party (n 22) 23.

point of view and to contest the decision.'[314] Recital 71 to the GDPR explains that these safeguards should include 'specific information to the data subject and the right to obtain (…) an explanation of the decision reached after such assessment'.[315]

2. Comparison with the Proposal for a Directive on consumer credits

As can be observed, the consumer's rights to obtain human intervention and contest the credit score, as introduced in the Proposal, derive from Article 22 GDPR. By contrast, Article 22 GDPR does not contain a 'right to explanation', or at least it is not referred to as such, which is significant from the viewpoint of the explanation that the consumer is entitled to obtain under the GDPR and the Proposal.

Scholars such as Selbst and Powles have argued that what recital 71 to the GDPR is referring to is the controller's obligation to provide 'meaningful information about the logic involved [in the process referred to in Article 22(1)]'[316] as per Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR.[317] Following Article 29 Working Party's recommendations in regard to providing meaningful information under Articles 13–15 GDPR, this limits the explanation to information such as 'how any profile (…) is built',[318] 'why this profile is relevant to the automated decision-making process',[319] and 'how it is used for a decision concerning the data subject'.[320]

Conversely, Article 18(6)(b) of the Proposal introduces a right to 'obtain (…) a clear explanation of the assessment of creditworthiness, *including* on the logic and risks involved',[321] thus suggesting that the right to explanation encompasses both an information and access right under Articles 13–15 GDPR and a right to an explanation of how a particular credit score was reached. This is also made clear in recital 48 to the Proposal, which defines it as a right to a 'meaningful explanation of the assessment made *and* of the functioning of the automated processing'.[322] Article 18(6)(b) of the Proposal also states that the right to explanation allows the consumer to obtain an explanation of the 'significance and effects [of the automated processing of personal data] on the *decision*'[323] – and not only of the significance and envisaged

---

[314] GDPR, art 22(3).
[315] GDPR, rec 71.
[316] GDPR, art 13(2)(f); art 14(2)(g); art 15(1)(h).
[317] Andrew D. Selbst and Julia Powles, 'Meaningful Information and the Right to Explanation' (2017) 7(4) International Data Privacy Law 233, 235.
[318] Article 29 Data Protection Working Party (n 22) 31.
[319] ibid.
[320] ibid.
[321] Proposal for a Directive on consumer credits, art 18(6)(b) (emphasis added).
[322] Proposal for a Directive on consumer credits, rec 48.
[323] Proposal for a Directive on consumer credits, art 18(6)(b) (emphasis added).

effects[324] on a decision that has not yet been taken. Hence, the Proposal seems to establish a right to a 'subject-based explanation'[325] instead of a right to a general explanation of the ML model's functioning or 'model-based explanation',[326] which is more compatible with the operative part of the GDPR.

3. Gaps in the protection of fundamental rights

The GDPR does not address the possibility of automation bias and selective adherence, so the effectiveness of the right to obtain human intervention remains questionable, and while the Proposal seems to give teeth to the right to a subject-based explanation, the right may fall short of mitigating the risks of algorithmic credit scoring. This is because recital 48 to the Proposal refers to the 'main variables' and not to the ML model's inferences and their impact on the credit score. While the Proposal does not exclude a subject-based explanation based on the ML model's inferences,[327] the legislator has omitted a reference to any element of the algorithmic credit scoring process that would support a more impactful interpretation of the right to explanation, namely as a right to an explanation of how the ML model generated the credit score based on one's algorithmic identity.

Given that recital 63 to the GDPR states that 'the right to know (…) the logic involved in any automatic personal data processing'[328] 'should not adversely affect the rights or freedoms of others, including trade secrets',[329] this could be seen as an effort to strike a balance between the interests of consumers and creditors. Indeed, creditors can argue that they have a legitimate interest in maintaining the confidentiality of such information and a legitimate expectation that such confidentially will be preserved, and that disclosure of the ML model's inferences would likely harm their interests.[330] This is because transparency towards individuals in the context of algorithmic credit scoring carries the risk for attempts of system manipulation.

---

[324] GDPR, art 13(2)(f); art 14(2)(g); art 15(1)(h).

[325] Lilian Edwards and Michael Veale, 'Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?' (arXiv:1803.07540, 2 July 2018) 1, 4 <https://arxiv.org/abs/1803.07540v2> accessed 1 June 2022.

[326] ibid.

[327] Proposal for a Directive on consumer credits, rec 48, stating 'among others'.

[328] GDPR, rec 63.

[329] ibid.

[330] Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure [2016] OJ L157/1, rec 14.

Kear explains that individuals often see gaming the system simply as taking control of their accurate representation[331] by '[telling] their creditworthiness-stories in ways that are legible to algorithms'.[332] As ML models generate credit scores based on individuals' algorithmic identity, which often misaligns with their perception of their creditworthiness,[333] they thus try to use information about the scoring logic to adapt their behaviour.[334] Since individuals' ability to control how creditworthy they appear means a higher risk for creditors' losses, as less creditworthy applicants could also adapt their behaviour to obtain a higher credit score, creditors can thus rely on trade secret protection to restrict individuals' access to information about ML models' logic.

While recital 63 to the GDPR does further state that 'the result of those considerations should not be a refusal to provide *all* information to the data subject',[335] such a refusal, according to the Proposal, would not happen if the consumer were provided with at least an explanation of how the choice of the *main* variables affected their credit score (e.g. 'our model uses credit history as one of the criteria and, due to your insufficient credit history, you scored poorly'). In light of this conclusion, it can also be said that while Article 15(1) GDPR provides for the right of access to personal data, including inferences in the form of 'opinions and assessments',[336] it does not guarantee access to algorithmic inferences.

Indeed, Article 15(4) and recital 63 to the GDPR state that the right of access 'shall not adversely affect the rights and freedoms of others.'[337] This is also emphasised in the European Data Protection Board's Guidelines 01/2022, which, while mentioning 'algorithmic results' as an example of inferred data to be provided under Article 15 GDPR,[338] stress that access to such data can be limited if it would adversely affect the economic interests of a private entity safeguarded through trade secret protection.[339] In the absence of the GDPR indicating what degree of disclosure of algorithmic inferences would still be compatible with trade secret protection while promoting respect for individuals' fundamental rights, it is thus hard to

---

[331] Mark Kear, 'Playing the Credit Score Game: Algorithms, "Positive" Data and the Personification of Financial Objects' (2017) 46(3-4) Economy and Society 1, 10 <https://www.tandfonline.com/doi/full/10.1080/03085147.2017.1412642?scroll=top&needAccess=true> accessed 7 June 2022.

[332] ibid 14.

[333] ibid 11.

[334] ibid 13–14.

[335] GDPR, rec 63 (emphasis added).

[336] *Nowak* (n 262) para 34; para 56.

[337] GDPR, art 15(4); rec 63.

[338] European Data Protection Board, 'Guidelines 01/2022 on data subject rights – Right of access' (version 1.0, 18 January 2022), para 96.

[339] ibid para 168.

envision creditors providing consumers access to algorithmic inferences following subject access requests.

It can thus be concluded that the protection afforded by the GDPR, complemented by the Proposal for a Directive on consumer credits regarding the right to explanation, does not sufficiently ensure respect for individuals' rights to non-discrimination, privacy, and data protection in algorithmic credit scoring. Neither the GDPR nor the Proposal, in fact, refer to an element of the ADM process in relation to the right to explanation that would support a more impactful interpretation of the right, which consequently also affects individuals' ability to effectively contest their credit score. The GDPR also lacks an indication of the extent of access to algorithmic inferences that would be compatible with trade secret protection while promoting respect for fundamental rights, which could prevent creditors' over-reliance on trade secret protection to avoid allowing individuals to access their personal data.

