MARCELA MATTIUZZO

Algorithmic Discrimination – The Challenge of Unveiling Inequality in Brazil

Masters Dissertation

Supervisor: Professor Virgílio Afonso da Silva

UNIVERSITY OF SÃO PAULO
FACULTY OF LAW
São Paulo
2019

MARCELA MATTIUZZO

Algorithmic Discrimination – The Challenge of Unveiling Inequality in Brazil

Masters dissertation submitted to the graduate program in law, at the School of Law at the University of São Paulo, within the field of concentration of Constitutional Law, under the supervision of Professor Virgílio Afonso da Silva.

UNIVERSITY OF SÃO PAULO
SCHOOL OF LAW
São Paulo
2019

Catalogação na Publicação

Serviço de Processos Técnicos da Biblioteca da

Faculdade de Direito da Universidade de São Paulo

Mattiuzzo, Marcela

Algorithmic Discrimination - The Challenge of Unveiling Inequality in Brazil / Marcela Mattiuzzo. -- São Paulo, 2019.
145 p. ; 30 cm.

Dissertação (Mestrado) – Programa de Pós-Graduação em Direito, Faculdade de Direito, Universidade de São Paulo, São Paulo, 2019.
Orientador: Virgílio Afonso da Silva.

1. Discriminação algorítmica. 2. Inteligência artificial. 3. Igualdade. 4. Big Data. 5. Governança algorítmica. I. Silva, Virgílio Afonso, orient. II. Título.

MATTIUZZO, M. **Algorithmic Discrimination –** The Challenge of Unveiling Inequality in Brazil. 2019. 145 f. Dissertação (Mestrado em Direito Constitucional) – Faculdade de Direito, Universidade de São Paulo, São Paulo, 2019.

Aprovado em:

Banca Examinadora

Prof(a). Dr(a). _____

Instituição: _____

Julgamento: _____

Prof(a). Dr(a). _____

Instituição: _____

Julgamento: _____

Prof(a). Dr(a). _____

Instituição: _____

Julgamento: _____

Acknowledgments

I would like to thank my adviser, Professor Virgílio Afonso da Silva, who indeed made this dissertation much better than it initially was. And also my "informal advisers," Artur Péricles Lima Monteiro, Flávio Marques Prol, Laura Schertel Mendes, and Victor Oliveira Fernandes, for your many inputs.

I am grateful to my family and to my friends, who provided the most important kind of assistance: encouragement. And to everyone at VMCA, especially Vinicius Marques de Carvalho, who worked extra hours to allow me to finish this work.

Quem sabe direito o que uma pessoa é? Antes sendo: julgamento é sempre defeituoso, porque o que a gente julga é o passado.

Guimarães Rosa

Abstract

MATTIUZZO, Marcela. **Algorithmic Discrimination** – The Challenge of Unveiling Inequality in Brazil. 2019. p. 145. Dissertation (Master Degree in Constitutional Law) – School of Law, University of São Paulo, São Paulo, 2019.

Abstract: the objective of this work is to provide some clarity on what the role of the Law can be in shedding light upon algorithmic discrimination, as well as how legal instruments could help minimize its risks, with a specific focus on the Brazilian jurisdiction.
To do so, it first engages in a debate about what algorithms indeed are, and how the emergence of the data-driven economy, Big Data, and machine learning have leveraged the use of automated systems. Next, it conceptualizes discrimination, and suggesting a typology of algorithmic discrimination that takes statistics into account to provide a rationalization of the debate.
It moves on to discussing the path towards enforcing legal norms against discriminatory outcomes running from the use of algorithms. Because legislation specifically aimed at fighting automated systems is still scarce (or application of the current legislation to the problem is contentious), it engages in a debate about the horizontal effects of fundamental rights – given that a relevant part of discriminatory practices occur among private parties, and the most basic defense an individual has against discrimination is the constitutional right to equality. It then analyzes ordinary legislation in three jurisdictions, the United States of America, Germany, and Brazil, that could also be enforced against discriminatory practices running from algorithms, with a special focus on the Brazilian legislation. The legislative debate concludes with the presentation of two concrete cases of algorithmic discrimination, one concerning the unemployment policy in Poland, and the other regarding credit scoring in Brazil. The cases are presented so that the applicability of Brazilian legislation to deal with algorithmic discrimination can be discussed.
The final chapter is focused on debating the path forward and what can and should be done by experts, legislators, and policymakers to foster algorithmic innovation without losing sight of its potential for discrimination. It first presents the literature on algorithmic governance and the many proposals for dealing with the problem – dedicating a specific section to the challenges brought about by machine learning – and then sets out an agenda for Brazil.

Keywords: algorithmic discrimination, artificial intelligence, equality, big data, algorithmic governance.

Resumo

MATTIUZZO, Marcela. **Discriminação Algorítmica** – O desafio em desvendar a desigualdade no Brasil. 2019. p. 145. Dissertação (Mestrado em Direito Constitucional) – Faculdade de Direito, Universidade de São Paulo, São Paulo, 2019.

Resumo: O objetivo desse trabalho é esclarecer qual é o papel que o Direito pode desempenhar no debate sobre a discriminação algorítmica, assim como de que maneira os instrumentos jurídicos podem auxiliar a mitigar os riscos discriminatórios desse tipo de prática, com foco especial na jurisdição brasileira.

Para isso, primeiro o trabalho propõe um debate sobre o que são algoritmos, e como a emergência da economia de dados, do Big Data e de técnicas de *machine learning* impulsionam o uso de sistemas automatizados. Em seguida, conceitua-se a discriminação, propondo-se uma tipologia para a discriminação algorítmica que leva em conta questões estatísticas, a fim de racionalizar a discussão.

A dissertação então parte para o debate sobre os caminhos para a aplicação de normas jurídicas em face de discriminação algorítmica. Dado que leis e normas especificamente voltados a esse tema ainda não são muito difundidas (e que a aplicação da legislação existente a essa questão é controversa), o trabalho propõe um debate sobre a eficácia horizontal dos direitos fundamentais – tendo em vista que boa parte das práticas discriminatórias via uso de algoritmos se dá entre partes privadas, e que a defesa mais básica que um indivíduo tem contra a discriminação é o direito constitucionalmente garantido à igualdade. Passa-se então a uma análise da legislação ordinária em três jurisdições, Estados Unidos da América, Alemanha e Brasil, legislação essa que pode também ser aplicada em casos de práticas discriminatórias levadas a cabo via algoritmos, dando especial destaque ao caso brasileiro. Esse debate legislativo é concluído com a apresentação de dois casos concretos, um que diz respeito à política de acesso a emprego na Polônia e outro que trata das práticas de *credit scoring* no Brasil. Os casos são apresentados de forma a se pensar a eventual possibilidade de uso de regras brasileiras para lidar com os temas discriminatórios que se colocam concretamente.

O capítulo final tem como foco o debate do caminho a ser trilhado, e qual pode e deve ser feito por especialistas, legisladores e aplicadores do direito para promover a inovação no campo algorítmico sem perder de vista seus potenciais impactos discriminatórios. Primeiro, apresenta-se a literatura sobre governança algorítmica e as muitas propostas que pretendem endereçar o tema – com especial atenção aos desafios apresentados pelo *machine learning* – e então delineia-se uma agenda para o Brasil sobre o assunto.

Palavras-chave: discriminação algorítmica, inteligência artificial, igualdade, big data, governança algorítmica.

Table of Contents

1 Introduction

The game of Go was developed in China over 2,500 years ago, it is the oldest board game that human kind is aware of, and counts with a legion of fans. The objective of the game is simple: each of the two players must occupate more board territory than the opponent in order to win, either by capturing opponent's stones or by surrounding empty space. The black and white stones are the tools the players must handle to reach their objective; and they both take turns placing them on the board. Apparently, the game is very simple, but in reality it can be terribly complex. The reason is the number of possible plays: whereas in chess the initial move by any player is limited to 20 moves, in Go it spikes to 361, and grows exponentially thereafter. To have an idea, the estimated number of possible board configurations in chess is 10^{120}, whereas in Go it is 10^{174}, meaning there are trillions more configurations in Go than in chess.

In 2016, DeepMind, a research project turned start-up, later acquired by Google's parent company Alphabet, challenged world-champion Lee Sedol to a game of Go.[1] The challenger: AlphaGo, a deep neural network developed by a group of programmers. Demis Hassabis, DeepMind's CEO and co-founder, a child chess prodigy turned computer scientist, entered into this endeavor after deciding that to test the true potential of the topic he had been studying since graduation from Cambridge University – artificial intelligence – he had to crack the game of Go by beating a professional player.

AlphaGo beat Sedol 4-1 in a 5-match challenge. As Hassabis explains in the AlphaGo documentary, originally the program learned how to play by watching over 100,000 games by strong amateurs, and its objective was to mimic human players. What gave it the real leap, however, was learning from its own mistakes. It has since improved its capacity by playing games against itself, and became the best Go player in the world.[2] This celebrated achievement of artificial intelligence is by no means the only application of algorithms to problem-solving,

[1] Before challenging Sedol, AlphaGo had already beat the European Champion, Fan Hui. The results of the European matches were published by DeepMind in Nature. See: SILVER, D. et al. **Mastering the game of Go with deep neural networks and tree search.** Nature, vol. 529, January 28, 2016. Available at: <https://storage.googleapis.com/deepmind-media/alphago/AlphaGoNaturePaper.pdf>. Access: January 10, 2019.
[2] SILVER, D. et al. **Mastering the game of Go without human knowledge**. Nature, vol. 550, January 19, 2017.

quite on the contrary; algorithms have been applied to a wide-array of matters with various outcomes, notably, they have been used by public authorities to inform decisions on various matters, ranging from sentencing to determining the best treatment for a given illness.

Imagine that one day, instead of leaving decisions of whether or not an individual should be imprisoned to a judge, an algorithm[3] was employed to calculate the risk of convicts' future conduct as the basis for the criminal sanction; that is, the prison term would vary depending on whether the algorithm identified convicts as presenting "low" or "high" risk to society. Imagine that to reach its decision the algorithm evaluated the convicts' responses to a wide array of questions that included whether the person's parents had ever been incarcerated, whether her friends had ever consumed illegal drugs, how often this person had been involved in fights at school, and so on. Imagine, lastly, that the mathematical formulas and weighting used by this algorithm were not public, such that there was no public access to the input used or outputs generated by it, and that convicts were thus unable to determine which aspects of their behavior led to their classification as high or low risk to society.

This scenario might sound like science fiction, but algorithms such as the one described above are already in use in several jurisdictions within the United States, and their implementation has also begun in Brazil.[4] Much about their use is questionable,[5] but this work

[3] One of the goals of this work is to clarify what is meant by an algorithm and how it carries out tasks such as risk assessment. In short, an algorithm is nothing more than a sequence of actions to be performed in a given order, which generates a specific result. One can have an algorithm for making dinner, an algorithm for taking a bath, an algorithm for walking to work every day, etc. Thus, algorithms do not have to be computerized, nor even need to be complex.

[4] The most popular tools are COMPAS and LSI-R (Level of Service Inventory Revised). In Brazil, the only instance of algorithmic use in sentencing so far is in the State of Minas Gerais, where an algorithm was used to identify the kinds of appeals presented by parties and to provide "model sentences" for review by judges. Other algorithms, however, are frequently used for purposes such as selecting Reporting Justices at the Brazilian Supreme Court. See: Tribunal de Justiça do Estado de Minas Gerais. **TJMG utiliza inteligência artificial em julgamento virtual**. November 07, 2018. Available at: <https://www.tjmg.jus.br/portal-tjmg/noticias/tjmg-utiliza-inteligencia-artificial-em-julgamento-virtual.htm#.XC1Vby2ZO1s>. Access: January 05, 2019. and Supremo Tribunal Federal. **Ministra Carmen Lúcia anuncia início de funcionamento do Projeto Victor, de inteligência artificial**. Notícias STF. August 30, 2018. Available at: < http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=388443>. Access: January 05, 2019.

[5] When it comes to the legality of their adoption, though the law varies from one jurisdiction to the next, democratic nations are unanimous in affirming the requirement of some form of reasoned motivation for court decisions, and that rationale cannot be entirely random, or must at least be sufficiently clear to allow the interested party to appeal it. In Brazil, Article 93, IX of the Constitution says that "all judgments of judicial authorities shall be public, and all decisions shall be motivated". In common law jurisdictions, the duty to give reasons may not exist when it

primarily focuses n the discriminatory risks of the output generated by algorithmic systems.[6] According to Julia Angwin and her team at ProPublica who revealed the use of algorithms in criminal sentencing and its legal and moral implications, COMPAS, the tool developed by the company Northpointe (later renamed Equivant) adopted by the justice system in states such as Florida, "turned up significant racial disparities (…) In forecasting who would re-offend, the algorithm made mistakes with black and white defendants at roughly the same rate but in very different ways". Whereas blacks were falsely flagged as high risk and potential re-offenders at twice the rate as white defendants, whites were more frequently deemed as low risk than black defendants.[7]

Though the use of algorithmic systems for sentencing has yet to reach the same intensity in Brazil as in the United States, discrimination through automation is already part of our reality and, given the extensive use of such tools in other jurisdictions, the debate over them will only

comes to administrative decisions, but judicial duty to give reasons is upheld. In: LO, H. **The Judicial Duty to Give reasons**. Legal Studies, vol. 20, issue 1, March 2000, p. 42 - 65.

[6] Other repercussions include the risks to data privacy and security, cf. section 2.3.2. There is a natural connection between the two, but in protecting individuals from discrimination we may not always be ensuring greater privacy.

[7] ANGWIN, Julia; LARSON, Jeff; MATTU, Surya; KIRCHNER, Lauren. **Machine Bias**. ProPublica, May 23, 2016. Available at: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Access: January 10, 2019., in which they state: "We also turned up significant racial disparities, just as Holder feared. In forecasting who would re-offend, the algorithm made mistakes with black and white defendants at roughly the same rate but in very different ways. The formula was particularly likely to falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants. White defendants were mislabeled as low risk more often than black defendants. Could this disparity be explained by defendants' prior crimes or the type of crimes they were arrested for? No. We ran a statistical test that isolated the effect of race from criminal history and recidivism, as well as from defendants' age and gender. Black defendants were still 77 percent more likely to be pegged as at higher risk of committing a future violent crime and 45 percent more likely to be predicted to commit a future crime of any kind."

grow. Algorithms have been deployed for making diagnoses,[8] job selection,[9] and for even welfare eligibility processes.[10]

The present work surveys the consequences of algorithmic discrimination, as well as the paths for the protection against discriminatory conduct. Four important observations must be made in this introduction: first, as will be further scrutinized throughout this dissertation, discrimination is not the only issue that may arise from algorithmic use. Problems such as privacy infractions may also run from algorithmic use but do not necessarily relate to discrimination, as section 2.3.2 will demonstrate. These problems are not within the scope of my work.

Second, if automation and its potential for discrimination are cause for concern, we must also remember that the world is full of generalizations, and that decision-making largely depends on generalizing to be feasible. The legal system, in particular, is full of necessary generalizations. Whenever we set out a rule, for example one that makes it illegal to drive above 80 km/h on a given road, or when we establish that only those over a given age have the right to vote, we are using, engaging, and reinforcing generalizations. Society as a whole accepts this manner of decision and rule-making, which leads us to suspect that the problem with profiling

---

[8] "Complex algorithms will soon help clinicians make incredibly accurate determinations about our health from large amounts of information, premised on largely unexplainable correlations in that data. […] With extraordinary accuracy, these algorithms were able to predict and diagnose diseases, from cardiovascular illnesses to cancer, and predict related things such as the likelihood of death, the length of hospital stay, and the chance of hospital readmission. Within 24 hours of a patient's hospitalization, for example, the algorithms were able to predict with over 90% accuracy the patient's odds of dying. These predictions, however, were based on patterns in the data that the researchers could not fully explain." In: BURT, A. and VOLCHENBOUM, S. **How Health Care Changes When Algorithms Start Making Diagnoses.** Harvard Business Review, May 08, 2018. Available at: <https://hbr.org/2018/05/how-health-care-changes-when-algorithms-start-making-diagnoses>. Access: January 10, 2019.

[9] "Many companies, including Vodafone and Intel, use a video-interview service called HireVue. Candidates are quizzed while an artificial-intelligence (AI) program analyses their facial expressions (maintaining eye contact with the camera is advisable) and language patterns (sounding confident is the trick). People who wave their arms about or slouch in their seat are likely to fail. Only if they pass that test will the applicants meet some humans." THE ECONOMIST. **How an algorithm may decide your career.** June 21, 2018. Available at: <https://www.economist.com/business/2018/06/21/how-an-algorithm-may-decide-your-career>. Access: January 10, 2019.

[10] Several examples of welfare automation in the United States are investigated in: EUBANKS, V. **Automating Inequality: How High-Tech Tools Profile, Police, and Punish de Poor**. St. Martin's Press. January 23, 2018. Eubanks provides examples on how welfare eligibility in Indiana, the homeless policy in Los Angeles, and others were automated.

and discrimination is not the practice of making generalizations in itself, but rather with the criteria used to create groups that are allowed (or not) to behave a certain way or access a specific good. Therefore, this dissertation takes the view, further scrutinized in section xxx, that discrimination is not always harmful, nor is it always illegal.

Third, the advent of machine learning has brought a number of complexities to this landscape. A system such as AlphaGo, which works with deep neural networks, is very different from a system like COMPAS, whose functioning is based on "old-fashioned" programming. The consequences for discrimination are severe, especially when it comes to identifying solutions. As section 4.1.1 further investigates, asking for transparency of machine learning algorithms is of little use, and accountability proposals are also challenging to implement.

Lastly, it should be clear from the start that this dissertation does not in any way suggest that innovation should be contained, for it always leads to unwanted outcomes. It merely intends to better understand and investigate one potentially harmful outcome, which is quite different. As with most of technology that came before algorithms, implementation always presents pros and cons; the challenge for society is to learn how to strike a balance that allows for the benefits and contains the disadvantages.