Given the broad exception under Article 22(2)(a) GDPR allowing creditors to fully automate credit scoring, how the safeguards to mitigate the risks of such decision-making are hardly effective, and the lack of support for a more effective right of access, the GDPR thus insufficiently mitigates the risks of algorithmic credit scoring. I now turn to the last section of this chapter, in which I will analyse the AI Act Proposal.

4.3 Artificial Intelligence Act Proposal

1. Introduction to the Proposal

On 21 April 2021, the European Commission presented the AI Act Proposal, which claims to harmonise the rules for the development, placement on the market, and use of AI systems,[340] namely software developed with techniques and approaches such as ML, with the ability to generate outputs such as content, predictions, recommendations, or decisions.[341] Despite this claim, the AI Act Proposal primarily focuses on the development of AI systems and contains almost no provisions concerning their use, which could be a consequence of the Proposal mostly building on the New Legislative Framework (NLF) regulating the safety of ICT products.[342] The latter, in fact, forms the basis for the Proposal's most extensive part (Part

---

[340] AI Act Proposal, Explanatory Memorandum, p 3; art 1.

[341] AI Act Proposal, art 3(1); Annex I, point (a).

[342] DIGITALEUROPE, 'DIGITALEUROPE's initial findings on the proposed AI Act' (6 August 2021) 2 <https://www.digitaleurope.org/resources/digitaleuropes-initial-findings-on-the-proposed-ai-act/> accessed 2 June 2022.

III),[343] which establishes a special regime for AI systems posing a 'high risk to the health and safety or fundamental rights of natural persons'.[344] These include AI systems intended to be used for the assessment of consumers' creditworthiness,[345] and so the AI Act Proposal regulates ML algorithms as the tool for algorithmic credit scoring.

The Proposal refers to the body that develops an AI system or has it developed to place it on the market or put it into service under its name or trademark as the 'provider',[346] while the body using it under its authority is the 'user'.[347] To 'place it on the market' means to first make an AI system available on the Union market,[348] whereas to 'put it into service' refers to the supply of an AI system for first use directly to the user or for own use on the Union market for its intended purpose, namely the use for which the AI system is intended by the provider and specified in its instructions for use.[349]

In the context of algorithmic credit scoring, the creditor is thus considered the user; however, if the creditor were to substantially modify the ML model, i.e. to make changes that have not been pre-determined by the provider in relation to the model's performance as it continues learning,[350] the creditor would be considered the provider[351] instead of the initial provider.[352]

2. Extent of ensured AI safety

The majority of obligations in relation to the requirements for credit-scoring ML models are borne by providers. Starting with obligations regarding the design and development of ML models, the provider must design and develop the ML model in such a way as to enable the automatic recording of events or 'logs' during its operation[353] and ensure that its functioning is sufficiently transparent to allow for the interpretation of the outputs and human oversight,[354] including through the development of human-machine interface tools.[355]

The AI Act Proposal defines human oversight as a measure aimed at 'preventing or minimising the risks to [individuals'] health, safety or fundamental rights'[356] by 'minimising

---

[343] AI Act Proposal, Explanatory Memorandum, p 4.
[344] AI Act Proposal, Explanatory Memorandum, p 13.
[345] AI Act Proposal, art 6(2); Annex III, point 5(b).
[346] AI Act Proposal, art 3(2).
[347] AI Act Proposal, art 3(4).
[348] AI Act Proposal, art 3(9).
[349] AI Act Proposal, art 3(11)–(12).
[350] AI Act Proposal, art 43(4).
[351] AI Act Proposal, art 28(1)(c).
[352] AI Act Proposal, art 28(2).
[353] AI Act Proposal, art 16(a); art 12.
[354] AI Act Proposal, art 16(a); art 13; art 14.
[355] AI Act Proposal, art 14(1).
[356] AI Act Proposal, art 14(2).

the risk of erroneous or biased AI-assisted decisions'.[357] As per Article 14(3) of the Proposal, this shall be achieved through measures enabling the human-in-the-loop (HITL) to 'fully understand the capacities and limitations'[358] of the ML model and to 'remain aware of (…) ('automation bias')'.[359] These can be identified and built into the ML model, or they can be identified before the ML model is placed on the market or put into service by the provider, and their implementation is left to the user.[360]

Providers also have obligations in relation to the data used for the development of ML models. With a view to ensuring that the training, validation, and testing datasets are sufficiently relevant, representative, and free of errors in view of the ML model's intended purpose,[361] the provider must thus put in place a data quality management system comprising procedures for each data operation.[362]

Moving on to documentation-related obligations, the provider must specify in the instructions for use accompanying the ML model any known or foreseeable circumstances that may affect the expected level of the ML model's accuracy and any known or foreseeable circumstances that may lead to risks to fundamental rights.[363] The instructions for use must also include information on human oversight measures, such as technical measures to facilitate the interpretation of the outputs.[364]

In addition, the provider must prepare technical documentation for the ML model[365] containing information, inter alia, on its design[366] and datasheets describing the training methodologies, techniques, and datasets.[367] The design specifications comprise choices regarding the definition of the target variable, class labels, features to be used for ML,[368] and the choice of the model,[369] whereas the datasheets include information on the processes of data collection and data labelling.[370]

---

[357] AI Act Proposal, Explanatory Memorandum, p 11.
[358] AI Act Proposal, art 14(4)(a).
[359] AI Act Proposal, art 14(4)(b).
[360] AI Act Proposal, art 16(a); art 14(3).
[361] AI Act Proposal, art 16(a); art 10(1); (3); rec 44.
[362] AI Act Proposal, art 17(1)(f); rec 44.
[363] AI Act Proposal, art 16(a); art 13(2)–(3)(b).
[364] AI Act Proposal, art 13(3)(d).
[365] AI Act Proposal, art 18(1).
[366] AI Act Proposal, Annex IV, point 2(b).
[367] AI Act Proposal, Annex IV, point 2(d).
[368] AI Act Proposal, Annex IV, point 2(b), mentioning 'key design choices including the rationale and assumptions made' and 'main classification choices'.
[369] AI Act Proposal, Annex IV, point 2(b), mentioning 'decisions about any possible trade-off'; point 2(g), mentioning 'the validation and testing procedures used'.
[370] AI Act Proposal, Annex IV, point 2(d).

The electronic instructions for use and a scanned copy of the EU technical documentation assessment certificate with the conclusions of the examination of the technical documentation performed by the competent authority[371] under Article 97 of the Directive 2013/36/EU[372] must also be entered into the EU database for stand-alone high-risk AI systems by the provider,[373] which aims to promote public-facing accountability.[374]

In comparison with the ones for providers, the obligations that are borne by creditors as users are extremely limited. These include, for instance, a time-limited obligation for users to keep the automatically generated logs to the extent they are under their control[375] and an obligation to use the ML model in accordance with the instructions of use.[376] The latter also serve as the basis for the user's monitoring of the ML model's functioning.[377]

3. Gaps in the protection of fundamental rights

Although the AI Act Proposal contains almost all of the herein identified tool-centred mechanisms that would work toward tackling the risks of algorithmic credit scoring, it lacks an obligation for the user to rely on an interpretable model or to alternatively use techniques such as model-agnostic methods to explain and understand the operation of a black-box model.[378] More specifically, the AI Act Proposal does not establish a regime for the interpretability of existing ML models that will not undergo significant changes in their design or intended purpose. If the ML models already in use by creditors will thus not undergo significant changes, they will be exempted from the rules of the AI Act[379] and so will not necessarily be developed in accordance with the transparency requirements under Article 13(1). In fact, it has been shown that organisations using AI are generally not '"actively addressing" the risk associated with explainability'.[380] The AI Act Proposal thus does not

---

[371] AI Act Proposal, Annex VIII, point 8; point 10; Annex VII, point 4.3.; point 4.6.

[372] AI Act Proposal, art 43(2).

[373] AI Act Proposal, art 60(2).

[374] Michael Veale and Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act—Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach' (2021) 22(4) Computer Law Review International 97, 112.

[375] AI Act Proposal, art 29(5).

[376] AI Act Proposal, art 29(1).

[377] AI Act Proposal, art 29(4).

[378] Christoph Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* (2nd edn, 2022) <https://christophm.github.io/interpretable-ml-book> accessed 2 June 2022.

[379] AI Act Proposal, art 83(2).

[380] Paul B. de Laat, 'Algorithmic Decision-making Employing Profiling: Will Trade Secrecy Protection Render the Right to Explanation Toothless?' (2022) 24(2) Ethics and Information Technology 1, 3 <https://link.springer.com/article/10.1007/s10676-022-09642-1> accessed 8 June 2022.

eliminate the risk to the respect for individuals' right to non-discrimination arising from ML models' intrinsic opacity, which undermines the identification of algorithmic discrimination.

A similar mismatch between the objective of ensuring respect for individuals' fundamental rights and the obligations under the AI Act Proposal also arises in relation to the requirements for human oversight measures. If their implementation is left to the user, the appointment of a person who truly understands the ML model's limitations and is aware of the possibility of automation bias, in fact, depends on the provider's instructions. To the extent that the instruction manual does not require the user to appoint a HITL with the necessary knowledge to prevent them from blindly following the outputs, the effectiveness of human oversight, and thus the minimisation of the risk for erroneous or biased decisions, is up to the user's commitment to mitigating the risks to individuals' fundamental rights. The Proposal also fails to acknowledge the possible tendency of selectively relying on the AI system's output. Since the Proposal for a Directive on consumer credits and the GDPR do not address the risk for the HITL rubber-stamping the credit scores or selectively accepting them, the effectiveness of human intervention as a safeguard against the risks of algorithmic credit scoring thus remains questionable.

Finally, in imposing obligations on the provider concerning the design and development of and documentation for ML models, the AI Act Proposal does not address the fact that ML engineers and data scientists lack a legal background. As previously explained, in striving for machine fairness, ML engineers and data scientists limit themselves to applying a set of selected statistical fairness criteria, which may not be the same as the criteria that would be chosen by other stakeholders, such as regulators and the public. In identifying circumstances that may lead to risks to fundamental rights, the hazards identified by ML engineers and data scientists may also fail to encompass all the hazards that would be pinpointed by legal and ethical experts, non-governmental organisations, and similar actors. For this reason, it is essential that experts or civil society be involved in the preparation of data and the development of ML models.

The need to recognise their role in the construction, training, and deployment of ML algorithms can also be viewed in the broader light of what Kaminski refers to as a system of collaborative governance wherein 'private-public partnerships [are deployed] towards public governance goals'[381] and which is built on public-facing and expert-facing accountability.[382]

---

[381] Kaminski (n 10) 1559.

[382] ibid 1563; 1607.

As Kaminski notes, the need for a system of collaborative governance stems from the fact that individuals have a limited technical, legal, or economic capacity to uncover discrimination in an ADM process[383] and may not invoke their rights.[384] Such a system is thus all the more necessary in the context of algorithmic credit scoring, which affects individuals' ability to fully participate in society or improve their standard of living.

Data governance or 'defining, applying and monitoring the patterns of rules and authorities for directing the proper functioning of, and ensuring the accountability for, the entire life-cycle of data and algorithms'[385] can be seen as an aspect of collaborative governance. The Proposal's provisions concerning data requirements, the obligation to put in place a data quality management system, the conformity assessment procedure, and the EU database for stand-alone high-risk AI systems are among those together establishing a data governance framework. This framework, however, significantly excludes external experts and civil society, which carries the risk for threats to individuals' fundamental rights not being adequately identified and acted upon.

Even the EU database for high-risk AI systems, aimed precisely at promoting public-facing accountability, may not effectively contribute to the public's ability to identify threats to individuals' fundamental rights. The AI Act Proposal, in fact, does not seem to envision it including information on the risks revealed after the implementation of AI systems,[386] which would significantly contribute to the public's ability to uncover AI systems that do not comply with the requirements laid down in the Proposal.

Furthermore, the enforcement of the AI Act Proposal's rules on the basis of the database could be significantly hampered due to the lack of a complaint mechanism.[387] The Proposal, in fact, contains no measures that would directly help affected individuals,[388] namely mechanisms to lodge a complaint against the user for non-compliance with the rules of the AI Act Proposal or seek a judicial remedy.[389] And although the Proposal states that 'effective redress for affected persons will be made possible by ensuring transparency and traceability of the AI systems',[390] it does not ensure transparency with respect to individuals, but only in relations

---

[383] ibid 1558–1559.
[384] ibid 1581.
[385] Marijn Janssen and others, 'Data Governance: Organizing Data for Trustworthy Artificial Intelligence' (2020) 37(3) Government Information Quarterly 1, 2 <https://www.sciencedirect.com/science/article/abs/pii/S0740624X20302719> accessed 9 June 2022.
[386] Article 60(3) refers to data listed in Annex VIII, which does not include post-market monitoring reports.
[387] Veale and Zuiderveen Borgesius (n 374).
[388] ibid 111.
[389] ibid.
[390] AI Act Proposal, Explanatory Memorandum, p 11.

between providers and users. Ensuring transparency in relations between creditors and consumers is thus left to consumer law and the GDPR, where transparency vis-à-vis individuals is hampered by the limited scope of the right to explanation and the right of access to personal data.

Considering the lack of a regime for the interpretability of existing ML models that will not undergo significant changes, an obligation on the user to ensure that the HITL is aware of the possibility of automation bias and selective adherence, and the minimal role of external experts and civil society in the construction, training, and deployment of ML algorithms, the protection against the risks of algorithmic credit scoring afforded by the AI Act Proposal does not sufficiently ensure respect for individuals' fundamental rights.

V: Process- and tool-centred solutions for ensuring respect for individuals' fundamental rights

The previous chapters identified the risks of algorithmic credit scoring to respect for individuals' rights to non-discrimination, privacy, and data protection, and presented the EU legislative framework for the practice. As the legislative analysis showed, there are still significant gaps in the currently applicable CCD and the GDPR in terms of ensuring respect for individuals' fundamental rights, as the CCD is based on traditional credit scoring and the GDPR most notably lacks strong safeguards in the case of automated individual decision-making. Recent EU legislative proposals do not appear to address these shortcomings sufficiently, especially given the absence of an alternative data regime in the Proposal for a Directive on consumer credits and the minimal role of external experts and civil society in the construction, training, and deployment of ML algorithms envisioned by the AI Act Proposal. The following solutions thus build on the currently applicable or proposed rules for the process of, or tool for, algorithmic credit scoring and, if implemented, would provide for a higher level of protection of individuals.

5.1 Collaborative data governance

The AI Act Proposal does not address the fact that ML engineers and data scientists lack a legal background and thus the circumstances that may lead to risks to fundamental rights identified by them may fail to encompass all the hazards that would be pinpointed by legal and ethical experts, non-governmental organisations, and similar actors. For this reason, it is essential that independent experts be involved in the preparation of data and the development of high-risk ML models, such as models enabling algorithmic credit scoring, which can be viewed in the broader light of the need for a system of collaborative governance built on public-facing and expert-facing accountability,[391] an aspect of which is data governance.