For no other reason the discussion is increasingly present in the private sector and also among public authorities. In March 2018, the Council of Europe, through its Committee of experts on Internet Intermediaries, finalized and published a study on the human rights dimensions of automated data processing techniques, focusing primarily on algorithms and possible regulatory implications.[11] The objective of the study is to

> map out some of the main current concern from the Council of Europe's human rights perspective, and to look at possible regulatory options that member states may consider to minimise adverse effects, or to promote good practices.[12]

[11] COUNCIL OF EUROPE PORTAL. **Algorithms and Human Rights: a new study has been published.** March 22, 2018. Available at: <https://www.coe.int/en/web/freedom-expression/-/algorithms-and-human-rights-a-new-study-has-been-published>. Access> January 10, 2019.
[12] Ibid, p. 4.

Similarly, in May 2018, the Science and Technology Committee of the House of Commons in the United Kingdom published a report on decision-making by algorithms, in order to "identify the themes and challenges that the proposed Center for Data Ethics & Innovation [which shall be created by the British government shortly[13]] should address as it begins its work". The report calls attention to the challenges brought about by automated decisions, especially bias, as well as the possible solutions for such hurdles, mainly involving mechanisms for transparency and accountability.[14]

The United States is also taking part in the debate. In 2016, the Federal Trade Commission issued a report on the consequences of Big Data and its inclusionary and exclusionary uses.[15] The FTC states that

> Big data analytics can provide numerous opportunities for improvements in society. In addition to more effectively matching products and services to consumers, big data can create opportunities for low-income and underserved communities. […] At the same time, workshop participants and other have noted how potential inaccuracies and biases might lead to detrimental effects for low-income and underserved populations. For example, participants raised concerns that companies could use big data to exclude low-income and underserved communities from credit and employment opportunities.[16]

Australia was one of the first countries to expressly address the issues brought about by big data, automation, and algorithms. Back in 2007, the Australian government had already issued its Better Practice Guide for automated decision-making in the public administration. The Guide is an effort to consolidate and provide concrete solutions for the best practice principles of the Automated Assistance in Administrative Decision-Making Report issued by the Attorney-General in 2004. It states that "[a]utomated systems have been used for some users in areas of

---

[13] DEPARTMENT FOR DIGITAL, CULTURE, MEDIA & SPORT of the UK Government. **Consultation on the Centre for Data Ethics and Innovation**. November 20, 2018. Available at: <https://www.gov.uk/government/consultations/consultation-on-the-centre-for-data-ethics-and-innovation/centre-for-data-ethics-and-innovation-consultation>. Access: January 10, 2019.

[14] SCIENCE AND TECHNOLOGY COMMITTEE of the House of Commons. **Algorithms in decision-making.** Fourth Report of Session 2017-19. May 23, 2018. Available at: <https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf>.

[15] FEDERAL TRADE COMMISSION. **Big Data, A Tool for Inclusion or Exclusion? – Understanding the Issues**. January, 2016, p. 1. Available at: <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf>. Access: January 01, 2019.

[16] Ibid, p. i.

financial entitlement for citizens and other agency customer groups, such as welfare, family and veteran support benefits."[17]

Brazil, on its part, has so far been less explicit in this debate. The Brazilian Strategy para Digital Transformation issued in 2018 recognizes the relevance of the data-driven economy and sets forth strategies the government should pursue to extract value from this new environment, namely: (i) promote the approval of incentives to attract data centers to Brazil; (ii) improve the National Policy on Government Open Data, and foster the adoption of tools, systems, and processes based on data; (iii) promote cooperation among authorities and the harmonization of normative frameworks regarding data, in order to facilitate the inclusion of Brazilian companies in the global market; (iv) promote cooperation among government representatives, universities, and companies, as to facilitate the knowledge and technology exchange; (v) stimulate the adoption of cloud services as part of the public administration's services and data storage systems; and (vi) evaluate the potential social and economic effects of disruptive technologies, such as artificial intelligence and big data, proposing policies that aim at mitigating their negative effects and simultaneously maximize their positive effects.[18]

It is in this context that this work aims at answering the following question: *What ought be the role of the Law in illuminating and addressing algorithmic discrimination in the Brazilian context?*

Chapter 2 provides the basis upon which this work will be built. It explains in further detail how the emergence of Big Data was crucial in allowing algorithms to evolve. It also highlights what algorithmic discrimination is, and what it is not. It is in this chapter that I propose a typology for algorithmic discrimination, with the goal of contributing to the discussion and hopefully helping bring more clarity to the topic. Chapter 3 delineates the legal framework in which algorithmic discrimination will be discussed. It presents the debate in

---

[17] AUSTRALIAN GOVERNMENT. **Automated Assistance in Administrative Decision-Making – Better Practice Guide.** February, 2007. Available at: <https://www.oaic.gov.au/images/documents/migrated/migrated/betterpracticeguide.pdf>. Access: January 01, 2019.

[18] MCTIC. **Estratégia Brasileira para a Transformação Digital (E-Digital).** Brasília, 2018, p. 64-65. Available at: <http://www.mctic.gov.br/mctic/export/sites/institucional/estrategiadigital.pdf>. Access: January 12, 2019.

constitutional terms, and also in terms of ordinary legislation, focusing primarily (though not solely) in the case of Brazil. It applies the enforcement debate to two concrete cases, one which took place in Poland, regarding unemployment public policy, and the other in Brazil, regarding credit scoring. Chapter 4 is focused on debating policy solutions, by presenting a review of the literature on algorithmic governance as it currently stands, and discussing an agenda for the Brazilian jurisdiction.

2 Algorithms, Models, and Discrimination

This chapter aims to answer two preliminary questions: first I will explain what comprises an algorithmic system and, second, I will show how its use can lead to discriminatory outcomes. The initial challenge will hence be explaining in plain language what an algorithm is and does. My description must (i) sufficiently convey a basic understanding of data science as applied to algorithmic systems, and (ii) contemplate the emergence of Big Data as part of the advances in this science. To reach the second objective of this chapter, I will delve deeper into algorithmic discrimination, clarifying what constitutes discrimination in the algorithmic context according to the literature and identifying aspects that make algorithmic discrimination peculiar. To do so, I (iii) examine what, as applied to algorithms, sets discrimination and generalization apart in the Brazilian legal context, and (iv) advance a tentative typology for algorithmic discrimination, which will later be taken up in my policy proposals to fight discriminatory outcomes.

2.1 The Algorithmic System

Before analyzing a given topic, academic researchers usually being by defining the object under study. In this case, that means explaining precisely what an algorithmic system is – and, perhaps more importantly, what it is not. Instead of offering formal computational definitions of algorithms or entering the discussion of whether or not algorithms can ever be rigorously defined,[19] my approach is inspired by that of Thomas Cormen, author of one of the most famous (and useful) books on the topic. I will present the definition in a less formalized, but hopefully more helpful manner.

[19] See BLASS A. and GUREVICH, Y. **Algorithms: A Quest for Absolute Definitions**. Bulletin of European Association for Theoretical Computer Science. vol. 81, 2003. Available at: <https://web.eecs.umich.edu/~gurevich/Opera/164.pdf>. Access: January 01, 2019. and GUREVICH, Y. **What Is an Algorithm?**. In: BIELIKOVÁ M., FRIEDRICH, G., GOTTLOB, G., KATZENBEISSER, S., TURÁN, G. (eds) *SOFSEM 2012: Theory and Practice of Computer Science*. SOFSEM 2012. Lecture Notes in Computer Science, vol. 7147. Springer, Berlin, Heidelberg, 2012. Available at: <https://www.researchgate.net/publication/221512843_What_Is_an_Algorithm>. Access: January 01, 2019. For a historical overview of algorithms and their definition, see: BULLUNCK, M. **Histories of algorithms: Past, present and future**. Historia Mathematica, Elsevier, 2015, 43 (3), p. 332 - 341.

An algorithm is, first and foremost, a set of steps or instructions to accomplish a task. Whether that task is responding to the query entered by a user into a search engine or brushing your teeth, both can be summarized as an algorithm. In this work, I will limit the use and definition of algorithms to those that are computable – meaning they can be read by computational devices. Cormen argues that the main difference between computable and non-computable algorithms involves in computational intolerance of imprecise data.[20] An algorithm for brushing your teeth illustrates this nicely. Such an algorithm could be expressed as follows: "Get toothbrush. Get toothpaste. Squeeze toothpaste onto toothbrush. Add water. Insert toothbrush into mouth and brush all teeth for 2 minutes. Spit the saliva and toothpaste into the sink. Rinse the mouth."

It is quite clear that even if this set of steps was transformed into code that a computational device could "read", it is unlikely that the device could accomplish the task because some steps are loosely defined and leave a wide margin for interpretation. For example, it is not clear how one should get the toothbrush or what this action entails. Does it mean I have to get close to the toothbrush, or that I should grab it with my hands? It is this room for interpretation that makes such algorithms useless for computers; algorithms must be faultlessly explicit and precise to be understood by a computational device, and that aspect, besides being useful for definitional purposes, is crucial to the debate that is mapped out in this dissertation, for policy proposals on algorithmic governance must never lose sight of that key characteristic.

Yet something more than a formal definition of algorithm is needed for this work. Because algorithms have become a topic of interest for many groups and fields beyond computer science, the term has come to convey a wide array of ideas that require exploration. As Gillespie puts it,

> Perhaps *algorithm* is coming to serve as the name for a particular kind of sociotechnical ensemble, one of a family of systems for knowledge production or decision making: in this one, people, representations, and information are rendered as data, are put into systematic/mathematical relationships with each

[20] "You might be able to tolerate it when an algorithm is imprecisely described, but a computer cannot. (…) So a computer algorithm is a set of steps to accomplish a task that is described precisely enough that a computer can run it." In: CORMEN, T. H., **Algorithms Unlocked**. MIT Press, 2013, p. 1.

> other, and then are assigned value based on calculated assessments about them.[21]

Moreover, Gillespie claims that "[c]onclusions described as having been generated by an algorithm wear a powerful legitimacy, much the way statistical data bolster scientific claims."[22] In a way, algorithms and the results rendered by them confer a particular kind of legitimacy, such that thy are often considered more reliable than conclusions derived by traditionally subjective human analysis and decision-making processes. Algorithms, in this sense, have become a synonym for a higher form of decision-making because they rely on strict procedures or the "formalization of social facts into measurable data," which "distances its human operators from both the point of contact with others and the mantle of responsibility for the intervention they make."[23]

Yet statistical accuracy and mathematical reliability do not suffice on their own to qualify decisions as better or worse. Quantification is surely important and the definition of procedures is often necessary for decision-making to be reliable, but it would be unwise to conclude that subjective human knowledge is therefore useless or of lesser value in terms of understanding and knowledge. Humans strive for objectivity and clarity, and algorithms often embody both aspirations, but even if we accept that some degree of objectivity is essential, we still must determine how objective algorithms can actually be. Now that they play significant functions in our daily lives, as we shall see in the following pages, doubts regarding their capacity for objectiveness have also been expressed.[24]

2.2 The Emergence of Big Data and Machine Learning

Before going into more detail about algorithms, objectivity and discrimination, it is important to describe the path that has led us to this point where the use of such tools is now

[21] GILLESPIE, T. **Chapter 2- Algorithm**. In: PETERS, B. (Ed.). *Digital Keywords: a Vocabulary of information society and culture.* Princeton: Princeton University Press. 2016, p. 22.
[22] Ibid, p. 23-24.
[23] Ibid, p. 26.
[24] More on objectivity an algorithms' limitations in section 2.3.4.1.

embedded into our lives. The process of "datification,"[25] as some authors call it, has become an inescapable part of our lives, bringing increased efficiencies, but also creating unprecedented risk for the misuse of statistical models and data mining.

This story starts with the concept (or idea) of Big Data itself. "Big" might lead the reader to believe that the amount of the data involved is the main aspect that sets Big Data apart. Although size is certainly an important part of the equation, another fundamental operational characteristic of Big Data is "the ability to render into data many aspects of the world that have never been quantified before."[26] As the FTC notes, an easy way to remember the essential aspects of Big Data uses three Vs: volume, velocity, and variety. Volume is the "vast quantity of data that can be gathered and analyzed effectively"[27]; velocity is the "speed with which companies can accumulate, analyze, and use new data"[28]; and variety is "the breadth of data that companies can analyze effectively."[29]

Mayer-Schönberger & Cukier define Big Data in terms of what they believe are its three main characteristics: comprehensiveness, acceptance of less accurate information, and shifting focus away from causality towards correlation. In their view, until recently most human empirical research had to rely on sample strength for its conclusions. Sampling involves significant problems both in terms of collecting and processing information, often because no tools capable of capturing vast amounts of information were available, or because the tools available could not analyze such immense datasets. The information available, consequently, was not comprehensive. The solution to this problem was relying on samples of the phenomenon one intended to study. As the authors put it, "sampling was a solution to the problem of

---

[25] MAYER-SCHÖNBERGER, V. and CUKIER, K. **The Rise of Big Data: How It's Changing the Way We Think**. Foreign Affairs, vol. 92, no. 3, May/June, 2013, p. 28-40. Available at: <https://www.jstor.org/stable/23526834?seq=1#page_scan_tab_contents>. Access: January 01, 2019.
[26] Ibid, p. 29.
[27] FEDERAL TRADE COMMISSION. **Big Data, A Tool for Inclusion or Exclusion? – Understanding the Issues**. January, 2016, p. 1. Available at: <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf>. Access: January 01, 2019.
[28] Ibid, p. 2.
[29] Ibidem.

information overload in an earlier age, when the collection and analysis of data was very hard to do."[30]

Scientists soon understood the advantage of random sampling, and started focusing on building stronger sample sets through randomness rather than simply increasing the sample size. Yet random sampling has its limits as well, and proved to be a second-best alternative to analyzing the entirety of data. For Mayer-Schönberger & Cukier, the main limitation of random sampling – the one that most clearly contrasts with Big Data – is that it still requires the researcher to precisely define the goal of the research and then identify the variables that must to be randomized in a sample for it to produce statistically valid results.

To explain it using an example, if one wanted to verify the effectiveness of a program that provides microloans to women in a specific population in the interest of improving their families' economic standing, an entire series of precautions must be taken.[31] Control and treatment groups would need to be established; factors other than the microcredit that could also improve economic standing would have to be controlled for; measurements would have to be made and applied consistently over a considerable amount of time; and so on. Big Data circumvents all of this. First, there is no need for a predefined problem, for the logic followed is reversed; rather than collecting a dataset based on a question that needs answering, you look blankly at the dataset in search of patterns the provide information. Second, obtaining comprehensive information is now feasible. As Mayer-Schönberger & Cukier say:

> Using all the data makes it possible to spot connections and details that are otherwise cloaked in the vastness of the information. […] An investigation using big data is almost like a fishing expedition: it is unclear at the outset not only whether one will catch anything but *what* one may catch.[32]

Intrinsically connected to comprehensiveness is inaccuracy. Once we start working with a dataset of *n = all* (or something approaching that) with the tools currently available to assemble

---

[30] MAYER-SCHÖNBERGER, V. and CUKIER, K. **Big Data: A Revolution That Will Transform How We Live, Work, Think**. Houghton Mifflin Harcourt, 2013, p. 23.
[31] BANERJEE, A. et al. **The miracle of microfinance? Evidence from a randomized evaluation**. Northwestern University of Economics and NBER. March, 2014. Available at: <https://economics.mit.edu/files/5993>. Access: January 01, 2019.
[32] MAYER-SCHÖNBERGER, V. and CUKIER, K. **Big Data: A Revolution That Will Transform How We Live, Work, Think**. Houghton Mifflin Harcourt, 2013, p. 27-29.

and analyze such datasets, the chances that the dataset contains inexact information increases. It is likely that a dataset of all of Facebook users' addresses contains a significant number of mistakes. When using the logic of randomized sampling, such inaccuracy is simply unacceptable for carefully curated treatment and control groups, where even one inaccurate piece of information could have a significant impact, because the representativeness or weight of the information is much larger. In the world of Big Data, however, that is not the case. Even if 50 million Facebook user addresses are incorrect, because the dataset comprises approximately 2.23 billion users,[33] the distortion totals less than 3%. Mayer-Schönberger & Cukier emphasize that:

> It is a trade-off. In return for relaxing the standards of allowable errors, one can get ahold of much more data. It isn't just that 'more trumps some', but that, in fact, sometimes 'more trumps better'. (…) Any particular reading may be incorrect, but the aggregate of many readings will provide a more comprehensive picture.[34]

The most important consequence of this dynamic is its heavy reliance on probability. Since constructing massive datasets is now possible, and because these datasets are so comprehensive, one can collect information without starting with a precise question in mind, and thus look differently at relationships between all the variables, even between ones that seem entirely unrelated.

Until recently, problems were traditionally approached by formulating and testing theories. As Kellstedt & Whitten put it, scientific analysis, especially in the so-called exact sciences, usually starts with a causal theory.[35] What scientists do is create a hypothesis that proves or disproves their causal theory. They then move on to create empirical tests – often

---

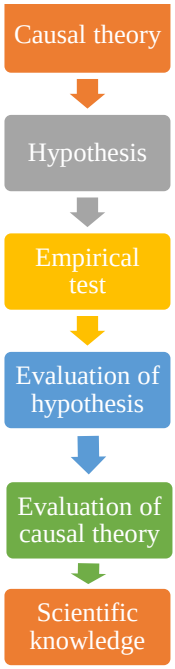
[33] Estimated number of Facebook users as of August 2018.

[34] MAYER-SCHÖNBERGER, V. and CUKIER, K. **Big Data: A Revolution That Will Transform How We Live, Work, Think**. Houghton Mifflin Harcourt, 2013, p. 33-34.

[35] "What do political scientists do and what makes them scientists? A simple answer to this question is that, as other scientists, political scientists develop and test theories. A theory is an attempt to conjecture about the causes of a phenomenon of interest. The development of causal theories about the political world requires thinking of familiar phenomena in new light. Thus, the building of a theory is part art and part science." In: KELLSTEDT, P. M. and WHITTEN, G. **The Fundamentals of Political Research**. New York: Cambridge University Press, 2009.

based on randomized samples – to verify the validity of their hypothesis. This leads to conclusions not only about the hypothesis, but sometimes also about the causal theory itself.