Although the AI Act Proposal establishes a data governance framework, this framework significantly excludes experts. The first tool-centred solution in this respect is thus ensuring that jurists are involved during data preparation and ML model development so that ML engineers and data scientists are not alone in defining concepts such as discrimination,[392] which should be acknowledged in the recitals to the AI Act Proposal. For instance, the proposed training, validation, and testing datasets could be subject to review by an independent board of experts, including legal experts. As individuals' credit score affects their ability to fully participate in society or improve their standard of living, such a solution is all the more essential in the context of algorithmic credit scoring.

The AI Act Proposal also strives for public-facing accountability through the establishment of an EU database for stand-alone high-risk AI systems;[393] however, this database might not effectively contribute to the public's ability to identify threats to fundamental rights, as the Proposal does not seem to envision it including information about risks revealed after the implementation of AI systems. The enforcement of the AI Act's rules on the basis of the database could also be significantly hampered due to the lack of a complaint mechanism in the Proposal.[394] Thus, another tool-centred solution is listing post-market monitoring discoveries of sources of risks to fundamental rights as data to be entered into the database, whereas a process-centred solution providing a mechanism to lodge a complaint against the user for non-compliance with the rules of the AI Act Proposal, in the case of algorithmic credit scoring, to the competent authority under Article 97 of the Directive 2013/36/EU.

---

[391] See n 382.
[392] Kaminski (n 10) 1575.
[393] See n 374.
[394] See n 387.

5.2 Alternative data regime

Related to the development of a ML model and its subsequent use is also the second process-centred solution, which concerns the use of alternative data in algorithmic credit scoring. Considering the risks that using such data poses to respect for individuals' fundamental rights, the operative part of the Proposal for a Directive on consumer credits, currently lacking a data regime, should contain a list of the types of data that can be used as input variables for the assessment of creditworthiness. The list should exclude behavioural data such as data about individuals' social network or any other data that are not *clearly* related to individuals' ability to reimburse credit. To the extent that alternative data can be used, the operative part of the Proposal or the recitals should also call creditors to justify the use of alternative data for the assessment of applicants with credit history.

5.3 Meaningful explanation and access to personal data

As transparency towards individuals in the context of algorithmic credit scoring carries the risk for attempts of system manipulation, creditors can rely on trade secret protection to restrict individuals' access to information about ML models' logic. The Proposal for a Directive on consumer credits and the GDPR thus do not provide individuals with a right to an explanation of how the ML model generated the credit score based on their algorithmic identity, which would allow them to understand completely why and how the ML model classified and judged them. However, if explanations under Article 18(6)(b) of the Proposal are to be sufficiently meaningful[395] to enable individuals to contest their credit score, they should at least know the *particularly relevant* algorithmic inferences made on the basis of the main variables, which should be made clear in the Proposal.

Considered as such could be, for instance, algorithmic inferences that explain the difference between an applicant's credit score and that of someone else and which are 'abnormal' or unusual in light of the (inferred) data about the applicant. Explanations encompassing the main variables used to calculate credit scores and this type of algorithmic inferences drawn from those variables, provided in a manner allowing individuals to comprehend the information, could thus be considered meaningful.

---

[395] Proposal for a Directive on consumer credits, rec 48.

The reasoning behind this conclusion is that 'good' or human-friendly explanations of decisions, as Molnar explains, are contrastive, selected, social, and focus on the abnormal.[396] First, contrastive explanations can be considered human-friendly because humans are prone to counterfactual thinking,[397] e.g. instead of an applicant wanting to know how their credit score was calculated, they are interested in the factors that could have improved it. A good explanation thus allows an applicant to understand the factors that determined the difference between their and someone else's credit score, which could also be an ideal credit score pertaining to a fictitious applicant.[398] Second, an explanation needs to be selected among different possible explanations of the factors that led to a particular event,[399] so a good explanation is limited to factors that can be considered 'abnormal' or unusual, the elimination of which would have significantly altered a particular result.[400] Explanations based on abnormal algorithmic inferences thus not only reduce the risk for individuals gaming the system by not disclosing all that negatively affected a credit score but only what stands out given the (inferred) data about an applicant, but they can also be considered good explanations. Finally, a good explanation is one that accounts for the social context,[401] which means that the explanation is tailored to the receiver's technical knowledge on the subject matter or lack thereof. A good explanation is thus one where the explainer, among other things, uses tailored language to explain how the ML model generated the credit score.

Similar to the limits of the right to explanation, the right of access does not enable individuals to access all algorithmic inferences about themselves. However, Article 15 GDPR should at least enable them access to those algorithmic inferences that are particularly relevant, which could be provided in the form of a summary of those data in an intelligible form.[402] As Wachter and Mittelstadt note, given the ECJ's view on the scope of Article 15 GDPR[403] and on data protection law[404] as not being intended to guarantee 'the greatest possible transparency of the decision-making process',[405] including of information on which the decisions are

---

[396] Molnar (n 378).
[397] ibid.
[398] ibid.
[399] ibid.
[400] ibid.
[401] ibid.
[402] Joined Cases C–141/12 and C–372/12 *YS* (C–141/12) *v Minister voor Immigratie, Integratie en Asiel and Minister voor Immigratie, Integratie en Asiel* (C–372/12) *v M and S* [2014] EU:C:2014:2081, para 70(2).
[403] Sandra Wachter and Brent Mittelstadt, 'A Right to Reasonable Inferences: Re-thinking Data Protection Law in the Age of Big Data and AI' (2019) 2019(2) Columbia Business Law Review 494, 546–547.
[404] ibid 499; 527.
[405] Case C–28/08 P *European Commission v The Bavarian Lager Co. Ltd.* [2010] EU:C:2010:378, para 49.

based,[406] a right of access to personal data in the form of algorithmic inferences calls for a broader interpretation of the remit of data protection law.[407] Most importantly, to materialise access to particularly relevant algorithmic inferences, the ECJ would have to recognise that such access does not factually impact an organisation's rights and freedoms.[408]

In short, with a view to filling the gaps in the legislation and thus ensuring respect for individuals' fundamental rights in algorithmic credit scoring, two additional process-centred solutions can be thought of: the right to explanation encompassing particularly relevant algorithmic inferences, which could be made clear in the Proposal for a Directive on consumer credits, and the right of access providing for access to such inferences, with the ECJ recognising that access to such data following a subject access request would generally not impact an organisation's rights and freedoms.

5.4 Meaningful human oversight and model interpretability

The GDPR does not clarify the level and quality of human involvement necessary for ADM not to be considered fully automated, nor do the Proposal for a Directive on consumer credits and the GDPR explain what kind of human intervention is sufficient to comply with Articles 18(6)(a) of the Proposal and 22(3) GDPR. While the AI Act Proposal does impose an obligation on the provider to develop the ML model so as to allow for meaningful human oversight, if the implementation of oversight measures is left to the user, their effectiveness depends on its commitment to mitigating the risks to fundamental rights. Given that the GDPR and the Proposals collectively do not address the possibility of not only automation bias but also selective adherence, the effectiveness of human oversight is thus questionable due to the risk for the HITL rubber-stamping the credit scores or selectively accepting them, which could lead to an increase in systemic bias.[409]

The AI Act Proposal could address this first by imposing an obligation on creditors as users to guarantee human oversight in accordance with Article 14(4) of the Proposal instead of with the instructions for use and by acknowledging therein the possible tendency of not only automatically or over-dependently relying on the AI system's output[410] but also selectively. However, as effective human oversight requires an intrinsically interpretable ML model or the

---

[406] ibid.
[407] Wachter and Mittelstadt (n 400) 580.
[408] European Data Protection Board (n 338), para 170; *Nowak* (n 262) paras 60–61.
[409] Kaminski (n 10) 1594.
[410] AI Act Proposal, art 14(4)(b).

use of model-agnostic methods, a tool-centred solution is also needed, namely ensuring that creditors use a ML model in algorithmic credit scoring whose functioning is sufficiently transparent.

In this respect, the AI Act Proposal requires providers to ensure that the operation of their ML models is sufficiently transparent to enable the interpretation of the systems' outputs; however, it exempts from its rules existing high-risk ML models that will not undergo significant changes. Therefore, to ensure that such ML models allow for meaningful human oversight, the AI Act Proposal should establish a special regime for them, which would subject them to interpretability requirements and their users to human oversight in line with Article 14(4) of the Proposal.

VI: Conclusion

6.1 Research outcome

The aim of this thesis was to identify gaps in EU legislation in regard to ensuring respect for individuals' fundamental rights in algorithmic credit scoring and therefrom suggest EU-level solutions, targeting both the process of and tool for it. In line with this aim, the research question was formulated as: '*What risks does algorithmic credit scoring pose to the respect for individuals' rights to non-discrimination, privacy, and data protection, and what process- and tool-centred solutions to mitigate them can be envisioned at the EU level?*'

6.1.1 Impact on the interpretation of rights

The first question that had to be answered was '*What technology or tool enables algorithmic credit scoring, and how does the use of that tool impact the interpretation of the rights to non-discrimination, privacy, and data protection as protected in the EU?*'

Chapter II thus explained the difference between traditional and algorithmic credit scoring in terms of the data being analysed and the technology behind it, namely machine learning (ML) algorithms. The first section briefly explained exactly how ML models are developed with a view to predicting the likelihood of defaulting on credit repayments and expressing it in a credit score.

The focus of this chapter, however, was on understanding the extent of EU-level protection of (the objects protected by) the rights to non-discrimination, privacy, and data protection, and how the use of ML algorithms affects their interpretation. The first sub-section of the second section thus concluded that a ML model's outputs could disadvantage certain individuals or groups more than others without a justified reason based on grounds like the type of web browser they use and indirect proxies for protected characteristics, and this would be classified as algorithmic bias but not as algorithmic discrimination.

The subsequent sub-section first conceptualised the rights to privacy and data protection as rights that protect individuals' development of their identity and personality through the protection of their informational privacy as an overarching aspect of privacy. This sub-section then concluded that informational privacy takes on a new meaning in the context of the use of ML algorithms, i.e. as one's ability to be aware of their algorithmic identity and have the power to contest it. This conclusion followed the finding that individuals cannot control ML algorithms' inferences based on seemingly neutral behaviour and actions and group membership (algorithmic classification) in advance and thus cannot, in the as yet understood sense of it, control how they present themselves to the world through information about them.

6.1.2 Risks to respect for rights

The second sub-question was also the first pillar of the main research question and read as '*How does algorithmic credit scoring affect individuals' access to credit, their private life, and personal data, and why does that pose a risk to the respect for their fundamental rights?*'

Chapter III found that choices regarding the definition of creditworthiness as the target variable and its associated class labels, construction of the training dataset, features to be used for ML, and the choice of the ML model can lead or contribute to discrimination in algorithmic credit scoring and thus possibly lead to the denial of access to credit. Furthermore, this section found that human oversight may not be an effective safeguard against the risk for discrimination due to the possibility for the human-in-the-loop rubber-stamping the credit scores or selectively accepting them, and that algorithmic discrimination may be difficult to identify and act upon in a certain case due to the ML model's intrinsic opacity or the protection of its logic as a trade secret.

As to the effects on individuals' private life and personal data, the subsequent section found that ML models' inferences re-define individuals' algorithmic identity based on correlations that can be spurious and algorithmic classification. This led to the conclusion that the denial of access to credit-scoring ML models' inferences poses a risk to individuals' free (external) identity-building and control over (the accuracy of) their personal data, and thus to the respect for their rights to privacy and data protection.

This section further found that the use of data about individuals' social network in algorithmic credit scoring can undermine individuals' free personality-building and development of social relations, as it can dissuade them from associating with those whom they consider un-creditworthy and can also conflict with the processing of their personal data in a manner that is not unjustifiably detrimental and is in line with the data minimisation principle.

Lastly, this section concluded that the processing of individuals' personal data in a way that corresponds to their expectations is also at risk when other types of alternative data are used. What led to this conclusion was the finding that ML models can reveal insights that go beyond the limits of human observation and which are less intuitive than those based on credit data. This was also found to carry the potential of dissuading individuals from engaging in any activity they believe could potentially negatively affect their credit score and thus a risk to the respect for their right to personal development and, accordingly, their right to privacy.

6.1.3 Gaps in legislation and solutions

The last sub-question, which also concerned the second pillar of the main research question and read as '*What process- and tool-centred solutions could be employed with a view to filling the gaps in the legislation and thus ensuring respect for fundamental rights in algorithmic credit scoring, and where could they be regulated?*', was a response to the third sub-question, namely '*What EU legislation regulates the process of, or the tool for, algorithmic credit scoring, and what gaps can be identified in the legislation in regard to ensuring respect for individuals' fundamental rights?*'

This thesis explained why algorithmic credit scoring triggers the application of various pieces of legislation, namely the CCD, which will be replaced by the Proposal for a Directive on consumer credits, the GDPR, and the AI Act Proposal, which were analysed in Chapter IV by considering the mechanisms that would work toward tackling the risks to the respect for individuals' fundamental rights set out in Chapter III.

The analysis showed that there are still significant gaps in the currently applicable CCD and the GDPR in this respect, as the CCD is based on traditional credit scoring and the GDPR most notably lacks strong safeguards in the case of automated individual decision-making. The analysis also revealed that the Proposal for a Directive on consumer credits and the AI Act

Proposal do not sufficiently address these shortcomings, especially given the absence of an alternative data regime in the Proposal for a Directive on consumer credits and the minimal role of external experts and civil society in the construction, training, and deployment of ML algorithms envisioned by the AI Act Proposal. Accordingly, several process- and tool-centred solutions were suggested in Chapter V that build on the existing mechanisms in EU legislation.

The solution for a collaborative data governance system builds on the data governance framework established by the AI Act Proposal by proposing that training, validation, and testing datasets for high-risk ML models be subject to review by an independent board of experts, post-market monitoring discoveries of sources of risks to fundamental rights be listed as data to be entered into the EU database for stand-alone high-risk AI systems, and that a complaint mechanism for non-compliance with the rules of the AI Act Proposal be established.

Related to the development of ML models and their subsequent use is also the solution of an alternative data regime, which builds on the regulation of algorithmic credit scoring in the CCD and the Proposal for a Directive on consumer credits by suggesting the Proposal to include a list of usable types of data for the assessment of creditworthiness in its operative part, excluding inter alia data about individuals' social network, and to call on creditors to justify the use of alternative data for the assessment of applicants with credit history.

The subsequent section then presented two solutions for making the rights to explanation and of access to personal data as contained in the Proposal for a Directive on consumer credits and the GDPR more meaningful in terms of enabling individuals to effectively contest their credit score: the Proposal for a Directive on consumer credits making clear that an explanation is to be based, inter alia, on particularly relevant algorithmic inferences and the ECJ recognising that access to such data following a subject access request would generally not impact an organisation's rights and freedoms. As concluded in this section, algorithmic inferences explaining the difference between an applicant's credit score and someone else's that are abnormal in light of the data about the applicant can be considered particularly relevant.