Figure 1 – The path of scientific research



Source: Kellstedt & Whitten (2008).

Big Data turns this logic upside down by moving away from causality towards correlation. Rather than focusing on *why* something happens in certain conditions – the basis for the scheme put forward by Kellstedt & Whitten – the focus is on observing *what* happens. In observing a gigantic dataset of various types of information, what shines through is not causality, but rather correlation, the quantification of a statistical relationship between two values. Let us imagine we are able to observe all the internet search queries made in by a population of 100 million people over the period of a year. Because we are not sure what information we will be able to extract from all these data points, intuitively it may be more productive to focus on the correlation between values in the dataset rather than on a specific question. In doing so, we may discover a relationship – whose cause we were not previously interested in, such as, for example, a correlation between search queries about flu symptoms, and its occurrence.[36] We understand that there is no real causal relation between typing down a symptom of the flu and effectively contracting the disease. Still, the information is relevant because it indicates that flu symptoms search queries may be a good proxy for the flu. That is why Mayer-Schönberer & Cukier emphasize that "[c]orrelations let us analyze a phenomenon not by shedding lights on its inner workings but by identifying a useful proxy for it."[37]

[36] Which is what Google did in establishing the Google Flu mechanism, as will be further scrutinized in section 2.3.4.

[37] MAYER-SCHÖNBERGER, V. and CUKIER, K. **Big Data: A Revolution That Will Transform How We Live, Work, Think**. Houghton Mifflin Harcourt, 2013, p. 53.

A proxy in this case, just as proxies in the perhaps more familiar legal context, refers to someone (or, in this case, some piece of information) to represent someone or something else. Finding and using proxies allows us to make better-informed predictions. Turning back to our previous example, one may be able to anticipate the rates of flu infestation based on search queries about flu symptoms – and search queries are much easier to observe and capture than the actual spread of the illness. Naturally, this result is essentially probabilistic. Because there is no causal relationship between the variables, the chances that they vary at similar rates can only be expressed as a probability, based on previous observations, but not with any certainty.

Big Data, therefore, represents a game-changer. In Boyd and Crawford's words,

> Big Data is notable not because of its size, but because of its relationality to other data. Due to efforts to mine and aggregate data, Big Data is fundamentally networked. Its value comes from the patterns that can be derived by making connections between pieces of data, about an individual, about individuals in relation to others, about groups of people, or simply about the structure of information itself.[38]

The source of such information can be anywhere or anything, and so too its object.

Another decisive technological development that is particularly pertinent to this study is the evolution of machine learning. Machine learning is a branch within the broader field of artificial intelligence, or AI. Whereas AI is primarily interested in bringing "intelligence," understood broadly, to machines, machine learning is the part of this field that proceeds by giving machines equipped with advanced software and processors access to data to see what conclusions and adaptations based on experience the machines make themselves. Clearly the rise of Big Data has also allowed machine learning to flourish as well: the increased data available is part of the reason that the learning capacity of machines has increased exponentially.

Pedro Domingos has explained why machine learning is so revolutionary for computer science: "Traditionally, the only way to get a computer to do something – adding two numbers to flying an airplane – was to write down an algorithm explaining how, in painstaking detail."

[38] BOYD, D. and CRAWFORD, K. **Six Provocations for Big Data.** A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, September 2011. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431>. Access: January 01, 2019.

As clarified earlier, algorithms have little tolerance to imprecise statements, hence thorough detail is an important part of coding.

> But machine-learning algorithms, also known as learners, are different: they figure out on their own, by making inferences from data. And the more data they have, the better they get. Now we don't have to program computers; they program themselves.[39]

Domingos goes on to compare data and correlations in machine learning to the bricks in a house, the necessary but not sufficient element upon which "learner" algorithms are built. The particularity of learners is that they not only rely on programs to draw correlations from data, they also adapt their programming to find other correlations and patterns without explicit instructions on what to look for and how. Earlier, brushing your teeth was expressed as an example of an algorithm. To transform the action into a computable algorithm, one would need to create many lines of code detailed enough for a computer to comprehend and process. Learner algorithms need much less effort. They could access the vast amount of information publicly available, such as personal videos of different people brushing their teeth in different circumstances, and from those data points work inductively to learn by itself the pattern of "brushing your teeth". The amount of prior human interference and programming, therefore, could be potentially limited to providing the desired dataset.

2.2.1 The risk of discrimination

Although many advantageous effects have been observed, the use of Big Data and machine learning can also lead to problems involving the main topic of this dissertation: discriminatory or discriminatory-like practices. Barocas & Selbst lay out the problem as follows:

> Approached without care, data mining can reproduce existing patterns of discrimination, inherit the prejudice of prior decision makers, or simply reflect the widespread biases that persist in society. It can even have the perverse result of exacerbating existing inequalities by suggesting that historically disadvantaged groups actually deserve less favorable treatment.[40]

As the authors point out, discriminatory outcomes may arise not only from intentional discrimination, but also, and perhaps primarily, due to unconscious or unpredicted biases. They

---

[39] DOMINGOS, P. **Master Algorithm**. Basic Books Inc. New York, 2018. p. xi.
[40] BAROCAS, S. and SELBST, A. D. **Big Data's Disparate Impact.** p. California Law Review, vol. 671, 2016, p. 674. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899>. Access: January 01, 2019.

go on to explain why this is so, focusing on four common steps followed in Big Data and machine learning processing: (i) the definition of target variables and class labels; (ii) the determination of training data; (iii) feature selection; and (iv) the use of proxies. Because the use of Big Data combined with machine learning "automates the process of discovering useful patterns,"[41] for these mechanisms to function in ways that serve the needs and desires of humans, target variables must be established, that is, the goal or target towards which the learner is oriented. Barocas & Selbst also state that translating a real-world problem into a computable definition is the most complex part of data mining and call attention to the fact that "through this necessarily subjective process of translation, data miners may unintentionally parse the problem in such a way that happens to systematically disadvantage protected classes."[42]

In these processes, one must also rely on class variables; that is, the values that the target variable may come to represent. The more complex the problem tackled, the more difficulty defining these variables will be. If the question can be translated into a binary equation (yes or no, hot or cold, up or down), the job tends to be simpler. The authors differentiate between simpler and more complex scenarios by comparing the algorithms responsible for flagging spam emails to algorithms used to assign credit scores. Whereas a message is either spam or is not – and therefore the class variables in this case are only two and only two mutually exclusive categories – determining creditworthiness is far more complicated, for it cannot be directly measured nor easily translated into mutually exclusive categories.[43] The same would hold for algorithms aimed at finding the "best" candidate for a job, or the "ideal" movie for a given user of a streaming service. In both cases, there is no readily available definition of "best" or "ideal," and "danger resides in the definition of the class label itself and the subsequent labeling of examples from which rules are inferred."[44] If the definition of the "best" candidate somehow

[41] Ibid, p. 677.
[42] Ibid, p. 678.
[43] "There is no way to directly measure creditworthiness because the very notion of creditworthiness is a function of the particular way the credit industry has constructed the credit issuing and repayment system. That is, an individual's ability to repay some minimum amount of an outstanding debt on a monthly basis is taken to be a nonarbitrary standard by which to determine in advance and all-at-once whether he is worthy of credit." In: BAROCAS, S. and SELBST, A. D. **Big Data's Disparate Impact.** p. California Law Review, vol. 671, 2016, p. 679. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899>. Access: January 01, 2019.
[44] Ibidem.

disproportionally affects individuals in a certain group, the algorithm will likely produce discriminatory outputs.

Additionally, because algorithms learn by aggregating data, another decisive aspect of their reliability and potential for discrimination is the training data they are exposed to. If the dataset on brushing teeth presented to the learner is insufficient, it will not perform the assigned task as well as it should. Similarly, if the data is biased, the learner will identify and reproduce biased patterns. Barocas & Selbst are careful in pointing out that this danger can arise in two different regards:

> (1) if data mining treats cases in which prejudice has played some role as valid examples to learn from, that rule may simply reproduce the prejudice involved in these earlier cases; or (2) if data mining draws inferences from a biased sample of the population, any decision that rests on these inferences may systematically disadvantage those who are under- or overrepresented in the dataset.[45]

These risks play out in the assignment of class labels to examples, especially when the assignment or weighting is incorrectly calculated. Going back to credit scores and evaluating candidates for jobs, should a person who missed 4 payments in the last 5 months be passed over even if she has never missed a payment before? What about someone who has no history of tardiness and has never missed a day of work, but has recently been diagnosed with a chronic disease? Labelling becomes paramount, for the training data provides the "ground truth" for learners. If carried out carelessly, labelling will not only skew results, but also create mistakes that are hard to locate subsequently and therefore to correct – turning again to emails and spam messages, if an email from a given sender is incorrectly labelled spam, the problem will only be fixed if the user identifies the error and manually corrects it. The algorithm will not learn the mistake by itself. If in the case of spam the problem seems benign (though I do not want to downplay the embarrassment of missing important messages because of incorrect labelling), the matter is far more critical when economic livelihood (obtaining a loan or job) is at stake.

Training data also poses data collection issues. If the dataset misrepresents a certain group of individuals, or possesses wrong data for one specific individual, it can

[45] Ibid, p. 681.

disproportionately affect the group or person. It is noteworthy that some populations and individuals remain largely outside the reach of Big Data; that is, their lifestyles do not generate the type and volume of data mostly captured by this phenomenon. As Jonas Lerman puts it:

> Big data poses risks also to those persons who are *not* swallowed up by it—whose information is not regularly harvested, farmed, or mined. (Pick your anachronistic metaphor.) Although proponents and skeptics alike tend to view this revolution as totalizing and universal, the reality is that billions of people remain on its margins because they do not routinely engage in activities that big data and advanced analytics are designed to capture.[46]

Lerman questions the validity of the assertion by Mayer-Schönberger & Cukier that in the world of Big Data, *n = all*. He highlights that even in cases where this is almost true, the excluded population will disproportionately suffer the effects of being left out. It should not come as a shock that the populations most affected by this problem tend to be those already historically disadvantaged. Lack of access to the internet, for example, is one of the many factors that contribute to this distortion, as are access and involvement in the formal economy. Again, considering the use of learners to deem creditworthiness, one could anticipate that very poor people will likely have difficulty obtaining loans because they have less access to financial institutions (many do not have bank accounts, credit cards and so on). They might not be turned down because the data reveals them to be bad at repaying debt, but simply because the financial institution does not have enough information on them to ascertain creditworthiness.

It is also important to note that this problem may present itself when the population in question is not completely absent from the pool of data, but is under- or overrepresented. Barocas & Selbst point out that "to ensure data mining reveals patterns that hold true for more than the particular sample under analysis, the sample must be proportionately representative of the entire population, even though the sample, by definition, does not include every case."[47] This statement throws the validity of Mayer-Schönberger & Cukier's earlier affirmations back into question, for it shows the need to rely on "old-fashioned" methodology such as sampling for representative and truthful patterns. An example that clearly reveals the problem involves

[46] LERMAN, J. **Big Data and Its Exclusions**. Stanford Law Review Online, 2013, p. 56.
[47] BAROCAS, S. and SELBST, A. D. **Big Data's Disparate Impact.** p. California Law Review, vol. 671, 2016, p. 686. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899>. Access: January 01, 2019.

incarceration. Blacks are severely overrepresented in the incarcerated community in the United States and in Brazil. If the police behavior on the streets is based on the incorrect belief that black citizens are proportionately more likely to commit crimes, the police will be more sensitized to the perceived risk and likely find and arrest more of them in a dynamic similar to a self-fulfilling prophecy. Yet there is no causal relation between race and proclivity toward crime, and using race as a predictor will only aggravate the sample bias. No one doubts that white commit crimes, but it is sometimes forgotten that when police enforcement is concentrated in black neighborhoods, it will naturally find more black criminals.

Another critical step is feature selection. Selecting features means deciding what will ultimately be part of the predictor model and what will not. Barocas & Selbst observe that groups or individuals that are not well represented by the selected features may be affected in this phase. In their words,

> members of protected classes may find that they are subject to systematically less accurate classifications or predictions because the details necessary to achieve equally accurate determinations reside at a level of granularity and coverage that the selected features fail to achieve.[48]

Once more, this is a problem that could be solved by the traditional methodological safeguards for sampling, which ensure equal representation of individuals and groups in the samples. Such controls are not frequently carried out, however, because they are not cost efficient. Decision makers may accept the tradeoff of lower accuracy for lower costs, and, again, the most affected groups tend to be those who have already historically suffered discrimination.

Lastly, algorithms rely on proxies. The use of proxies is not a problem in and of itself; quite on the contrary, proxies are an efficient and often necessary mechanism which allow us to use an easily observable characteristic as a predictor for another phenomenon that cannot be as easily measured. The risk for discrimination arises, however, "when the criteria that are genuinely relevant in making rational and well-informed decisions also happen to serve as reliable proxies for class membership."[49] The problem is technically described as redundant

[48] Ibid, p. 688.
[49] Ibid, p. 691.

encoding, which means using more than one visual characteristic to represent a variety of data. If race or gender are features that strongly correlate to the relevant characteristic one intends to observe, it is natural for people pertaining to that race or gender to be disproportionally affected, even if race or gender are not in themselves what one sets out to observe. For instance, it may be fair to assume that if one is looking for suitable candidates for a position, one could use the university where they graduated from as an indicator of education quality. If, for whatever reason, the schools considered the best by the examiner also have greater numbers of white students enrolled, redundant encoding will result: race will be encoded into schooling background. If this process is automated and the algorithm fails to detect the problem, which is likely, it may use race as a proxy for schooling background whenever that background is not in the dataset, and thus disadvantage black people.

This is one reason that caution must be taken to avoid so-called spurious correlations. Turning back to a previous example, just as search queries may show a correlation between the occurrence of the flu and the typing of flu symptoms into a search website, algorithms could also determine the correlation between the consumption of cheese and the number of people who died by becoming tangled in their bedsheets.[50]

It seems intuitive that this later case is somewhat different from the previous one, but why? Statistically speaking, there is no way of telling the real difference between these two examples unless we move on to analyzing causal effects. Formally, a correlation is spurious whenever two elements are not related, but seem to be, either due to coincidence or to the so-called common response variable (a third variable that interferes with both others and causes them to vary correspondingly, but remains hidden and therefore causes the confusion). In other words, unless we analyze causality, it is impossible to tell if the correlation is spurious.[51]

---

[50] As exemplified in: VIGEN, T. **Spurious correlations.** Available at: <http://www.tylervigen.com/spurious-correlations>. Access: January 01, 2019.

[51] In Simon's words, "To test whether a correlation between two variables is genuine or spurious, additional variables and equations must be introduced, and sufficient assumptions must be made to identify the parameters of this wider system. If the two original variables are causally related in the wider system, the correlation is "genuine." In: SIMON, H. A. **Spurious Correlation: A Causal Interpretation**. Journal of the American Statistical Association, vol. 49, September 1954, p. 467. Available at: <http://digitalcollections.library.cmu.edu/awweb/awarchive?type=file&item=33513>. Access: January 01, 2019.

There are, however, ways to verify if a correlation is spurious without engaging in causal analysis. One of them is to expand the dataset. Where it is somehow possible to observe all occurrences of two given variables, and in all such occurrences they present a correlation, it is less likely for the correlation to be spurious. Spuriousness nevertheless remains a possibility, and always will be unless causal analysis is carried out, as the case of schooling background exemplifies. Access to university is likely to remain a problem in larger sets of data, because the access of blacks to universities is, at least in Brazil, a social problem. Big Data, being primarily concerned with correlations, fails to provide strong causal explanations. As it is part of human nature to look for causation and to assume it exists whenever correlation is present, one must be vigilantly aware of what Big Data can and cannot offer, for otherwise we run the risk of taking the information as something it simply is not.

2.3 Discrimination and Profiling - Generalizations under the law

The final step before delving deeper into a typology for algorithmic discrimination is to briefly establish the difference between profiling (or any form of generalization) and discrimination. This distinction is necessary for two reasons. First, because it must be clear for the reader that generalizations are part of our everyday lives, and have been long before the emergence of the data-driven economy. Second, because it will narrow the scope of my work by illustrating why, although profiling may pose many legal controversies, it is not always discrimination and, even when controversy exists, it is not always related to discrimination. Often, profiling can be a practice that somehow impairs other rights, such as privacy, which falls outside the focus of this dissertation.

2.3.1 Striving for Particularization - Is individualized decision-making superior?

It is not a given that discriminating is inherently bad. We constantly discriminate in our everyday lives, for discriminating is one of the means by which decision-making in a context of imperfect information becomes feasible. In a very broad sense, discrimination can be understood as any practice that associates things, people, or situations on the basis of certain characteristics that we are interested in observing or in taking into consideration. When we decide to buy our children a cat instead of a dog, we may be basing our decision on a myriad of factors, and may

discriminate by considering a "cat" part of a category of animals that requires less care than "dogs," though in practice it may turn out that certain cats are far more trouble than certain dogs. The same holds true for our trust in standardized tests as reliable measures of children's aptitudes. Children who score highly become part of a group that is considered to have developed the abilities needed to progress in their education, while children with low scores are not deemed to have mastered those skills. Although in the aggregate this discriminatory judgment may be true in most cases – and that is why we resort to standardized testing – it does not hold true in all scenarios and for all children. Some students who are not fully equipped to move on pass the test, and some who are ready do not.

But why, then, do we discriminate, when we know the result of such discrimination is imperfect? Simply put, we face a trade-off between efficiency and accuracy: we live in a world of imperfect information where there is no way of knowing whether the particular cat that we would like to adopt will be any less trouble than a particular dog. To know for sure, we would need to take both home, observe them for some time, and then make a decision. Still, and this is of fundamental importance, even after the trial period, *the decision would still involve generalization*. It would be based on a generalization about the attitude of that given animal and the likelihood that the behavior observed for a short time is an accurate predictor of the animal's future behavior.