Finally, the solutions for meaningful human oversight and model interpretability build on the mechanisms contained in the AI Act Proposal by proposing that a special regime for existing high-risk ML models that will not undergo significant changes and thus be exempted from the rules of the AI Act Proposal be established, which would subject them to interpretability requirements and their users to human oversight in line with Article 14(4) of the Proposal, such oversight also being established as a separate general obligation for users and the relevant provision acknowledging the possibility of selective adherence to the AI system's output.

6.2 Final thoughts

This thesis presented solutions for ensuring respect for fundamental rights in algorithmic credit scoring; yet, it did not address the simplest and perhaps most effective solution of not using ML but instead rule-based algorithms for credit scoring. Given the social and democratic importance of access to credit herein emphasised, the use of AI to assess individuals' creditworthiness and thus AI determining their ability to fully participate in society or improve their standard of living, in fact, may not be a reality that democracies upholding the values and objectives of the Rule of Law should embrace.

That said, technological progress is inevitable, and allowing ML algorithms to assess individuals' creditworthiness can be beneficial to society, as it can lead to credit-granting decisions based on credit scores that more accurately reflect individuals' likelihood to repay their debt obligations in full, thus preventing over-indebtedness. Perhaps the focus should thus shift toward the benefits of the use of this technology, not only for creditors but also for consumers. Which of these two arguments prevails, however, cannot be concluded without further consideration.

Bibliography

Primary sources

1. Case-law

1.1 Court of Justice of the European Union

- David L. Parris v Trinity College Dublin and Others (Case C-443/15) [2016] EU:C:2016:897
- European Commission v The Bavarian Lager Co. Ltd. (Case C–28/08 P) [2010] EU:C:2010:378
- Jyske Finans A/S v Ligebehandlingsnaevnet, acting on behalf of Ismar Huskic (Case C-668/15) [2017] EU:C:2017:278
- Komisia za zashtita na potrebitelite v Evelina Kamenova and Okrazhna prokuratura – Varna (Case C–105/17) [2018] EU:C:2018:808
- Peter Nowak v Data Protection Commissioner (Case C-434/16) [2017] EU:C:2017:994
- YS v Minister voor Immigratie, Integratie en Asiel and Minister voor Immigratie, Integratie en Asiel v M and S (Joined Cases C–141/12 and C–372/12) [2014] EU:C:2014:2081

1.2 European Court of Human Rights

- Bărbulescu v Romania App no 61496/08 (ECtHR, 5 September 2017)
- Botta v Italy App no 21439/93 (ECtHR, 24 February 1998)
- P.G. and J.H. v the United Kingdom App no 44787/98 (ECtHR, 25 September 2001)
- Peck v the United Kingdom App no 44647/98 (ECtHR, 28 January 2003)
- Pfeifer v Austria App no 12556/03 (ECtHR, 15 November 2007)
- Satakunnan Markkinapörssi Oy and Satamedia Oy v Finland App no 931/13 (ECtHR, 27 June 2017)
- Von Hannover v Germany App no 59320/00 (ECtHR, 24 June 2004)

2. Legislation

2.1 Council of Europe

Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended)

2.2 European Union

Primary law

- Charter of Fundamental Rights of the European Union [2010] OJ C83/2
- Consolidated Version of the Treaty on European Union [2012] OJ C326/1
- Consolidated Version of the Treaty on the Functioning of the European Union [2016] OJ C202/1

Secondary law

- Commission, 'Proposal for a Directive of the European Parliament and of the Council on consumer credits' COM (2021) 347 final
- Commission, 'Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts' COM (2021) 206 final
- Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin [2000] OJ L180/22
- Directive 2008/48/EC of the European Parliament and of the Council of 23 April 2008 on credit agreements for consumers and repealing Council Directive 87/102/EEC [2008] OJ L133/66
- Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure [2016] OJ L157/1
- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 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

Secondary sources

1. Articles

Aggarwal N, ‘The Norms of Algorithmic Credit Scoring’ (2021) 80(1) Cambridge Law Journal 42

Alon-Barkat S and Busuioc M, ‘Human-AI Interactions in Public Sector Decision-making: ‘Automation Bias’ and ‘Selective Adherence’ to Algorithmic Advice’ (preprint, arXiv:2103.02381, 28 January 2022) 1 <https://arxiv.org/abs/2103.02381> accessed 16 May 2022

Barocas S and Selbst A D, ‘Big Data’s Disparate Impact’ (2016) 104(3) California Law Review 671

Bayamlıoğlu E, ‘The Right to Contest Automated Decisions Under the General Data Protection Regulation: Beyond the So-called “Right to Explanation”’ (2021) Regulation & Governance (special issue) 1 <https://onlinelibrary.wiley.com/doi/abs/10.1111/rego.12391> accessed 1 June 2022
– – and Leenes R E, ‘The “Rule of Law” Implications of Data-driven Decision-making: A Techno-Regulatory Perspective’ (2018) 10(2) Law, Innovation and Technology 295

Berman E, ‘A Government of Laws and Not of Machines’ (2018) 98(5) Boston University Law Review 1277

Binns R and Veale M, ‘Is That Your Final Decision? Multi-stage Profiling, Selective Effects, and Article 22 of the GDPR’ (2021) 11(4) International Data Privacy Law 319

Bruckner M A, ‘The Promise and Perils of Algorithmic Lenders’ Use of Big Data’ (2018) 93(1) Chicago-Kent Law Review 3

Burrell J, ‘How the Machine “Thinks”: Understanding Opacity in Machine Learning Algorithms’ (2016) 3(1) Big Data & Society 1 <https://journals.sagepub.com/doi/full/10.1177/2053951715622512> accessed 14 May 2022

Busuioc M, ‘Accountable Artificial Intelligence: Holding Algorithms to Account’ (2020) 81(5) Public Administration Review 825

de Laat P B, ‘Algorithmic Decision-making Employing Profiling: Will Trade Secrecy Protection Render the Right to Explanation Toothless?’ (2022) 24(2) Ethics and Information Technology 1 <https://link.springer.com/article/10.1007/s10676-022-09642-1> accessed 8 June 2022

de Vries S A, ‘Balancing Fundamental Rights with Economic Freedoms According to the European Court of Justice’ (2013) 9(1) Utrecht Law Review 169 <https://www.utrechtlawreview.org/articles/abstract/10.18352/ulr.220/> accessed 3 May 2022

Edwards L and Veale M, ‘Enslaving the Algorithm: From a “Right to an Explanation” to a “Right to Better Decisions”?’ (arXiv:1803.07540, 2 July 2018) 1 <https://arxiv.org/abs/1803.07540v2> accessed 1 June 2022

Fan J, Han F and Liu H, ‘Challenges of Big Data analysis’ (2014) 1(2) National Science Review 293

Fenwick M and Jurcys P, ‘From Cyborgs to Quantified Selves: Augmenting Privacy Rights with User-centric Technology and Design’ (2022) 13(1) Journal of Intellectual Property, Information Technology and E-Commerce Law 20

Ferretti F, ‘The Legal Framework of Consumer Credit Bureaus and Credit Scoring in the European Union: Pitfalls and Challenges—Overindebtedness, Responsible Lending, Market Integration, and Fundamental Rights’ (2013) 46(3) Suffolk University Law Review 791

Greenstein S, ‘Preserving the Rule of Law in the Era of Artificial Intelligence (AI)’ (2021) Artificial Intelligence and Law (Online first articles) 1 <https://link.springer.com/article/10.1007/s10506-021-09294-4> accessed 21 March 2022