The law, it must be highlighted, works in much the same manner. It is constantly generalizing and creating categories, which often times lead to discriminations of some sort. The most common examples are that of the minimum age for driving, drinking, and voting. Legislation in most countries establishes the age of 18 years-old as the threshold for several activities, Brazil included. No one aged 17 or less can legally drive a car or drink alcoholic beverages. And no one aged 15 or less can vote – voting is optional for those between 16 and 18 years-old, but becomes mandatory thereafter. One could reasonably question these rules, claiming they are unfair generalizations because many people at the age of 17 are entirely fit to drive a car, and many above 18 are unable to do so. The same goes for voting and drinking. The observation is accurate, but it is nonetheless how most countries have decided to regulate these

matters. Usually there are no major protests or popular outbursts claiming it is unfair to discriminate on the basis of age in order to determine who is lawfully able to drive a car.

This leads us to believe that there is no fundamental problem with generalizations themselves, but rather with certain uses of generalizations, and with the content of what is generalized. Society understands and accepts the necessity to make do with a lesser degree of accuracy in a great many situations in order to turn decision-making into a feasible effort. Notwithstanding, there are also many cases where we do not welcome the trade-off of certainty for efficacy, and believe individualized case-by-case decision-making is a better (or fairer) option.

There is a wide-spread notion that judging people on their individual merits and characteristics is fairer than generalizing, and this idea does indeed have important grounds. It largely runs from (i) our correct understanding that a person should never be judged by someone else's deeds, as the Brazilian Constitution states,[52] but rather given her own actions or omissions and (ii) our belief that individualization is a superior form of assessment, and should be used whenever possible, leaving generalization as a second-best alternative to less relevant scenarios.

In regard to (i), it is worth stressing that individualization of punishment and sentencing should indeed remain as far away as possible from any attempts of generalization. Naturally, it would not only be wrong but also illegal to send someone to prison based on the notion that the group to which that specific person pertains is more prone to committing crimes, even if that person has never been involved in any act of the sort, and it would be equally problematic to abstain from imprisoning someone because they belong to a group deemed less prone to committing the same crimes. Still, in discussing these scenarios we are talking about a very particular field of law – criminal law – that not only requires the highest degree of proof, precisely due to the seriousness of the sanctions, but also the highest degree of individualization.

[52] The original in Portuguese reads: Artigo 5º, XLV: "nenhuma pena passará da pessoa do condenado, podendo a obrigação de reparar o dano e a decretação do perdimento de bens ser, nos termos da lei, estendidas aos sucessores e contra eles executadas, até o limite do valor do patrimônio transferido".

There is a myriad of cases where the bar against generalization is much lower, as illustrated previously by the minimum ages for drinking, driving and voting.

In regard to society's belief that case-by-case decision-making is more accurate, the situation is quite different. As mentioned, there is a natural tendency to believe any decision based on individualization is superior to judgments based on some form of generalization. Frederick Schauer explains that this understanding is often misguided, not because it is inconsistent with the value of individualization, but because the process by which one reaches an individualized decision is usually flawed. He uses the example of a traffic accident to clarify his point: one car is hit by another vehicle at an intersection but the other vehicle does not stop. It is not clear what vehicle hit the car, but the car's driver alleges that she saw it was a bus. All buses in the city belong to the same company, and this company contests the driver's allegation. Since there was a passenger in the car at the time of the accident, he is called to testify on the driver's behalf. It turns out that Mr. Wilson – the passenger – is blind. He testifies to the court that the other vehicle certainly sounded like a bus, and is corroborated by expert witnesses.

Schauer makes an important observation about this example:

> [I]n considering what to make of Wilson's perceptions, we would naturally think that the validity of these perceptions depends on a process of generalizations and noncertain inference. Wilson has perceived some sounds in the past, and they have turned out to be buses. (…) As a result, Wilson's inference from this sound to this conclusion (it is a bus at this distance) is an inference based on most but not necessarily all sounds of this type's having turned out in the past to be buses. This is a nonspurious but nonuniversal generalization – most but not all sounds like this are buses – that undergirds what appears to be a direct and thus individualized perception.[53]

Schauer's point is that we often lose sight of the fact that much of what we consider to be individualized decision-making is in fact based on generalizations from past experiences, just as in the case of Mr. Wilson, and therefore often omit several relevant variables that could otherwise be taken into consideration. He therefore concludes that our preference for this kind of decision-making is based on two mistakes:

> an overconfidence in the empirical reliability and even the very directness of direct evidence, and an underappreciation of the essential continuity between

[53] SCHAUER, F. **Profiles, Probabilities, and Stereotypes**. Belknapp Press. April, 2016, p. 102-103.

> so-called indirect or statistical evidence and evidence that that on its face appears to be more individualized and thus less statistical.[54]

This dissertation does not contest that if individualization was indeed possible and if it was not riddled with the defects pointed out by Schauer, it would be a superior form of decision-making. However, because it is not (and cannot be), generalizing (and thus discriminating) is a valid form of decision-making. The problem lies primarily on the *substance* of the generalization, not in the process of generalizing itself.[55]

If my premise holds, the next step is examining the areas where generalizing is acceptable. To do so, I turn to another distinction made by Schauer between rational and irrational generalizations. The distinction is important because irrational generalizations are rejected as unacceptable,[56] whereas rational generalizations may be permitted. There are two forms of irrational generalizations: (i) empirically sound generalizations about immaterial traits – the type that "reliably predicts something in which we have no interest in,"[57] and (ii) empirically unsound generalizations about material traits, which are "is aimed at something we are indeed concerned about but has no tendency to indicate or predict it."[58]

True correlations exist for the irrational generalizations of type (i), but that correlation is useless for the purposes of the intended analysis. There is a correlation between body height and age, for example, but it would be irrational to justify any rule establishing a minimum age for voting on the basis of a desire to exclude short voters. Type (ii) irrationality can be observed when the trait used to analyze a given variable is arbitrarily chosen. When establishing the minimum age for voting, one could say that taller people tend to be better prepared to vote, because they are also older and more experienced. This does not make it acceptable to use height as a criterion to vote, however, because the trait "height" has no bearing on one's ability to make informed decisions.

---

[54] Ibid, p. 106.
[55] The process will be relevant inasmuch as a generalization is reached by use of irrational methods. It seems clear that if something is irrational, it cannot be accepted as a valid conclusion.
[56] As mentioned, though the substance of generalizing leads to the more contentious debate, process will be relevant if it results in irrational conclusions.
[57] Ibid, p. 133.
[58] Ibidem.

After excluding such irrational uses of discrimination as never permissible, we are left with rational statistical discriminations, which involves statistically sound and measure material traits. To the question if all statistically rational forms of discrimination will be allowed, the answer is still no. Schauer explains that "even a genuine statistical correlation […], rather than being seen as a *justification* for discrimination, might better be understood as a *product* of discrimination."[59] In other words, if faced with a statistically rational generalization, one must analyze whether the empirical soundness of the observation results from the generalization, in which case it will not be permissible.

Striving for individual and particularized decision-making is natural and responds to ingrained principles of our legal systems and other entrenched social values. It would be, however, naive to believe the methods by which so-called individualized decision-making is performed are far superior to actuarial methods and generalizations. That is why, instead of focusing on the process of generalizing and profiling as a problem in itself, one should emphasize the substance of the generalization. That is the aim of the rest of this chapter. Because I am particularly concerned with one of the consequences of generalization – discrimination – I will first briefly go over some of the other issues that may arise from profiling, for it may be useful to clarify some of the other consequences of this process.

2.3.2 Profiling and Other Legal Matters

It is rather intuitive that profiling or generalizing may, albeit not necessarily, be harmful.[60] Still, the process of generalizing may be harmful not because it violates the right to equality but because it infringes on other rights. If a search engine creates a profile to track my behavior and direct advertisements to me, and for whatever reason its algorithms come to the conclusion I enjoy spending my holidays at the beach, as long as the only outcome is an increased number of advertisements for trips to Rio de Janeiro that show up on my browser, the claim that a right is being violated is hard to sustain. After all, what do the adds prevent me from

[59] Ibid, p. 139.
[60] As the previous section has showed, however, this is not a determining factor in defining generalizations.

accomplishing? Seeing other advertisements? Hardly. Am I in any way worse off than before? It is a difficult claim to make.

Even if this discriminatory practice does not violate equality standards, it may turn out to be problematic for other reasons.[61] Maybe I was not aware that data about my preferences was being collected. Maybe I never consented to it. Or maybe I did, but never for its use for advertisement purposes. This may be a relevant legal issue, but it is not related to the substance of generalization. Similar scenarios should be addressed by personal data protection legislation. Naturally, if personal data regulation is robust, protection against algorithmic discrimination tends to be less urgent. Take the infamous case of Cambridge Analytica as an example.

Cambridge Analytica was accused of harvesting data from Facebook users – in many instances without consent – and deploying it for political use, most notably by U.S. presidential campaign of Donald Trump.[62] The public discussion focused on whether or not Facebook took appropriate measures to counter Analytica's conduct, but the first problem with the case is not discrimination, but rather privacy. Did users consent to the use of their data by the company? The claim is that Analytica not only harvested data from those who explicitly agreed to the terms and conditions of its app, but also from any and all Facebook friends of such users, who never expressed their interest or their consent. Although questions regarding discrimination came up subsequently when asking how Analytica used the data and whether profiles were created on the basis of race, gender, ideology and so forth, these questions are in some ways independent from the immediate privacy concerns, even if strong personal data and privacy legislation would result in fewer cases of algorithmic discrimination. Still, however clear the connection between algorithmic discrimination and personal data/privacy, the two are quite distinct. The distinction

[61] The Study on the Human Rights Dimensions of Automated Data Processing Techniques (in particular algorithms) and possible regulatory implications, by the Council of Europe, highlights impacts that may arise from algorithmic use in regards to due process, privacy, freedom of expression, freedom of assembly and association, the right to free elections and so forth. See in: COUNCIL OF EUROPE PORTAL. **Algorithms and Human Rights: a new study has been published.** March 22, 2018. Available at: <https://www.coe.int/en/web/freedom-expression/-/algorithms-and-human-rights-a-new-study-has-been-published>. Access> January 10, 2019.

[62] THE GUARDIAN. **The Cambridge Analytica Files – A year-long investigation into Facebook, data, and influencing elections in the digital age.** Available at: <https://www.theguardian.com/news/series/cambridge-analytica-files>. Access: January 01, 2019.

also gains added relevance when, as will be discussed in chapter 3, differentiating between discrimination and other legal matters is essential in discussing enforcement.

2.3.3 Algorithmic Discrimination: A Proposed Typology

Since generalization is not in itself a problematic practice, my objective in this section is to propose a typology of algorithmic discriminatory practices, not merely for the sake of classification, but because such categories, if well developed, may be useful for understanding, criticizing, and suggesting policy proposals to deal with the problem. It should also be noted that this typology is informed by the Brazilian legal system and the legal instruments available in this jurisdiction, although my intention is to provide a useful framework for other jurisdictions, at least as a point of reference.

The first step is once again emphasizing that discrimination owing to algorithmic use is based on statistics. Statistics, for their part, may render sound or unsound results. Any form of discrimination that is not statistically sound should always be considered unlawful, as stated in more detail in section 2.3.1. The question of legality remains open, however, with regards to statistically sound generalizations, for they may or may not be lawful. The typology that follows intends to identify (a) when cases of unsoundness arise, and (b) cases of unjustified sound generalizations.

a) Scenarios of unsoundness

Two premises should be clarified before taking up the scenarios that follow. First, all categories of discrimination considered unsound are taken as unlawful for the purposes of this dissertation, even though Brazilian legislation (and legislation in other jurisdictions as well) is not always explicit about their illegality.[63] Second, the generalizations set out by the algorithmic system in these categories lead to discrimination either because of empirical inexactitude or incorrect identification of the relevant traits. In these scenarios what is questionable is the model, the dataset, or the methods by which inferences are draw from the combination of algorithm and data.

[63] I will delve deeper into this point in chapters 3 and 4.

(i) Discrimination by faulty collection or design

One form of statistical unsoundness arises from any and all mistakes that represent failures in the collection of data or in the design of the algorithm on the part of the engineers. This category is therefore not concerned with verifying whether the algorithm or the dataset are biased, it is mostly concerned with empirical weaknesses, which may be either:

- *Data-bound* – mistakes in the data that is captured or used by the algorithmic system. Incorrect or outdated data are the most common sources of such problems; or
- *Algorithm-bound* – errors in the algorithm itself. These can include faulty coding, unintentional failures to account for part of the database, and so on. If an algorithm somehow ignores data due to an engineering mistake, such as failure to include relevant information in its analysis, unsoundness arises.

Data-bound discrimination by design can be verified in cases such as the "No Fly" computer matching system in the United States. The Terrorist Screening Center (TSC) is responsible for keeping a list of people who are prohibited from coming aboard commercial aircrafts with the goal of preventing terrorist attacks. As Citron notes, however,

> over half of the tens of thousands of matches sent to TSC between 2003 and January 2006 were misidentifications. These mistakes stem from faulty information stored in the "No Fly" databases.[64]

Algorithm-bound discrimination, on the other hand, can be illustrated by the problems faced by the Colorado Benefits Management System (CBMS) in the United States. Citron mentions that one of such problems was programmers' failure to correctly translate policy into computer code. For instance,

> CBMS incorporated an incorrect rule that discontinued food stamps to individuals with past drug problems in violation of Colorado law […] Contrary to federal law, CBMS provided food stamps to college students who did not work the required twenty hours a week.[65]

(ii) Discrimination by reproduction

---

[64] CITRON, D. K. **Technological Due Process**. 85 Wash. U. L. Rev. 1249, 2008, p. 1274. Available at: <http://openscholarship.wustl.edu/law_lawreview/vol85/iss6/2>. Access: January 01, 2019.
[65] Ibid, p. 1268-1269.

By discrimination by reproduction I refer to discriminatory outcomes originating in samples whose biases are reproduced by the algorithm. Unlike the samples comprising the first category, the data in them is not outdated or incorrect, but rather the dataset as a whole is somehow compromised, be it because it misrepresents the intended population, or because the algorithm is programmed to select only part of the dataset, thus generating questionable results. Again, in this scenario, empirical unsoundness arises.

A good example comes up with algorithms designed for recruitment. If a tech firm creates an algorithm aimed at choosing "suitable candidates" for a data scientist position, and that algorithm is programmed to look for people whose profiles are similar to those already hired by the firm, it will likely give enormous preference to men instead of women. The algorithm itself is not mistakenly programmed, nor is the data available "wrong" in the sense of the previous category. But the sample is biased. There are more men than women in data science jobs, so even though women are no less suited for these positions, the effects of a long history of gender discrimination are still felt in the way that competence and talent are construed, even by "gender-blind" algorithms.

As we will see later on, this category is particularly cumbersome when dealing with machine learning algorithms, for in this case the dataset is not merely an input, it is also what trains the program and allows for the creation or modification of the algorithm, which means distortions may be multiplied at each iteration. This type of problem is particularly hard to identify, for the issue lays in the target variable selected by the algorithmic system, which for an outside observer can be extremely laborious to identify.

(iii) Discrimination by misleading correlation

In this case, the model works well, both in terms of the dataset and the algorithm, and there are no inherent biases in the data, but the careless use of correlations leads to

misclassification.[66] The case differs from the previous category in that it does not reproduce a bias, it simply ignores some of the characteristics of a given person, usually because those characteristics are not available in the database, and takes the characteristics that are available as the complete picture. More often than not, this results in people being included in generalizations or categories where they do not belong or do not accurately describe them.

For an illustration, let us say Jane Doe goes to the drugstore once a week to buy a specific kind of medicine commonly used to treat cardiac patients. It turns out that insurance companies use this information as an indicator of heart disease to decide how much she should be charged for healthcare insurance, so she may be categorized into a group whose coverage is more expensive. If Jane buys the medicine, however, because it is also an effective remedy for bunions, not because she has a cardiac condition, the algorithm will have classified her as pertaining to a group to which she may very well not belong. It will do so despite correct data – she does indeed buy the medicine; despite correct code – assuming there is nothing wrong with the health insurer's algorithm; and despite a non-biased sample – there is no bias in assuming that people who buy cardiac medication have a cardiac condition, but it nonetheless results in an incorrect generalization. The problem is that the algorithm assumes there is a correlation between "buying cardiac medicine" and "having a cardiac condition" and fails to account for other possible explanations such as off-label use of the drug.

b) Scenarios of unjustified sound discrimination

Unlike the previous categories, the questionable legality in scenarios of unjustified sound discrimination does not come from statistical error, but rather from the lack of legal justification of a result rendered by a statistically sound process. This difference is key both for the process of identifying the problem and for addressing it, as will be further explored in chapter 4.

[66] Such incorrect classification can be due to spurious correlations, but not necessarily. The problem can also run from the fact the algorithmic system does not have enough information on the individual, and as such classifies her given the information it possesses.

(iv) Discrimination by use of sensitive information

The source of the problems covered by this category lies not in the calculation of probabilities, since the results rendered by the algorithmic system are correct, but rather in the type of information used to reach such the results. Legislation in many jurisdictions, including Brazil, holds that using some specific characteristics is unlawful in and of itself, even if these characteristics are somewhat helpful in reaching statistically sound results.

The Brazilian legislation describes sensitive data in two separate instances, the Credit Information Act (Law 12,414/2011) and the General Data Protection Act (GDPA). The GDPA[67] defines sensitive personal data as data related to racial or ethnic origin, religious conviction, political opinion, membership in unions or religious, philosophical or political organizations, data referring to health or sexual life, as well as individual genetic or biometric data. It also establishes a higher legal threshold for the use of that data (Article 11). Article 3, §3, II of the Credit Information Act, for its part, goes even further by clearly stating that, for the purposes of credit scoring, the use of sensitive information is prohibited.[68]

Relying on sensitive categories for profiling purposes, regardless the potential usefulness of these categories, implies ignoring that there often are alternative traits that, if used, would be equally as reliable and less damaging to the individuals and groups concerned. A common and useful example is that of admission into the military. For many years, women were not admitted into military schools or military service, in Brazil and elsewhere.[69] Usually, the explanation for this limitation was that women are weaker than men, especially in terms of upper-body strength. Men's superior upper-body strength is a scientifically proven fact that bears significance with the job in question, and for that reason, military leaders have justified using gender as a common proxy for admission into the military. Upper-body strength, however, is certainly not the only quality of a good soldier, and other relevant traits, such as the ability to work as a team, resilience, or even psychological stability, cannot be inferred from gender.