Hildebrandt M, 'Understanding Law and the Rule of Law: A Plea to Augment CS Curricula—Why Law Matters for Computer Scientists and Other Folk' (2021) 64(5) Communications of the ACM 28
– – 'Privacy As Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning' (2019) 20(1) Theoretical Inquiries in Law 83

Hitomi K, 'Automation —Its Concept and a Short History' (1994) 14(2) Technovation 121

Hurley M and Adebayo J, 'Credit Scoring in the Era of Big Data' (2016) 18(1) Yale Journal of Law and Technology 148

Hutchinson T, 'The Doctrinal Method: Incorporating Interdisciplinary Methods in Reforming the Law' (2015) 8(3) Erasmus Law Review 130

Janssen M and others, 'Data Governance: Organizing Data for Trustworthy Artificial Intelligence' (2020) 37(3) Government Information Quarterly 1 <https://www.sciencedirect.com/science/article/abs/pii/S0740624X20302719> accessed 9 June 2022

Kaminski M E, 'Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability' (2019) 92(6) Southern California Law Review 1529
– – 'The Right to Explanation, Explained' (2019) 34(1) Berkeley Technology Law Journal 189

Kear M, 'Playing the Credit Score Game: Algorithms, "Positive" Data and the Personification of Financial Objects' (2017) 46(3-4) Economy and Society 1 <https://www.tandfonline.com/doi/full/10.1080/03085147.2017.1412642?scroll=top&needAccess=true> accessed 7 June 2022

Keats Citron D and Pasquale F, 'The Scored Society: Due Process for Automated Predictions' (2014) 89(1) Washington Law Review 1

Kokott J and Sobotta C, 'The Distinction Between Privacy and Data Protection in The Jurisprudence of the CJEU and the ECtHR' (2013) 3(4) International Data Privacy Law 222

Koops B-J, ‘Privacy Spaces’ (2018) 121(2) West Virginia Law Review 611
– – and Leenes R E, ‘“Code” and the Slow Erosion of Privacy’ (2005) 12(1) Michigan Telecommunications and Technology Law Review 115
– – and others, ‘A Typology of Privacy’ (2017) 38(2) University of Pennsylvania Journal of International Law 483

Kordzadeh N and Ghasemaghaei M, ‘Algorithmic Bias: Review, Synthesis, and Future Research Directions’ (2021) 31(3) European Journal of Information Systems 1 <https://www.tandfonline.com/doi/full/10.1080/0960085X.2021.1927212> accessed 5 May 2022

McDermott Y, ‘Conceptualising the Right to Data Protection in an Era of Big Data’ (2017) 4(1) Big Data & Society 1 <https://journals.sagepub.com/doi/10.1177/2053951716686994> accessed 10 May 2022

Mittelstadt B, ‘From Individual to Group Privacy in Big Data Analytics’ (2017) 30(4) Philosophy & Technology 475

Mulligan D K and others, ‘This Thing Called Fairness: Disciplinary Confusion Realizing a Value in Technology’ (2019) 3 Proceedings of the ACM on Human-Computer Interaction 1 <https://dl.acm.org/doi/abs/10.1145/3359221> accessed 5 May 2022

Selbst A D and Powles J, ‘Meaningful Information and the Right to Explanation’ (2017) 7(4) International Data Privacy Law 233

Surden H, ‘Machine Learning and Law’ (2014) 89(1) Washington Law Review 87

van der Sloot B, ‘Privacy as Personality Right: Why the ECtHR’s Focus on Ulterior Interests Might Prove Indispensable in the Age of “Big Data”’ (2015) 31(80) Utrecht Journal of International and European Law 25 <https://utrechtjournal.org/articles/10.5334/ujiel.cp/> accessed 9 May 2022

Veale M and Binns R, ‘Fairer Machine Learning in the Real World: Mitigating Discrimination Without Collecting Sensitive Data’ (2017) 4(2) Big Data & Society 1 <https://journals.sagepub.com/doi/abs/10.1177/2053951717743530> accessed 16 May 2022

Veale M and Zuiderveen Borgesius F, ‘Demystifying the Draft EU Artificial Intelligence Act—Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach’ (2021) 22(4) Computer Law Review International 97

Wachter S and Mittelstadt B, ‘A Right to Reasonable Inferences: Re-thinking Data Protection Law in the Age of Big Data and AI’ (2019) 2019(2) Columbia Business Law Review 494

Waldman A E, ‘Power, Process, and Automated Decision-making’ (2019) 88(2) Fordham Law Review 613

Wei Y and others, ‘Credit Scoring with Social Network Data’ (preprint, 2015) 35(2) Marketing Science 1 <https://repository.upenn.edu/marketing_papers/383/> accessed 20 May 2022

Zuiderveen Borgesius F and Gerards J, ‘Protected Grounds and the System of Non-discrimination Law in the Context of Algorithmic Decision-making and Artificial Intelligence’ (working draft, 2021) 1 <https://works.bepress.com/frederik-zuiderveenborgesius/65/> accessed 6 May 2022

2. Books

Bali R and Sarkar D, *R Machine Learning by Example* (Packt 2016)

Brownlee K and Stemplowska Z, ‘Financial Inclusion, Education, and Human Rights’ in Tom Sorell and Luis Cabrera (eds), *Microfinance, Rights and Global Justice* (Cambridge University Press 2015)

Davies G, ‘The Consumer, the Citizen, and the Human Being’ in Dorota Leczykiewicz and Stephen Weatherill (eds), *The Images of the Consumer in EU Law* (Hart Publishing 2016)

Gerards J, *General Principles of the European Convention on Human Rights* (manuscript, Cambridge University Press 2018)

Greer S, Gerards J and Slowe R, *Human Rights in the Council of Europe and the European Union* (Cambridge University Press 2018)

Haig B D, ‘Spurious Correlation’ in Neil J. Salkind (ed), *Encyclopedia of Measurement and Statistics* (online version, Sage Publications 2007) <https://www.researchgate.net/publication/315829671_Spurious_correlation> accessed 20 May 2022

Hendriks F, *Vital Democracy: A Theory of Democracy in Action* (Oxford Scholarship Online 2010)

Molnar C, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* (2nd edn, 2022) <https://christophm.github.io/interpretable-ml-book> accessed 2 June 2022

Pollicino O and De Gregorio G, ‘Constitutional Law in the Algorithmic Society’ in Hans-W. Micklitz and others (eds), *Constitutional Challenges in the Algorithmic Society* (Cambridge University Press 2021)

Ramsay I, ‘Changing Policy Paradigms of EU Consumer Credit and Debt Regulation’ in Dorota Leczykiewicz and Stephen Weatherill (eds), *The Images of the Consumer in EU Law* (Hart Publishing 2016)

Sellers M N S, ‘What Is the Rule of Law and Why Is It So Important?’ in James R. Silkenat, James E. Hickey Jr. and Peter D. Barenboim (eds), *The Legal Doctrines of the Rule of Law and the Legal State (Rechtsstaat)* (Springer 2014)

Zalnieriute M, Bennett Moses L and Williams G, ‘Automating Government Decision-making: Implications for the Rule of Law’ in Siddharth Peter de Souza and Maximilian Spohr (eds), *Technology, Innovation and Access to Justice: Dialogues on the Future of Law* (Edinburgh University Press 2021)

3. Other secondary sources

3.1 Article 29 Data Protection Working Party/European Data Protection Board/European Data Protection Supervisor

Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (No 17/EN, 6 February 2018)

European Data Protection Board, 'Guidelines 01/2022 on data subject rights – Right of access' (version 1.0, 18 January 2022)

European Data Protection Board, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (version 2.0, 20 October 2020)

European Data Protection Supervisor, 'Opinion 11/2021 on the Proposal for a Directive on consumer credits' (26 August 2021)