[67] It should be noted that the GDPA will only come into force in 2020.
[68] Cf. section 3.2.3.2 below.
[69] Brazilian women were only admitted into the Military Academy of Agulhas Negras, the most important military school in the country, in 2016. It was only in 2014 that they were allowed to enroll for voluntary military service.

Gender, therefore, is very often not the most effective proxy for applicant selection and it disproportionally affects a category of individuals who often possess several other relevant characteristics that make for a good soldier.[70]

In passing, results obtained without the use of sensitive data have a second advantage, in that they tend to be less discriminatory. Avoiding liability for discrimination is in itself a valued outcome. A different aspect of the problems in this category that deserves further attention is the definition of what is considered sensitive information. People usually associate sensitive categories with the *input* entered into or gathered by algorithms, which indeed matches the most common use of the term in the history of antidiscrimination law. It should also be noted, however, that the *output* of the algorithm may be just as sensitive and could therefore fall within the same prohibitions. When an algorithm takes non-sensitive inputs such as what one "likes" on social media, the songs frequently listened to, the social events usually attended, plus the similar preferences of one's friends, and aggregates all of this information, the algorithm may generate outputs that crosses into sensitive categories – such as assumptions regarding one's political beliefs. This is, in fact, precisely what Cambridge Analytica is accused of doing: gathering users' personal information on Facebook to build political/ideological profiles. Other examples include algorithms used to deduce sexual orientation based on facial photographs.[71]

Brazilian law remains unclear on the use of non-sensitive categories to draw conclusions regarding sensitive ones. Similarly, the legal doctrine and jurisprudence do not determine with any certainty on whether outputs will fall under the scrutiny of Article 3, §3, II of the Credit Information Act. Nonetheless, that they will have to answer this ambiguity seems obvious and perhaps necessary judging by the way algorithmic systems operate and the goals they are designed to achieve.

(v) Discrimination by association with the fulfilment of rights

[70] It is for no other reason that Schauer emphasizes the use of sensitive data proxies is often instrumental for unlawful discrimination and prejudice.

[71] LEVIN, S. **New AI can guess whether you're gay or straight from a photograph.** The Guardian. September 08, 2017. Available at: <https://www.theguardian.com/technology/2017/sep/07/new-artificial-intelligence-can-tell-whether-youre-gay-or-straight-from-a-photograph>. Access: January 01, 2019.

The inadequacy represented by this category derives from the relationship between the data used by the algorithm and the enjoyment of a protected right. If there is a strict connection between the two and if the right in question is severely impaired, the use of the data is all the more likely to be discriminatory.[72] In contrast to the previous category, it is not necessary for the data to be sensitive for a problem to occur, what is problematic is the impact of profiling. For the result to be considered discriminatory, two more factors must be examined: whether the classification (i) relies on endogenous characteristics or (ii) it singles out groups that have historically suffered discrimination.

Features used in decision-making can be exogenous or endogenous. In the first case, the variable that renders a group distinct from the other is external, whereas in the second, it is internal and might feed back into the differentiation.[73] When the characteristic under consideration has endogenous effects, it is more likely to lead to discrimination. If, however, the trait is exogenous, the result is less likely to be discriminatory.

Once again turning to gender to exemplify this distinction, let us look at two situations. The first is that of employers who may be prone to hire men rather than women because they associate the group "women" with tougher career choices – they usually have to choose between work and family, and do not always choose work. The employer may well adopt this generalization into his or her decision making-process without any inherent prejudice against women, but rather simply as a cost-efficient way to select applicants, and thus end up selecting more men for the available positions. Now consider of a second statistic: the higher propensity of young male drivers to be involved in car accidents compared to young female drivers. This information is used by companies to price insurance differently for these two groups, and it frequently results in young men paying more for their coverage than young women.

[72] In Schauer's view, the problem in this case is not with discrimination per se but rather with exclusion.
[73] This result is what is usually referred to as a feedback loop. A feedback occurs when the output of a system - say an algorithm - is feed back into the system as an input. In other words, a given effect of the system returns as its cause. The result of less women being hired returns to the decision-making system as an input for the decision-maker and thus reinforces the conclusions it creates.

Although both situations take gender as the relevant trait for the differentiation, they bear a significant difference: in the first case, gender is an endogenous variable, whereas in the second it is exogenous. It may be that women have historically been more involved in child-raising and house-caring duties, but this can be a *result* of having fewer career opportunities. The consequence of the discrimination leads to the confirmation of the initial assumption. In contrast, in the car insurance context, gender is no longer an endogenous variable. It is exogenous, for nothing in the higher price paid for insurance is causally related to the higher number of car accidents involving young men. It is far more likely for the first situation to be considered discriminatory than for the second.

Imagine now that instead of singling out young male drivers statistically accurate data demonstrated that young female Muslims are more prone to car accidents. Let us also assume both gender and religion are exogenous variables. The problem is of a different nature, for it lies in the targeted population. Muslims and women are groups that have historically faced discrimination, and thus it is in society's best interest to avoid profiling that reinforces discriminatory patterns towards them. That is to say that, in the view defended here, if the data showed young male Protestants in the United States, or young male Catholics in Brazil were the most prone to car accidents and were therefore subjected to higher insurance rates the result would not be discriminatory, barring the existence of the problems set out in the previous categories.

2.3.4 Causation, Correlation, Objectivity and The Limitations of Algorithms

Oftentimes, the main problem of algorithms identified by the literature is their disregard for causation and blind faith in correlation. Chris Anderson, former Editor-in-Chief of Wired, wrote perhaps one of the most talked-about pieces defending the "end of theory" and causal analysis, claiming:

> This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. Out with every theory of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they

> do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves.[74]

Anderson was severely criticized by many, such as Financial Times analyst Tim Harford, who in a lecture at the Royal Statistical Society International Conference used the example of Google Flu to show how this promise is at best an oversimplification and at worst just plain wrong.[75] Google Flu Trends was a tool developed by the company to track the spread of the flu in different regions of the world – Google no longer publishes data from Flu Trends, but it did so until 2014. The algorithm mapped search keywords in the Google Search tool and matched them with datapoints, with the intent of predicting future behavior and anticipating where serious breakouts of the flu would occur. Google did not concern itself with the causes of the flu, but with factors correlated to it. It did not particularly care whether outbreaks were due to the weather, lack of preventive care, or a new virus; it was interested in finding and revealing patterns.

After a promising start, the project encountered difficulty.[76] As Lazer, Kennedy, King & Vespignani explain, Google Flu failed rather basic statistical lessons:

> However [enormous the scientific possibilities in big data], quantity of data does not mean that one can ignore foundational issues of measurement and construct validity and reliability and dependencies among data. The core challenge is that most big data that have received popular attention are not the output of instruments designed to produce valid and reliable data amenable for scientific analysis.[77]

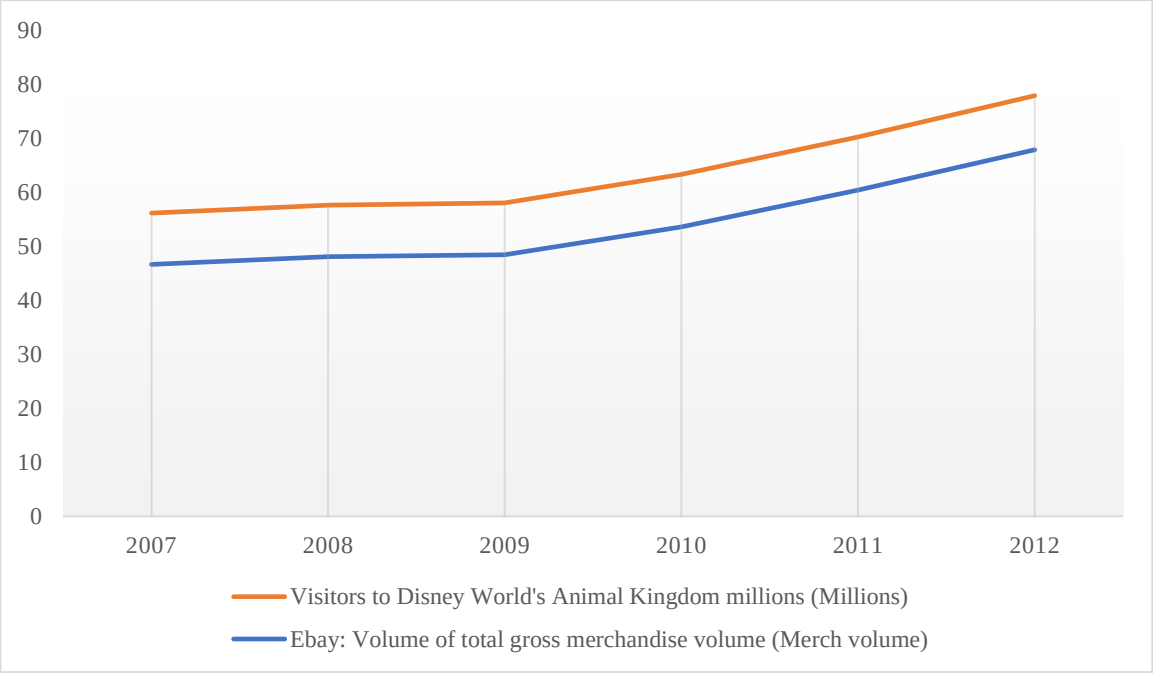
In other words, the fragility of a non-causal theory based merely on correlations is enormous. Not having or at least formulating a hypothesis for a correlation makes it very easy to break the correlations down. At the first sign of instability, the model falls apart. To clarify what the problem with confusing these two concepts is, let us turn back to the concept of spurious correlation. Tyler Vigen assembled a collection showing sets of data that bear no

---

[74] ANDERSON, C. **The End of the Theory: the Data Deluge Makes the Scientific Method Obsolete**. Wired. June 23, 2008. Available at: <https://www.wired.com/2008/06/pb-theory/>. Access: January 01, 2019.
[75] HARFORD, T. **Big data: are we making a big mistake?**. The Financial Times. 2014. Available at: <https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1740-9713.2014.00778.x>. Access: January 01, 2019.
[76] Explain that Google Flu Trends is no longer publicly available, but the tool itself is used by Google.
[77] LAZER, D. et. al. **The Parable of Google Flu: Traps in Big Data Analysis**. Science 343:1203-1205, March 14, 2014. Available at: <https://gking.harvard.edu/files/gking/files/0314policyforumff.pdf>. Access: June January 01, 2019.

relation to each other, but that nonetheless show similar patterns, in order to illustrate the problem in not differentiating causation from correlation. His intent was to present a "fun way to look at correlations and to think about data." Here is the result he compiled for eBay total gross merchandise volume and visitors to Disney Worlds Animal Kingdom, from years 2007 to 2012[78]:

Graph 1 – Visitors to Disney World vs. E-Bay gross merchandise volume.



Source: Vigen.

Common sense leads us to disbelieve that the number of people who visited Animal Kingdom has anything to do with the volume of eBay's total gross merchandise sales. Still, the curves move together, revealing a pattern. Assuming causation in a situation such as this can lead to skewed outcomes and also to discriminatory results.

[78] VIGEN, T. **EBay's Total Gross Merchandise Volume Correlates With Visitors to Disney Worlds Animal Kingdom.** Tylervigen, Spurious Correlations. Available at: <http://tylervigen.com/view_correlation?id=28571>. Access: January 01, 2019.

Yet despite the limitations of working with correlations, there is good reason to believe that because the enormity of the data gathered today, and because causal explanations are not trivial nor represent the cornerstone of every decision made prior to algorithms, the use of correlations may effectively and safely supplants our need for causation in many scenarios. Google Flu trends itself can be an example. If a tool is useful in accurately predicting whether region A will suffer a flu outbreak, that information may be relevant and inform public policy even if the algorithm cannot capture causation. The same goes for applications of this mechanism to commercial relations. If Amazon's algorithm is good enough so it can somewhat accurately predict which books a costumer may be interested in purchasing, the company does not need to know why that person is interested in that title or author, it is enough for it to know that the product is of interest to the client.

There are two issues that are far more complicated, pervasive, and central to the use of algorithms that should be explored, for they are decisive in designing public policies able to deal with this matter: (i) the so-called objectivity of algorithms and (ii) algorithms' limitation in exercising prudence.

2.3.4.1 Objectivity

As previously stated,[79] it is very common – and to some extent understandable – for algorithmic systems to be considered an objective and often even superior form of decision-making. It is not unusual to find arguments that claim algorithmic decisions are better not just because they are efficient and economical, but also because they are free of human bias or even perform better than humans.[80]

Results rendered by machines enjoy an aura of unquestioned accurateness, probably because people associate the processes with mathematics and so imagine that decisions are

[79] See section 2.3.4.1.

[80] CUMMINGS, M. L. **The Social and Ethical Impact of Decision Support Interface Design**. International Encyclopedia of Ergonomics and Human Factors. 2005. Available at: <https://pdfs.semanticscholar.org/a9b3/ec436508ebfa40f3a3f5b59231bece4f3e34.pdf>. Access: January 01, 2019. And PARASURAMAN, R. and MILLER, C. A. **Trust and etiquette in high-criticality automated systems**. Communications of the ACM – Human-computer etiquette, vol. 47 issue 4, April, 2004. Available at: <https://dl.acm.org/citation.cfm?id=975844>. Access: January 01, 2019.

based on precise calculations that render a reproducible output, one that as such could not be different precisely because it resulted from calculations, not from human preferences, the balancing of values, or anything of the sort.

Leaving aside the debate over whether mathematics is indeed more objective than the social sciences,[81] the fact is that algorithms are not as straightforward as college math, or even as very complicated functions. Even though they are essentially intertwined calculations, the basis for their decision-making is just as dependent on social sciences as court opinions seeking to determine whether writing and publishing a book that claims the Holocaust did not occur is covered by freedom of expression. In both situations, the process requires a goal to be set, and a path to be designed for achieving the goal. The decisions on what the goal should be and what that path to it will look like are not based on mathematics alone – and often not on mathematics whatsoever – but on the needs and desires of human beings. As Microsoft's Kate Crawford put it, numbers cannot speak for themselves, "data and data sets are not objective; they are creations of human design."[82] However efficient and cost-saving algorithmic systems may be, they are not by any means free of subjectivity and should not be seen as the bearers of non-biased results.

The connected claim is that, even if not completely unbiased, algorithms are far better than humans in reaching objective results. That is Alex Miller's argument, for example, when he states: "algorithms are less biased and more accurate than the humans they are replacing."[83] The problem with this argument is it puts forward a false debate: whether algorithms are overall better or worse decision-makers when compared to their human counterparts. Authors who highlight the perils of relying on algorithms do not claim that humans are not biased when making decisions, quite the contrary. The point is not to defend humans, but rather to call attention to the social perception of algorithmic decision-making as purely objective, which is false.

---

[81] I am aware that some question this claim, but I will assume it true for my purposes here.
[82] CRAWFORD, H. **The Hidden Biases in Big Data**. Harvard Business Review. April 01, 2013. Available at: <https://hbr.org/2013/04/the-hidden-biases-in-big-data>. Access: January 01, 2019.
[83] MILLER, A. P. **Want Less-Biased Decisions? Use Algorithms.** Harvard Business Review. July 26, 2018. Available at: <https://hbr.org/2018/07/want-less-biased-decisions-use-algorithms>. Access: January 01, 2019.

Moreover, unlike human decision-making, which is well known to be subject to bias, algorithmic decision-making is not perceived as subjective. Where human bias has been recognized, mechanisms to deal with potential bias have been developed. In the legal justice system, for example, one of the solutions for human bias are courts of appeal, which give the parties a chance to review and correct any mistakes that the judge might have made. Similar safety mechanism for algorithmic decision-making have yet to be devised. In fact, the goal of many authors who call attention to the limitations of algorithmic decision-making is not to simply abandon the practice, but to develop tools to deal with their inaccuracies. In other words, algorithms should not be abandoned altogether, but rather society must be educated to understand the nature of decisions rendered by algorithmic systems – what they are and can be, as well as what they cannot be.

2.3.4.2 Prediction, not Judgment

Let us for a moment assume that policy makers have adequately addressed all the problems identified above. The subjectivity of algorithms has been made sufficiently clear, mechanisms for addressing biases are in place, and so forth. Still, there remains one aspect of algorithmic decision-making of fundamental importance that neither of those tools are able to correct: the intrinsic limitation of algorithms when it comes to exercising prudence and judgement.

In *Prediction Machines,* by Agrawal, Gans & Goldfarb, what the advent of artificial intelligence and algorithms means is expressed concisely: better and faster prediction. According to the authors, these new machines bring us the ability to employ prediction in many more scenarios at cheaper prices. They define prediction as "the process of filling in missing information. Prediction takes information you have, often called 'data', and uses it to generate information you don't have."[84] Better prediction, for its part, leads to better decision-making.

[84] AGRAWAL, A., GANS, J. and GOLDFARB, A. **Prediction Machines: The Simple Economics of Artificial Intelligence**. Harvard Business Press, April 17, 2018. p. 24. - The authors engage in a very interesting explanation on how precisely machine learning changed prediction, comparing more traditional regression methods to machine learning.