3.2 Council of Europe

Monitoring/Advisory bodies of the Council of Europe

- European Commission for the Efficiency of Justice, 'European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment' (31st plenary meeting, Strasbourg, 3–4 December 2018)
- Venice Commission, 'Rule of Law Checklist' (Study No 711/2013, 18 March 2016)

Registry of the European Court of Human Rights

- 'Guide on Article 8 of the European Convention on Human Rights—Right to respect for private and family life, home and correspondence' (updated on 31 August 2021, Council of Europe/European Court of Human Rights 2021)
- 'Guide on Article 14 of the European Convention on Human Rights and on Article 1 of Protocol No. 12 to the Convention—Prohibition of discrimination' (updated on 31 August 2021, Council of Europe/European Court of Human Rights 2021)
- 'Guide to the Case-Law of the European Court of Human Rights—Data protection' (updated on 31 December, Council of Europe/European Court of Human Rights 2021)

3.3 Other guidelines/Reports/Studies

Advisory Council on International Affairs, 'The Will of the People? The Erosion of Democracy under the Rule of Law in Europe' (No 104, June 2017) <https://www.advisorycouncilinternationalaffairs.nl/documents/publications/2017/06/02/the-will-of-the-people> accessed 3 May 2022

Bollier D, *The Promise and Peril of Big Data* (The Aspen Institute 2010)

Commission, 'White Paper on Artificial Intelligence: A European approach to excellence and trust' (White Paper) COM (2020) 65 final

Complaint from the European Center for Digital Rights (NOYB) (31 July 2020) 2 <https://noyb.eu/en/credit-scoring-negative-credit-rating-generated-without-data> accessed 21 May 2022

DIGITALEUROPE, 'DIGITALEUROPE's initial findings on the proposed AI Act' (6 August 2021) 2 <https://www.digitaleurope.org/resources/digitaleuropes-initial-findings-on-the-proposed-ai-act/> accessed 2 June 2022

European Telecommunications Standards Institute, 'Experiential Networked Intelligence (ENI); Definition of Data Processing Mechanisms' (ETSI GR ENI 009 V1.1.1, June 2021) <https://www.etsi.org/committee/1423-eni> accessed 15 June 2022

Gerards J and Xenidis R, *Algorithmic Discrimination in Europe: Challenges and Opportunities for Gender Equality and Non-discrimination Law* (Publications Office of the European Union 2021)

Leslie D and others, 'Artificial Intelligence, Human Rights, Democracy, and the Rule of Law: A Primer' (The Alan Turing Institute 2021) <https://www.turing.ac.uk/research/publications/ai-human-rights-democracy-and-rule-law-primer-prepared-council-europe> accessed 21 March 2022

Raso F and others, ‘Artificial Intelligence & Human Rights: Opportunities & Risks’ (Berkman Klein Center Research Publication 2018) <https://cyber.harvard.edu/publication/2018/artificial-intelligence-human-rights> accessed 21 March 2022

Valant J, *Consumer Protection in the EU* (Publications Office of the European Union 2015)

World Bank Group, ‘Credit Scoring Approaches Guidelines’ (2 April 2020)

Zuiderveen Borgesius F, ‘Discrimination, Artificial intelligence, and Algorithmic Decision-making’ (Council of Europe, Directorate General of Democracy 2018) <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73> accessed 5 May 2022

3.4 Websites

Aggarwal N, ‘Law and Autonomous Systems Series: Algorithmic Credit Scoring and the Regulation of Consumer Credit Markets’ (*University of Oxford/Faculty of Law*, 1 November 2018) <https://www.law.ox.ac.uk/business-law-blog/blog/2018/11/law-and-autonomous-systems-series-algorithmic-credit-scoring-and> accessed 21 March 2022

Amazon Web Services, ‘What Is Data Labeling for Machine Learning?’ (*aws*) <https://aws.amazon.com/sagemaker/data-labeling/what-is-data-labeling/> accessed 15 May 2022

Brownlee J, ‘What Is Deep Learning?’ (*Machine Learning Mastery*, 16 August 2019) <https://machinelearningmastery.com/what-is-deep-learning/> accessed 3 May 2022

Card D, ‘The “Black Box” Metaphor in Machine Learning’ (*Medium*, 5 July 2017) <https://dallascard.medium.com/the-black-box-metaphor-in-machine-learning-4e57a3a1d2b0> accessed 16 June 2022

CloudFactory, ‘The Ultimate Guide to Data Labeling for Machine Learning’ (*cloudfactory)* <https://www.cloudfactory.com/data-labeling-guide> accessed 15 May 2022

Cohen R B, ‘What Is Redundant Encoding And Should I Care?’ (*Lexology*, 29 November 2016) <https://www.lexology.com/library/detail.aspx?g=d4f54bc5-704d-4ad3-9047-500493cdc41d> accessed 16 May 2022

Council of Europe, ‘Intersectionality and Multiple Discrimination’ (*Council of Europe*) <https://www.coe.int/en/web/gender-matters/intersectionality-and-multiple-discrimination> accessed 6 May 2022

DeCew J, ‘Privacy’ (*Stanford Encyclopedia of Philosophy*, 18 January 2018) <https://plato.stanford.edu/entries/privacy/> accessed 16 June 2022

European Center for Digital Rights, ‘Credit Scoring: Negative Credit Rating Generated Without Data’ (*noyb*, 4 August 2020) <https://noyb.eu/en/credit-scoring-negative-credit-rating-generated-without-data> accessed 19 May 2022

European Center for Digital Rights, ‘Data Voodoo: Credit Ranking Agency CRIF Creates Credit Rating Out of Thin Air’ (*noyb*, 4 August 2021) <https://noyb.eu/en/data-voodoo-credit-ranking-agency-crif-creates-credit-rating-out-thin-air> accessed 21 May 2022

European Parliament, ‘Big Data: Definition, Benefits, Challenges (infographics)’ (*European Parliament*, 29 March 2021) <https://www.europarl.europa.eu/news/en/headlines/society/20210211STO97614/big-data-definition-benefits-challenges-infographics> accessed 21 March 2022

Information Commissioner’s Office, ‘Automated Decision-making and Profiling’ (*ico*) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/> accessed 21 March 2022

Information Commissioner’s Office, ‘Principle (a): Lawfulness, Fairness and Transparency’ (*ico*) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/lawfulness-fairness-and-transparency/> accessed 24 May 2022

Information Commissioner’s Office, ‘Principle (c): Data Minimisation’ (*ico*) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/data-minimisation/> accessed 27 March 2022

Merriam-Webster Dictionary, ‘Creditworthy’ (*Merriam-Webster*) <https://www.merriam-webster.com/dictionary/creditworthy> accessed 21 March 2022

Merriam-Webster Dictionary, ‘Decision’ (*Merriam-Webster*) <https://www.merriam-webster.com/dictionary/decision> accessed 21 March 2022

Merriam-Webster Dictionary, ‘Decision-making’ (*Merriam-Webster*) <https://www.merriam-webster.com/dictionary/decision-making> accessed 21 March 2022

Merriam-Webster Dictionary, ‘Social Network’ (*Merriam-Webster*) <https://www.merriam-webster.com/dictionary/social%20network> accessed 20 May 2022

Oxford Reference, ‘Documentary Research’ (*Oxford Reference*) <https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095724431> accessed 21 March 2022

White & Case LLP, ‘Algorithms and Bias: What Lenders Need to Know’ (*White & Case*, 20 January 2017) <https://www.whitecase.com/publications/insight/algorithms-and-bias-what-lenders-need-know> accessed 10 May 2022