Prediction, however, cannot be confused with intelligence. The authors explicitly pose that question, and, although the idea that they may be the same has some (very few) defenders,[85] the vast majority of specialists believes there is much more to intelligence than prediction. That seems particularly true when one thinks of the fundamental aspects of legal systems, and of the enforcement of legal rules. An important component in legal thinking is the ability to take real-life scenarios and conform them to existing legal institutions. Classifying human behavior into categories seems intrinsically relevant for legal purposes, and this process probably involves prediction to some extent. Yet there is an additional aspect of legal thinking that is probably even more central to legal analysis: exercising judgment.[86]

Judgement has much less to do with prediction and much more to do with the exercise of prudence. Judgement is about perceiving the nuances of situations and being able to evaluate pros and cons, to balance values and to provide adequate responses to situations. That is strikingly different from what algorithms do nowadays. Machine learning has enabled computers to learn patterns, but not to criticize patterns according to an order of values, much less to make decisions based not on standards, but on an evaluation of what is right or wrong. Currently, programming algorithms to exercise judgment is far beyond the reach of technology. Not surprisingly, the discussions of ethics and algorithms often focus not on "teaching" algorithms values, but on erecting barriers to prevent them from taking certain actions.[87] The reason is simple: because we are unable to teach machines how to reach a fair decision, we either program the fair decision into their systems or we exclude that decision from their purview.

---

[85] Jeff Hawkins is one of the most prominent of such authors. See in: HAWKINS, J. and BLAKESLEE, S. **On Intelligence: How a New Understanding the Brain Will Lead to the of the Creation of Truly Intelligent Machines.** Times Books, 2004.

[86] Danielle Citron puts this same idea in different words. She highlights that law can take two forms: rules or standards. While rules would be the realm of predictability, ex ante instructions for behavior, standards are better described as ex post judgments that should tailor an outcome to facts, and demand decision-makers to articulate their choices. "The emergence of automation threatens to overwhelm this debate by giving rules a huge, and often decisive, advantage on the basis of cost and convenience rather than the desirability of the substantive results they produce." In: CITRON, D. K. **Technological Due Process**. 85 Wash. U. L. Rev. 1249, 2008, p. 1303. Available at: <http://openscholarship.wustl.edu/law_lawreview/vol85/iss6/2>. Access: January 01, 2019.

[87] IAPP and THE UNITED NATIONS GLOBAL PULSE. **Building Ethics into Privacy Frameworks for Big Data and AI.** Available at: <https://iapp.org/media/pdf/resource_center/BUILDING-ETHICS-INTO-PRIVACY-FRAMEWORKS-FOR-BIG-DATA-AND-AI-UN-Global-Pulse-IAPP.pdf>. Access: January 01, 2019.

What should be clear at this point, as it permeates the entirety of this work, is that algorithms are not actually “intelligent,” nor do they provide answers for each and every question that has long haunted humanity. They are extremely useful tools that must be used without losing sight of their limitations. The same way using a hammer to cut paper will likely be ineffective and have less than optimal results, using algorithms to determine what is “fair” is doomed to failure, or at the very least will yield suboptimal results.

3 Algorithmic Discrimination in Practice – The Challenges in Enforcement

This work has so far described and put forward means by which discrimination by algorithmic systems can take place. It is now time to discuss when and how rights arising from such violations will indeed be enforceable. That is the main goal of this section.

Because the topic of algorithmic discrimination is recent and there is no abundance of legislation addressing it, and also because there is not necessarily a coincidence of personal data/privacy concerns and algorithmic discrimination, the legal discussion inevitably ends up being about (i) how this conduct affects constitutionally recognized rights such as equality. Because much of algorithmic discrimination happens by the hands of private actors, the result is a very traditional debate in constitutional law, the applicability of fundamental rights to horizontal relations.[88] It is not the goal of this dissertation to develop a thesis on this topic, but to consider the implications of the prevailing theories to the problem at hand. Therefore, the three prevailing views will be presented below. Once again, the ultimate focus of the text will be the Brazilian legal system, but I will provide some context on the debate in other jurisdictions and how other understandings about the horizontal application of fundamental rights may lead to different results in terms of enforceability.

After presenting the debate on constitutional terms, I will also delve deeper into (ii) ordinary legislation that tackles, or could be used to tackle, algorithmic discrimination. More specifically, the focus will be on antidiscrimination diplomas and on data protection regulation in the United States, Germany (and to some extent the European Union), and Brazil. As will be clarified, there currently is no abundance of legislation concretely targeted to algorithmic discrimination, which is to be expected given the novelty of this debate, but some acts and laws do bring up the topic and provide paths to be followed by enforcers.

[88] There is a long-standing discussion in constitutional tradition on whether individuals' constitutionally recognized rights are applicable only against the State or also among private parties. Is the right to due process, for example, only applicable when the decision-making is carried out by a public entity, or should it be observed in private decision-making as well? Similar questions arise when dealing with algorithmic discrimination. Can private companies carry out credit scoring however they please, regardless of the potential discriminatory outcomes? If the constitutional right to equality is applicable to private relations, it can be invoked against a practice such as this. That is why this discussion is relevant for my purposes.

Lastly, the text examines two cases of algorithmic discrimination – the use of algorithms in Poland to fight unemployment, and the case of credit scoring in Brazil – that will help clarify my point and hopefully show why the enforcement debate is indeed meaningful.

3.1 The Horizontal Effects of Fundamental Rights

Many instances of potential algorithmic discrimination are carried out by private parties. Automated selection of candidates to fill job positions in a company, creditworthiness of individuals, as well as the selection of advertisement to be shown to a person are all practices implemented in the private sector. In that light, if one wishes to discuss the potential applicability of constitutional rights and guarantees – such as the right to equality – to these scenarios, one must discuss the limits for the applicability of such rights between private agents.

The applicability of fundamental rights[89] to public-private relations is the very justification for the existence of such rights: their enforceability by the individual against the State establishes a protected sphere for every individual that not even the government can pierce. The question is different, however, when one thinks of private-private relations.

This discussion on whether fundamental rights are enforceable or have any bearing on private relations has emerged in many jurisdictions, with various outcomes. The sections that follow will briefly present the debate in three jurisdictions, which provide different answers to the question at hand: the United States of America, which adopts the so-called state action doctrine; Germany, which embraces the idea of "mittelbare Drittwirkung" (or indirect horizontal effects); and Brazil, whose Supreme Court understands fundamental rights are directly applicable to private relations.[90]

3.1.1 United States of America

The state action doctrine embodies the jurisprudence formulated by the American Supreme Court according to which the rights put forward in the Constitution, especially those

[89] In some instances, the expression "constitutional" rather than "fundamental" rights is used in the debate. For the purposes of this dissertation, these expressions will be taken as synonymous.
[90] That is not to say that no authors in all of these jurisdictions defend different views, but simply what is considered the prevailing theory in each country.

in the First and Fourteenth Amendment, are only enforceable against the State, never against individuals. In other words, the citizenship clause, the due process clause, the equal protection clause, the privileges and immunities clause, as well as the rights of freedom of speech, of the press, of religion and of assembly, are not applicable to horizontal relations.[91] The reasons advanced by the Supreme Court to support this understanding are two-fold: (i) first and foremost, limiting applicability is essential for private autonomy, and (ii) the U.S. federalist system that reserves the domain of private law to the individual states requires it.

There are three tests[92] that the Court applies to determine the suitability of the state action doctrine. The first, the "public function test", determines as public the exercise by private actors of functions that the State exclusively holds. The case that largely established this test is *Marsh v. Alabama*. In *Marsh*, the discussion centered on whether a Jehovah witness could profess her beliefs and distribute pamphlets in the streets of Chickasaw, a town in Alabama. The peculiarity of the case is the fact Chickasaw is entirely owned by a company, the Gulf Shipbuilding Corporation, which claimed permission to distribute religious material was needed before such action could be carried out by an individual. Marsh, for her turn, claimed that the restriction violated her First Amendment rights.

The opinion of the Supreme Court sided with Marsh, stating that

> Whether a corporation or a municipality owns or possesses the town, the public in either case has an identical interest in the functioning of the community in such manner that the channels of communication remain free. (…) When we balance the Constitutional rights of owners of property against those of the people to enjoy freedom of press and religion, as we must here, we remain mindful of the fact that the latter occupy a preferred position. (…) In our view, the circumstance that the property rights to the premises where the deprivation of liberty, here involved, took place were held by others than the public is not sufficient to justify the State's permitting a corporation to govern a community of citizens so as to restrict their fundamental liberties and the enforcement of such restraint by the application of a state statute.[93]

---

[91] The only universal exception is the Thirteenth Amendment, which prohibits slavery.

[92] Some authors say there are actually four strands. See: GARDBAUM, S. **The 'Horizontal Effect' of Constitutional Rights**. Michigan Law Review, vol. 102, UCLA School of Law Research Paper No. 03-14, pp. 388-459, 2003. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=437440>. Access: January 01, 2019.

[93] Marsh v. Alabama, 326 U.S. 501 (1946).

The second test focuses on unveiling "whether the state is significantly entangled with, or jointly participating in, the actions of a private actor."[94] Unlike the previous test, there is no need for a typically state-held function to be involved, all the test requires is a strong enough nexus between the private actor and the State. This debate was present in *Blum v. Yaretsky*[95], a case that scrutinized the applicability of due process by private nursing homes. Although patient care at such nursing homes was predominantly funded by the State, the nursing facilities were solely responsible for deciding the level of care to give patients, and therefore the facility to which the patients were placed.

As the court clarified, federal regulations require each nursing home to establish a utilization review committee (URC) of physicians whose functions include periodically assessing whether each patient is receiving the appropriate level of care, and thus whether the patient's continued stay in the facility is justified. The debate regarded the legality when a URC issued a decision without providing patients with notice, stating the reasons supporting the decision, or giving any opportunity for the patients to challenge the decision, as required by the due process clause. Ultimately, the court decided state action had not been proved and constitutional rights were not applicable for two reasons: (i) first, "the mere fact that a private business is subject to state regulation does not, by itself, convert its action into that of the State for the purposes of the Fourteenth Amendment,"[96] and (ii) "[t]he fact that the State responds to the nursing homes' discharge or transfer decisions by adjusting the patients' Medicaid benefits does not render it responsible for those decisions."[97]

Lastly, the third test checks whether the State actively encouraged the action by the private party. As Gardbaum observes, a common problem when dealing with the third test is frequent misapplication, "[f]or the cases sometimes treat this issue not as a threshold one of state action – i.e. whether the Constitution applies – but rather as a substantive one of

[94] GARDBAUM, S. **The 'Horizontal Effect' of Constitutional Rights**. Michigan Law Review, vol. 102, UCLA School of Law Research Paper No. 03-14, p. 412, 2003. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=437440>. Access: January 01, 2019.
[95] Blum v. Yaretsky, 457 U.S. 991 (1982).
[96] Ibid, p. 1003-1005.
[97] Ibid, p. 993.

constitutionality,"[98] as the discussion in *Reitman v. Mulkey*[99] demonstrates. The debate in *Reitman* primarily examined whether under California's Proposition 14 a landlord could refuse to rent a property to someone on the sole basis of their race. The Proposition stated:

> Neither the State nor any subdivision or agency thereof shall deny, limit or abridge, directly or indirectly, the right of any person, who is willing or desires to sell, lease or rent any part or all of his real property, to decline to sell, lease or rent such property to such person or persons as he, in his absolute discretion, chooses.

The Supreme Court of California, instead of focusing on whether the third state action test applied, decided the content of the Proposition to be unconstitutional for violating the Fourteenth Amendment and the Equal Protection Clause.

The decision that is usually considered to have established the boundaries of the third test is *Shelley v. Kraemer.*[100] In the city of St. Louis, Missouri, a neighborhood signed an agreement among its members restricting African-Americans and Asian-Americans from moving there. The Shelleys, unaware of the provisions, moved in to the neighborhood in 1945, and were sued by one of its residents, Louis Kraemer. The state court found that Kraemer was indeed in right and enforced the provision, considering it a private agreement immune from state action. The Supreme Court however found that state action took place when a court was called upon to enforce the private agreement, as was the case in *Shelley*, and therefore the private agreement violating any constitutionally protected rights such as the Fourteenth Amendment could not be enforced.

Regardless of the many exceptions the three tests grant, the prevailing understanding of comparative constitutional law is that the U.S. is where fundamental rights protections are least applicable to private relations.[101] In the words of Mark Tushnet: "standard U.S. constitutional

---

[98] GARDBAUM, S. **The 'Horizontal Effect' of Constitutional Rights**. Michigan Law Review, vol. 102, UCLA School of Law Research Paper No. 03-14, p. 413, 2003. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=437440>. Access: January 01, 2019.

[99] Reitman v. Mulkey, 387 U.S. 369 (1967).

[100] Shelley v. Kraemer, 334 U.S. 1 (1948).

[101] It should be noted, however, that there are authors who disagree. Gardbaum, for example, argues that "the issue of the scope of application of constitutional rights is resolved within comparative constitutional law by answering the following series of questions: (1) Are individuals as well as government actors bound by constitutional rights? (2) Do constitutional rights apply to private law or common law? (3) Are courts bound by constitutional rights? (4) Do constitutional rights apply to litigation between private individuals? The answer to the first of these questions

doctrine is that constitutional provisions do *not* have horizontal effect."[102-103] In that light, one could theoretically conclude that protection for individuals harmed by algorithmic discrimination is limited unless the discriminatory conduct is carried out by the government. As will be seen below,[104] while protection is lacking in the U.S. for individuals in many of the scenarios that arise out of the recent development of algorithmic systems, American citizens are not as vulnerable as one might imagine, be it because the state action doctrine allows for the application of fundamentally established values to some private scenarios, be it because of other historical developments in U.S. legislation.

3.1.2 Germany

German constitutional law understands differently the applicability of fundamental rights to horizontal relations: it applies what is known as "mittelbare Drittwirkung", or indirect horizontal effects theory, to such situations. The first author to formally defend this theory was Günter Dürig,[105] who was followed by others such as Konrad Hesse and also by the Bundesverfassungsgericht (the German Constitutional Court or GCC).

This approach, like the state action doctrine, holds that the argument that fundamental rights are directly applicable to private relations is unsustainable, for that would effectively eliminate private autonomy altogether. Furthermore, the theory's defenders usually argue that direct applicability would confer too much power to the judiciary, so that ultimately every

---

resolves only the issue of direct horizontal effect; the remaining ones address the issue of possible indirect horizontal effect. In the United States, however, the only question that is conventionally asked concerning the scope of constitutional rights is the first one, and the answer given (the state action doctrine) is supposed to supply all necessary answers to the general issue." In: GARDBAUM, S. **The 'Horizontal Effect' of Constitutional Rights**. Michigan Law Review, vol. 102, UCLA School of Law Research Paper No. 03-14, p. 411, 2003. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=437440>. Access: January 01, 2019.

[102] TUSHNET, M. **The issue of state action/horizontal effect in comparative constitutional law**. Oxford University Press and New York University School of Law, vol. 1, number 1, 2003, p. 81.

[103] Other authors disagree, and claim that the state action doctrine is nothing but an apparent limitation on applicability of fundamental rights to private relations. In Virgílio Afonso da Silva's words: "This trick consists of attributing to the State acts carried out by private actors, or of equating private acts to State action. As such, even without expressly accepting private agents' link to fundamental rights, it is possible to reach a comparable result in practice." In: SILVA, V. A. da. **A Constitucionalização do Direito: os direitos fundamentais nas relações entre particulares.** São Paulo: Malheiros Editores, 2014.

[104] See section 3.2.1.

[105] The most important reference is his work, see: DÜRIG, G. **Grundrechte und Zivilrechtsprechung.** Vom Bonner Grundgesetz zur gesamtdeutschen Verfassung: Festschrift zum 75. Geburtstag von Hans Nawiasky, 1956.

contract would be interpreted in light of the open-ended statements of the Constitution, rendering private law irrelevant, since everything would always be analyzed through the lens of constitutional values.[106]

What the theory does recognize, however, is (i) the subordination of the rest of the legal system to the Constitution, meaning that private law must not violate constitutionally established values, and where it does it should be deemed void, and (ii) that private law should always be interpreted in light of constitutional values. Therefore, the Constitution is not directly applicable to private matters, but its application is carried out through private law, for there is a constitutional obligation to protect fundamental rights that is imposed on the entire legal system.

The case that established this understanding before the GCC is *Lüth.*[107] Veit Harlan, a film producer and former Nazi, produced a movie in 1950 that prompted Erich Lüth, then president of the Press Association in Hamburg, to organize a boycott. The case was brought before the state court in Hamburg, which decided in favor of Harlan and the movie's distributor, claiming that §826 of the German Civil Code held Lüth liable for damages.[108] Lüth successfully appealed to the GCC, which reasoned that §826 should be interpreted in light of the constitutional values and fundamental rights set forth by the German constitutional system, which the Hamburg court had failed to do. Free speech, in this specific case, outweighed the distributor's economic interests.

This ruling by the GCC was not based on direct applicability of free speech to the case at hand, but rather on the examination of the constitutional values undergirding the case. A private dispute could not be resolved in direct contradiction of the constitutional order, and contradiction would result if the film producer Varian was granted damages. The solution was found by recognizing the hierarchical superiority of fundamental rights over private relations regulated by private law, and thus the need for private law to conform to constitutional values.

---

[106] Konrad Hesse claims that direct applicability menaces private law's identity. See: HESSE, K. **Verfassungsrecht und Privatrecht**. Müller Jur. Vlg. C. F., 1988, p. 43.
[107] BVerfGE, 7, 198.
[108] § 826 BGB: a person who, contrary to good customs, causes damages to another person, must compensate such damages.

It should be noted that recently the GCC ruled on a case that may have modified this position, which for its part may question Germany's standing as the prime example of indirect applicability. The case concerned a 16 year-old FC Bayern Munich fan, who in 2006 attended a soccer match against MSV Duisburg. The match took place at Duisburg's stadium. As the GCC clarifies:

> After the end of the match, verbal and physical altercations involving a group of FC Bayern Munich fans, among them the complainant, and fans of MSV Duisburg resulted in personal injury and damage to property. Subsequently, approximately 50 persons, including the complainant, were placed in police custody for the purposes of establishing their identities. The public prosecution office opened investigation proceedings on suspicion of rioting charges against the complainant. Following this, MSV Duisburg imposed a ban on the complainant at the suggestion of the local chief of police, prohibiting him from entering any stadium in Germany until June 2008. […] The complainant brought an action requesting that the nationwide stadium ban be lifted. After the initial application filed by the complainant had been rendered moot, the complainant modified his application in the appeal proceedings to an application seeking a declaration that the stadium ban had been unlawful. The initial action and the appeal on points of fact and law, as well as the appeal on points of law before the Federal Court of Justice (Bundesgerichtshof), were unsuccessful. With his constitutional complaint, the complainant claims a violation of his fundamental rights, contending that he was banned from entering stadiums on the basis of a mere suspicion, without viable justification or reasons.[109]

[109] BUNDESVERFASSUNGSGERICHT. **Decision in "Stadium Ban" proceedings clarifies the indirect horizontal effects of the right to equality in private law relations.** Press Release No. 29/2018, April 27, 2018. Available at: <https://www.bundesverfassungsgericht.de/SharedDocs/Pressemitteilungen/EN/2018/bvg18-029.html>. Access: January 12, 2019.

The official statement of the GCC is that the decision "clarifies" the indirect horizontal effects of fundamental rights in private relations, but its effects may prove otherwise.[110-111]

3.1.3 Brazil

Brazil lies in the opposite end of the spectrum from the United States. A significant number of authors, as well as the Brazilian Supreme Court (the Supremo Tribunal Federal or STF,)[112] argue that the fundamental rights set forth in the Constitution are not only an order of

---

[110] As per the official translation provided by the Bundesverfassungsgericht: "The standard of review applicable to the challenged decisions under constitutional law is informed by the doctrine of the indirect horizontal effects of fundamental rights (mittelbare Drittwirkung der Grundrechte). a) The challenged decisions concern a legal dispute between private actors relating to the scope of rights of ownership and possession vis-à-vis third parties under private law. According to the established case-law of the Court, fundamental rights may have a bearing on such disputes by way of indirect horizontal effects (cf. BVerfGE 7, 198 <205 and 206>; 42, 143 <148>; 89, 214 <229>; 103, 89 <100>; 137, 273 <313 para. 109>; established case-law). Fundamental rights do not generally create direct obligations between private actors. They do, however, permeate legal relationships under private law; it is thus incumbent upon the regular courts to give effect to fundamental rights in the interpretation of ordinary law, in particular by means of general clauses contained in private law provisions and legal concepts that are not precisely defined in statutory law. These effects are rooted in the decisions on constitutional values (verfassungsrechtliche Wertentscheidungen) enshrined in fundamental rights, which permeate private law in terms of "guiding principles" (cf. BVerfGE 73, 261 <269>; 81, 242 <254>; 89, 214 <229>; 112, 332 <352>); accordingly, the case-law of the Federal Constitutional Court has referred to the fundamental rights as an "objective order of constitutional values" (cf. BVerfGE 7, 198 <205 and 206>; 25, 256 <263>; 33, 1 <12>). In this context, the fundamental rights do not serve the purpose of consistently keeping freedom-restricting interferences to a minimum; rather, they are to be developed as fundamental values informing the balancing of the freedoms of equally entitled rights holders. The freedom afforded one right holder must be reconciled with the freedom afforded another. For this purpose, it is necessary to assess conflicting fundamental rights positions in terms of how they interact, and to strike a balance in accordance with the principle of practical concordance (praktische Konkordanz), which requires that the fundamental rights of all persons concerned be given effect to the broadest possible extent (cf. BVerfGE 129, 78 <101 and 102>; 134, 204 <223 para. 68>; 142, 74 <101 para. 82>; established case-law)."
See in: BUNDESVERFASSUNGSGERICHT. **Headnotes to the Order of the First Senate of 11 April 2018 – 1 BvR 3080/09.** Available at: <https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/EN/2018/04/rs20180411_1bvr308009en.html;jsessionid=34F6550A81C53BC9257433E87D7CF087.1_cid393>. Access: January 12, 2019.

[111] Some authors already argue that in practice the difference between direct applicability and indirect applicability as carried out by the GCC is mostly irrelevant. See: GRUNDAMNN, S. **Constitutional Values and European Contract Law**. Aspen Publishers, 2008, p. 5-8.

[112] An argument can be made that the STF's actual view of any topic is hard to identify, for the court's way of expressing opinions is very peculiar. Oftentimes Justices end up writing their "votes" based on different arguments, and the final decision does not clearly state which arguments prevailed or constitute the tribunal's ratio decidendi. I myself, along with many others, have previously pointed to this problem: MATTIUZZO, M. **Voto Vencido, Fundamentação Diversa e Fundamentação Complementar: um estudo sobre deliberação no Supremo Tribunal Federal.** 2011. SBDP. Available at: <http://www.sbdp.org.br/publication/voto-vencido-fundamentacao-diversa-e-fundamentacao-complementar-um-estudo-sobre-deliberacao-no-supremo-tribunal-federal/>. Access: January 05, 2019.; SILVA, V. A. Da. **De Quem Divergem os Divergentes: os Votos Vencidos no Supremo Tribunl Federal.** Direito, Estado e Sociedade, n. 47, p. 205-225, July/December, 2015. Available at: <http://direitoestadosociedade.jur.puc-rio.br/media/artigo09n47.pdf>. Access: January 07, 2019. MENDES, C. H. **Direitos fundamentais, separação de poderes e deliberação.** Tese (Doutorado em Ciência Política) – Faculdade de Filosofia, Letras e Ciências Humanas da Universidade e São Paulo, São Paulo, 2008. KLAFKE, G. F. **Vícios no Processo Decisório do Supremo Tribunal Federal**. SBDP 2010. Available at:

values against which private law should be interpreted, but may in fact be directly applied to horizontal relations.

The direct horizontal theory states that fundamental rights are applicable *erga omnes*. As one of its defenders, Daniel Sarmento, puts it,

> the supporters of the direct horizontal effects theory do not negate the existence of specificities in such applicability, nor the need to balance the fundamental right at stake with private autonomy. It is not, therefore, a radical doctrine, that may lead to freedom-restricting results, as claimed by its opponents, for it does not ignore individual freedom in private relations, it simply requires that such freedom be balanced in concrete scenarios.[113]

Addressing the same point, Ingo Sarlet states that there is nothing particularly problematic about the tension between private autonomy and other fundamental rights when it comes to private relations. He does not deny the potential specific problems of applicability, but emphasizes that conflicts among fundamental rights are extremely common in public-private relations as well.[114]

It is worth noting the main consequence that distinguishes the indirect and direct theories: under direct horizontal effects theory, there is no need for ordinary legislation of any kind to be present so the "order of values" can be applied to private relations. Fundamental rights immediately confer individuals rights against other individuals. That is not the same as saying, however, that all constitutionally established rights are applicable to any and all private

---

<https://www.sbdp.org.br/publication/vicios-no-processo-decisorio-do-supremo-tribunal-federal/>. Access: January 07, 2019.

However, though the STF does indeed fail in justifying the reasons to adopt direct application of fundamental rights, and the justifications often differ from Justice to Justice, it is possible to say that in the vast majority of cases it accepts such applicability. That is precisely what Sarmento states in SARMENTO, D. **Direitos Fundamentais e Relações Privadas**. Rio de Janeiro: Lumen Juris, 2004, p. 297.

[113] Ibid, p. 72.

Original in Portuguese reads: "Os adeptos da teoria da eficácia imediata dos direitos fundamentais nas relações privadas não negam a existência de especificidades nesta incidência, nem a necessidade de ponderar o direito fundamental em jogo com a autonomia privadas dos particulares envolvidos no caso. Não se trata, portanto, de uma doutrina radical, que possa conduzi a resultados liberticidas, ao contrário do que sustem seus opositores, pois ela não prega a desconsideração da liberdade individual no tráfico jurídico-privado, mas antes impõe que seja devidamente sopesada na análise de cada situação concreta."

[114] SARLET, I. W. **Direitos Fundamentais e Direito Privado: algumas considerações em torno da vinculação dos particulares aos direitos fundamentais**. Revista dos Tribunais Online. Available at: <http://www.direitocontemporaneo.com/wp-content/uploads/2018/03/SARLET-Direitos-fundamentais-e-direito-privado.pdf>. Access: January 01, 2019.

relations. It is saying that if a given fundamental right is applicable to a given private relation, then applicability does not require mediation by ordinary law.

The first to formally defend the idea of direct applicability of fundamental rights to private relations was Hans Carl Nipperdey, whose thesis claimed that "the dangers that surround fundamental rights in the modern world do not originate solely in actions by the State, but run also from social powers and third parties in general."[115] Nipperdey presided the Germany Federal Labor Court, a court that issued some decisions which supported his thesis, but did not convince the GCC to adopt the doctrine of direct horizontal effects.[116]

Virgílio Afonso da Silva has made an important observation in discussing horizontal effects theories, pointing out the relevance of the differences between efficacy, effectiveness, and applicability. According to Afonso da Silva, the claim by some authors that direct applicability of fundamental rights steams from Article 5, §1 of the Brazilian Constitution ("Norms which define fundamental rights and guarantees have immediate applicability"[117]) is inaccurate:

> [T]he simple constitutional statement that the defining norms of fundamental rights will have 'immediate applicability' says *absolutely nothing* about *which* relations will be subjected to their effects, that is, it does not bring any indication about the type of relation that should be disciplined by fundamental rights.[118]

Afonso da Silva provides an alternative path for fundamental rights applicability, departing from Robert Alexy's three-level thesis adapted for the Brazilian constitutional framework. Under his view, both indirect and direct applicability have relevant roles, for "only

---

[115] SARMENTO, D. **Direitos Fundamentais e Relações Privadas**. Rio de Janeiro: Lumen Juris, 2004, p. 245.

[116] See OETER, S. **Fundamental Rights and Their Impact on Private Law – Doctrine and Practice Under the German Constitution**. 12 Tel Aviv U. Stud. L. 7, 1994. Available at: <https://heinonline.org/HOL/LandingPage?handle=hein.journals/telavusl12&div=4&id=&page=>. Access: January 01, 2019.

[117] Original in Portuguese reads: § 1º As normas definidoras dos direitos e garantias fundamentais têm aplicação imediata.

[118] SILVA, V. A. da. **A Constitucionalização do Direito: os direitos fundamentais nas relações entre particulares.** São Paulo: Malheiros Editores, 2014, p. 58.
Original in Portuguese reads: "Mas a simples prescrição constitucional de que as normas definidoras de direitos fundamentais terão 'aplicação imediata' não diz absolutamente nada sobre quais relações jurídicas sofrerão seus efeitos, ou seja, não traz indícios sobre o tipo de relação que deverá ser disciplinada pelos direitos fundamentais."

a differentiated model is capable of addressing the many different situations in which fundamental rights affect private relations."[119]

Regardless how direct horizontal effects theory should be applied and what its limitations are, in practice the influence of variations of this understanding can be seen operating in the Brazilian judicial system. To cite the words of Supreme Court Justice Gilmar Mendes, in reference to the case that is considered to have established the validity of horizontal application by the STF:

> I am not currently concerned with discussing the way this Court's case law defines fundamental rights should be applied to regulate private relations. I am solely concerned with emphasizing that this Court has an identifiable history or constitutional jurisprudence professing the applicability of such rights to private relations.[120]

Paula Gorzoni researched the matter more extensively and her conclusions point in the same direction. She investigated the STF cases where horizontal applicability of fundamental rights was expressly debated[121] and concluded that the Court often recognizes fundamental rights in private relations. In her analysis of 18 cases that expressly debated horizontal applicability, only two resulted in non-applicability (and one did not have a decision on the merits of the case), in all other instances the Justices held fundamental to be indeed applicable to private relations.[122]

---

[119] Ibid, p. 145. Original in Portuguese reads: "somente um modelo diferenciado é capaz de enquadrar os diversos tipos de situações em que os direitos fundamentais produzem efeitos nas relações entre particulares".

[120] STF. Recurso Extraordinário 201.819-8, Rio de Janeiro, 2005, p. 607. Available at: <http://redir.stf.jus.br/paginadorpub/paginador.jsp?docTP=AC&docID=388784>. Access: January 09, 2019. Original in Portuguese reads: "Não estou preocupado em discutir no atual momento qual a forma geral de aplicabilidade dos direitos fundamentais que a jurisprudência desta Corte professa para regular as relações entre particulares. Tenho a preocupação de, tão-somente, ressaltar que o Supremo Tribunal Federal já possui histórico identificável de uma jurisdição constitucional voltada para a aplicação desses direitos às relações privadas."

[121] Gorzoni explains her methodology in detail in her work. See in: GORZONI, P. F. A. da C. **Supremo Tribunal Federal e a Vinculação dos Direitos Fundamentais nas Relações entre Particulares**. 2007, p. 7-11. Available at: <http://www.sbdp.org.br/publication/supremo-tribunal-federal-e-a-vinculacao-dos-direitos-fundamentais-nas-relacoes-entre-particulares/>. Access: January 13, 2019.

[122] "[G]eralmente, o STF vincula os direitos fundamentais nas relações entre particulares. Isso porque, dos 18 casos analisados, em apenas duas ocasiões não foram aplicados estes direitos e em outra não houve julgamento de mérito, sendo que nas outras 15 houve vinculação de alguma maneira. Tal fato confirma a hipótese inicial de que o tribunal já vinha decidindo questões entre particulares ao longo dos anos, apesar de somente com o RE 201.819/RJ ("Caso UBC") ter identificado expressamente a situação como um conflito entre sujeitos privados." In: Ibidem.

3.1.4 Relevance for algorithmic discrimination

As the state action doctrine and the indirect horizontal effects theory emphasize, the main objection to the direct application of fundamental rights to private relations is the sanctity of free will (or private autonomy). Some claim that, in assuming such rights are immediately binding on private relations, one may end up diluting constitutional law. Notwithstanding, many if not most jurisdictions concede some level of applicability of fundamental rights to private relations – and even the state action doctrine has (several) exceptions that make them enforceable against individuals.

The issue of private autonomy, however, remains central to the debate, and is particularly meaningful with regards algorithmic discrimination. Those who defend limiting private autonomy in favor of other fundamental rights are unanimous in affirming the importance of establishing criteria to evaluate the relevance of autonomy in concrete scenarios.[123] The prevailing standard for most authors is the degree of effective autonomy exercised by the individual.[124] Measuring effective autonomy is generally challenging,[125] but even more so in the cyberspace. In the digital landscape, the issues of consent permeate the legislation and are central in most instruments referring to personal data. Consent is not, however, a clear-cut matter free of all hurdles.

Given the character of the digital environment, it is often hard to determine whether an individual has truly expressed her consent (i.e., for the use of data that can be collected from her profile in a social network by a given third-party application) or if, pressed by the need or desire for access to a given functionality and the difficulty in comprehending the complicated terms and conditions usually put forward by companies, decided to signal agreement with the

---

[123] They are however not unanimous in affirming what this process entails. For example, whereas Daniel Sarmento defends balancing fundamental rights, Afonso da Silva explicitly states that "O que se faz, ao que parece sem exceções, é definir situações em que a autonomia privada deve ser mais respeitada e situações em que esse respeito poderá ser mais facilmente mitigado. Esse raciocínio – que é, de fato, correto – não é, contudo, um sopesamento". In: SILVA, V. A. da. **A Constitucionalização do Direito: os direitos fundamentais nas relações entre particulares.** São Paulo: Malheiros Editores, 2014, p. 155.

[124] See: Ibid, p. 369.

[125] Sarmento defends balancing as the solution. Afonso da Silva, on the other hand, defends the use of a "differentiated solution, one that is able to flexibly approach the distinct configurations of the problem." In: Ibid, p. 134.

specifications without fully comprehending them. Despite efforts by private parties to mitigate such issues, but they undoubtedly persist to various degrees.

One conclusion that can be drawn from the preceding discussion is that directly applying fundamental rights to private relations is not without controversy. Even if direct applicability is accepted, private autonomy remains a value that must be preserved and the hurdles in verifying effective consent are anything but trivial, which means that horizontal applicability will often be questioned. Because of all these hurdles, and in recognition of the potential for discriminatory outcomes, legislation has emerged in many jurisdictions aimed specifically at tackling discrimination and data protection. The next sections will scrutinize the pertinent legislation in each of the three jurisdictions, both to understand their reach and verify whether they represent feasible enforcement alternatives for the issues identified in chapter 2.

3.2 Antidiscrimination and Data Protection Legislation – What Lies Beyond Fundamental Rights

Alternatives and complements to constitutional rights applicability to address both general discriminatory issues and personal data have emerged in many jurisdictions. My goal in the following sections is thus to verify whether there is indeed ordinary legislation that offers a useful path for enforcement against *algorithmic* discrimination, especially in jurisdictions where the direct horizontal effects theory is not commonly practiced.

3.2.1 The United States and the Civil Rights Act

The most important element in American antidiscrimination law is the Civil Rights Act of 1964, which prohibits discrimination based on sex, gender, race, religion or origin. Though the Act is mostly directed towards public actors (state and municipal governments, "public accommodations," public schools, programs and activities receiving federal funds, and so forth), it encompasses the notable exception of Title VII, which prohibits discrimination by private employers.[126]

[126] As with most American legislation, the Title has very specific conditions. It applies to employers who have "fifteen (15) or more employees for each working day in each of twenty or more calendar weeks in the current or

Other statutes are the Fair Credit Reporting Act (FCRA), the Fair Housing Act, and the Age Discrimination Employment Act (ADEA).[127] FCRA is the legislation that regulates credit scoring and credit reporting in the U.S. This Act contains a provision that specifically allows consumers to request their credit reports and dispute incorrect information. As employers started making use of credit scores for hiring purposes, FCRA also included provisions aimed at protecting individuals – the employee must always be advised that her employer will make use of such reports and obtain written permission.

The Fair Housing Act aims to protect whoever wishes to buy or rent a property against potentially discriminatory behavior from sellers or landlords, who cannot refuse to sell or to rent based on race, gender, religion, disability, familial status or national origin. ADEA, for its turn, is generally directed towards discrimination against workers over 40 years-old – a provision that today seems obsolete, since a person's productive span has dramatically increased since 1967 when the act was passed into law. It also contains more specific provisions related to mandatory retirement.

In their attempt to determine whether these statutes could provide the legal means to prohibit algorithmic discrimination, Barocas & Selbst note two tests of Title VII that can be used to establish employer liability for discrimination: disparate treatment and disparate impact.[128] They conclude that "aside from rational racism and masking (with some difficulties),

---

preceding calendar year." Exceptions to this general rule also exist, mostly in the form of the so-called bona fide occupational qualifications.

[127] A more thorough analysis can be found in the report by the FTC. See in: FEDERAL TRADE COMMISSION. **Big Data, A Tool for Inclusion or Exclusion? – Understanding the Issues**. January, 2016. Available at: <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf>. Access: January 12, 2019.

[128] BAROCAS, S. and SELBST, A. D. **Big Data's Disparate Impact.** California Law Review, vol. 671, 2016, p. 694. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899>. Access: January 01, 2019.
As the FTC clarifies, "Disparate treatment occurs when a creditor treats an applicant differently based on a protected characteristic. For example, a lender cannot refuse to lend to single persons or offer less favorable terms to them than married persons even if big data analytics show that single persons are less likely to repay loans than married persons. Disparate impact occurs when a company employs facially neutral policies or practices that have a disproportionate adverse effect or impact on a protected class, unless those practices or policies further a legitimate business need that cannot reasonable be achieved by means that are less disparate in their impact." In: FEDERAL TRADE COMMISSION. **Big Data, A Tool for Inclusion or Exclusion? – Understanding the Issues**. January, 2016, p. iii. Available at: <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf>. Access: January 12, 2019.

disparate treatment doctrine does not appear to do much to regulate discriminatory data mining,"[129] for intention and knowledge are usually required by courts for one to be considered liable. Disparate impact, however, is thought to be a more fruitful standard. The authors explain that under disparate impact "a plaintiff must show that a particular facially neutral employment practice causes a disparate impact with respect to a protected class."[130] Still, the defendant may claim the practice is covered by business necessity, in which case, the plaintiff can counter by showing alternative practices which render less discriminatory results than the one adopted by the employer.

Barocas & Selbst state that the business necessity standard is at the core of disparate impact doctrine. The criterion was first presented in *Griggs v. Duke Power Co.*[131] but has since become a much broader parameter:

> Some courts require that the hiring criteria bear a 'manifest relationship' to the employment in question or that they be 'significantly correlated' to job performance. (…) In a subsequent case, however, the Third Circuit recognized that Title VII does not require an employer to choose someone 'less qualified' (as opposed to unqualified) in the name of nondiscrimination and noted that aptitude tests can be legitimate hiring tools if they accurately measure a person's qualification. (…) Thus, all circuits seem to accept varying levels of job-relatedness rather than strict business necessity.[132]

Because of these changes, and also given the way data mining operates, the authors conclude that disparate impact doctrine seems a more logical tool to fight algorithmic discrimination under Title VII, but emphasize that it still poses significant hurdles and that Title VII itself requires reform to more directly and safely address algorithmic discrimination.

Recognizing the deficiency of current legislation, initiatives have emerged in the United States to fight algorithmic discrimination. A prominent example is New York City, where Instruction n. 1696 was approved to establish an automated decision systems task force, with the goal of issuing recommendations for, among others,

---

[129] BAROCAS, S. and SELBST, A. D. **Big Data's Disparate Impact.** California Law Review, vol. 671, 2016, p. 694. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899>. Access: January 01, 2019., p. 701.
[130] Ibidem.
[131] Griggs v. Duke Power Co., 401 U.S. 424 (1971).
[132] BAROCAS, S. and SELBST, A. D. **Big Data's Disparate Impact.** p. California Law Review, vol. 671, 2016, p. 704-705. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899>. Access: January 01, 2019.

> development and implementation of a procedure that may be used by the city to determine whether an agency automated decision system disproportionately impacts persons based upon age, race, creed, color, religion, national origin, gender, disability, marital status, partnership status, caregiver status, sexual orientation, alienage or citizenship status.[133]

3.2.2 Germany, informational self-determination, and the European Union

As seen above,[134] Germany adopted the indirect horizontal effects theory. As in the case of the United States, one could conclude this would make protecting against algorithmic discrimination be rather challenging. That is peculiarly untrue, however, for Germany possesses particular characteristics that mitigate the supposed barrier on fundamental rights applicability, namely (i) the explicit recognition of a fundamental right to personal data protection in EU law, (ii) the theory of informational self-determination, and (iii) directives and regulations from the European Union regarding personal data and discrimination.

Some of the particular characteristics of the German jurisdiction referred to above derive from Germany's membership in the European Union. The treaties and legal instruments of the EU impose obligations upon national states in many instances, some of which involve antidiscrimination. In the words of Ellis & Watson,

> The picture which emerges from a consideration of the numerous sources of EU equality and non-discrimination law is a complex one. There are a number

---

[133] Other provisions include: (a) Criteria for identifying which agency automated decision systems should be subject to one or more of the procedures recommended by such task force pursuant to this paragraph; (b) Development and implementation of a procedure through which a person affected by a decision concerning a rule, policy or action implemented by the city, where such decision was made by or with the assistance of an agency automated decision system, may request and receive an explanation of such decision and the basis therefor; (c) Development and implementation of a procedure that may be used by the city to determine whether an agency automated decision system disproportionately impacts persons based upon age, race, creed, color, religion, national origin, gender, disability, marital status, partnership status, caregiver status, sexual orientation, alienage or citizenship status; (d) Development and implementation of a procedure for addressing instances in which a person is harmed by an agency automated decision system if any such system is found to disproportionately impact persons based upon a category described in subparagraph (c); (e) Development and implementation of a process for making information publicly available that, for each agency automated decision system, will allow the public to meaningfully assess how such system functions and is used by the city, including making technical information about such system publicly available where appropriate; and (f) The feasibility of the development and implementation of a procedure for archiving agency automated decision systems, data used to determine predictive relationships among data for such systems and input data for such systems, provided that this need not include agency automated decision systems that ceased being used by the city before the effective date of this local law. In: THE NEW YORK CITY COUNCIL. **Automated decision systems used by agencies – Law number 2018/049**. January 11, 2018. Available at: <https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0>. Access: January 01, 2019.

[134] See section 3.1.2.

> of instruments to which a court must have regard in deciding an issue within this area, and the European judicature, in seeking to resolve ambiguities and unclear matters, must have recourse to many different instruments. Nevertheless, in the current state of the law, there is only a limited list of grounds on which EU law actually contains an outright prohibition on discrimination. These are nationality, sex, part-time and temporary employment, racial or ethnic origin, religion or belief, disability, age, and sexual orientation.[135]

These rights are determined in the EU's founding treaties, in the decisions set forth by Europe's higher courts, and in Directives 2000/43/EC, 2000/78/EC, 2004/113/EC and 2006/54/EC. All of these instruments establish rules against discrimination, be it in matters of employment, regarding access to the supply of goods and services, or more generally against ethnic or racial discrimination. A more comprehensive proposal, which if implemented would govern the principle of equal treatment irrespective of religion, belief, disability, age or sexual orientation was presented in 2008, but has not yet not been approved by the Council.[136]

The Charter of Fundamental Rights of the European Union is of distinct relevance to EU law, particularly regarding algorithmic discrimination.[137] Article 8 of the Charter recognizes a fundamental right to personal data protection, and more specifically states such data "must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law". In other words,

> It innovates to the extent that it establishes that the elements mentioned deserve to be protected as elements of a fundamental right deserving protection per se, and that the protection is not exclusively granted to data in a way or another related to the right to respect for private life, but to personal data in general. In

---

[135] ELLIS, E. and WATSON, P. **EU Anti-Discrimination Law - Second Edition**. Oxford EU Law Library, 2012 p. 22.

[136] The proposal states quite clearly that it intends to "set out a framework for the prohibition of discrimination on these grounds and establish a uniform minimum level of protection within the European Union for people who have suffered such discrimination," given that the current directives "applies only to employment, occupation and vocational training". For more, see: COMMISSION OF THE EUROPEAN COMMUNITIES. **Council Directive on implementing the principle of equal treatment between persons irrespective of religion or belief, disability, age or sexual orientation**. Brussels, 2008. Available at: <https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52008PC0426>. Access: January 01, 2019.

[137] For a broader understanding of the Charter and how the EU came to adopt a list of fundamental rights, see The Emergence of Personal Data Protection as a Fundamental Right of the EU, Gloria González Fuster, chapter 6. It worth noting, as pointed out by the author, that the chair of the Convention responsible for the drafting of the Charter was Roman Herzog, former president of Germany and of the GCC who was "peculiarly familiar with the German Federal Constitutional Court's case law on the right to informational self-determination". In: FUSTER, G. G. **The Emergence of Personal Data Protection as a Fundamental Right of the EU**. Springer International Publishing, 2014, p. 194.

> this sense, it goes beyond the scope of the protection granted on the basis of the ECHR, and of the common constitutional traditions of the Member States.[138]

There has been much discussion on the integration of the EU Charter into Member States' legal systems. The Charter itself refers to it, establishing in Article 52(4) that "[i]n so far as this Charter recognises fundamental rights as they result from the constitutional traditions common to the Member States, those rights shall be interpreted in harmony with those traditions."[139] In the case of Germany, the discussion was not as prominent as in other national jurisdictions for the country had long given prominence to data protection. More specifically, Germany has recognized the right to informational self-determination since 1983, when the GCC handed down a decision in a case about the reform of the Census Act. In the decision, the court determined certain provisions of the act were unconstitutional because individuals have the right to determine not only whether personal data about them may be disclosed, but also how it can be used.[140-141]

As Laura Schertel Mendes points out, the GCC identifies three components of informational self-determination:

> First, the power to decide is included in the protection [conferred by the court], so the individual may choose by herself on the collection and use of personal information. The second feature, that the fundamental right to informational self-determination does not encompass a fixed and predefined protection sphere, runs from the first and brings the right further away from a private sphere protection. Third, the reference to the person is decisive in determining protection, as each piece of information that is considered personal is entitled to protection.[142]

---

[138] Ibid, p. 205.

[139] The debate ended when article 6(1) of the Treaty of the European Union (TEU), also known as the Lisbon Treaty, stated that "The Union recognises the rights, freedoms and principles set out in the Charter of Fundamental Rights of the European Union of 7 December 2000, as adapted at Strasbourg, on 12 December 2007, which shall have the same legal value as the Treaties." In: EUROPEAN COMMISSION. **Charter of Fundamental Rights: the Presidents of the Commission, European Parliament and Council sign and solemnly proclaim the Charter in Strasbourg**. Brussels, December 12, 2007. Available at: <http://europa.eu/rapid/press-release_IP-07-1916_en.htm>. Access: January 07, 2019.

[140] BVerfGE 65,1, Volkszählung.

[141] It is worth noting that the Brazilian GDPA included informational self-determination among its founding ideas, in Article 2, II.

[142] MENDES, L. S. **Habeas Data e autodeterminação informativa: os dois lados de uma mesma moeda**. In: Centro de Direito, Internet e Sociedade do Instituto Brasiliense de Direito Público (CEDIS/IDP) (Orgs). *Internet & Regulação* Saraiva. Forthcoming.

While informational self-determination is more generally concerned with protecting privacy than specifically aimed at fighting discrimination, the protection addresses the fluid nature of the individual's protected sphere and gives the individual the final word on whether data about herself may be used, and for which purposes. This makes it possible for German citizens to fight much of the discriminatory potential of algorithmic using Constitutional provisions, the current legislation, and understandings of the GCC.

Personal data protection received another boost after the approval of the General Data Protection Regulation (GDPR), which came into force in 2018. The GDPR replaced the former 1995 Directive on Data Protection – with the notable difference that under EU law, regulations, unlike directives, are immediately enforceable and do not depend on national legislators for implementation. The GDPR was certainly the object of much debate in both the policymaking and private sectors, and was largely considered a step forward for personal data protection. When it comes to discrimination, however, the advances are less categorical. The word only appears once in the text, in Recital 71,[143] though other provisions scattered throughout the regulation can be used to protect individuals from discriminatory practices.

[143] Recital 71 states: "The data subject should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing and which produces legal effects concerning him or her or similarly significantly affects him or her, such as automatic refusal of an online credit application or e-recruiting practices without any human intervention. Such processing includes 'profiling' that consists of any form of automated processing of personal data evaluating the personal aspects relating to a natural person, in particular to analyse or predict aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, where it produces legal effects concerning him or her or similarly significantly affects him or her. However, decision-making based on such processing, including profiling, should be allowed where expressly authorised by Union or Member State law to which the controller is subject, including for fraud and tax-evasion monitoring and prevention purposes conducted in accordance with the regulations, standards and recommendations of Union institutions or national oversight bodies and to ensure the security and reliability of a service provided by the controller, or necessary for the entering or performance of a contract between the data subject and a controller, or when the data subject has given his or her explicit consent. In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision. Such measure should not concern a child.
In order to ensure fair and transparent processing in respect of the data subject, taking into account the specific circumstances and context in which the personal data are processed, the controller should use appropriate mathematical or statistical procedures for the profiling, implement technical and organisational measures appropriate to ensure, in particular, that factors which result in inaccuracies in personal data are corrected and the risk of errors is minimised, secure personal data in a manner that takes account of the potential risks involved for the interests and rights of the data subject and that prevents, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health

Goodman claims there are two key principles that can be used to tackle algorithmic discrimination in the GDPR: data sanitization and algorithm transparency. The first is set out in Article 9[144] and determines "the removal of special categories from datasets used in automated decision making,"[145] whereas the second is enshrined in Articles 13(2)(f)[146] and 14, determining that "meaningful information about the logic involved, as well as the significance and the envisaged consequences" of automated decision-making must be provided for data subjects.[147]

Articles 21 and 22 add to the mix. Article 21 establishes a right to object, stating that the data subject can always object against the processing of her personal data "on grounds relating to his or her particular situation," and Article 22 determines individuals' rights in the context of automated decision-making. In such circumstances, the data subject can demand exclusion from automated decisions "which produces legal effects concerning him or her or similarly significantly affects him or her." As discriminatory outcomes easily fall among those that have

status or sexual orientation, or that result in measures having such an effect. Automated decision-making and profiling based on special categories of personal data should be allowed only under specific conditions."

[144] Article 9(1) states that "Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited." Paragraph 2 contains exceptions to this general rule, including cases in which explicit consent has been given, the data has been made public by the data subject herself, processing is required for adequate healthcare, and so forth.

[145] GOODMAN, B. W. **A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection**. 29th Conference on Neural Information Processing Systems. Barcelona, Spain. 2016, p. 2.

[146] Article 13(2): "In addition to the information referred to in paragraph 1, the controller shall, at the time when personal data are obtained, provide the data subject with the following further information necessary to ensure fair and transparent processing: (f) the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject."

[147] It is worth noting Goodman claims both data sanitization and algorithm transparency are insufficient tools to fight discrimination. Pertaining to sanitization, he states that "In short, *prohibiting the collection or processing of data revealing special category membership may worsen the problem it is intended to solve*. The methods proposed for identifying and reducing discrimination in algorithms are only effective if special category membership is indicated in the dataset (Feldman et al., 2015). Furthermore, variables with no theoretical interpretation whatsoever may be highly indicative of special category membership. However, there is no way to establish whether this is or is not the case without information about special category membership. Eliminating the collection of data revealing sensitive categories may, perversely, allow discrimination to continue and deepen by making it impossible to be detected in the first place." In: GOODMAN, B. W. **A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection**. 29th Conference on Neural Information Processing Systems. Barcelona, Spain. 2016, p. 3.

significant effects, automated profiling is a field where Germans enjoy considerable levels of protection.[148]

3.2.3 Brazil, the Right to Equality, and the GDPA

The circumstances of Brazil are very particular when it comes to antidiscrimination law. Article 5, *caput* of the Brazilian Constitution states that everyone is equal under the law, without distinction of any nature, granting to Brazilians and foreigner residents alike the inviolable rights to life, freedom, equality, safety, and property.[149]

Ordinary legislation, for its part, tackles the issue rather unsystematically. Rios & Silva mention some instruments which have provisions addressing discrimination, namely (i) Law 12,288/2010, also known as the Racial Equality Act, (ii) Law 7,716/1989, which criminalizes racial, ethnical, or religious prejudice, (iii) Law 12,711/2012, a statute establishing quotas for access to federal universities and federal higher education institutions, and (iv) Law 9,029/1995, which prohibits discriminatory practices in labor relations.[150]

Aiming specifically at algorithmic discrimination, other instruments not directed exclusively towards equality are pertinent, namely the Consumer Protection Code, the Credit Information Act, the Public Information Access Act, the Brazilian Internet Framework, and the recently approved GDPA. The relevance of these instruments, as will become clear, steams from the protection they extend to personal data, and, in certain cases, from provisions that can be used to fight discrimination.

3.2.3.1 The Consumer Protection Code

The first instrument in Brazil that directly addressed personal data is the Consumer Protection Code (Law 8,078/1990 or CDC, for its Portuguese acronym). The CDC arranged a

---

[148] More on the GDPR and its impacts for algorithmic discrimination will be discussed in section 4.2.

[149] Original in Portuguese reads: "Todos são iguais perante a lei, sem distinção de qualquer natureza, garantindo-se aos brasileiros e aos estrangeiros residentes no País a inviolabilidade do direito à vida, à liberdade, à igualdade, à segurança e à propriedade."

[150] RIOS, R. R. and SILVA, R. da. **Democracia e direito da antidiscriminação: interseccionalidade e discriminação múltipla no direito brasileiro**. Cienc. Cult., São Paulo, vol. 69, n. 1, p. 44-49, March, 2017. Available at: <http://cienciaecultura.bvs.br/scielo.php?script=sci_arttext&pid=S0009-67252017000100016&lng=en&nrm=iso>. Access: January, 01, 2019

framework to address privacy and data protection demands through principle-based norms that are broad enough to offer solutions to many new conflicts related to information technology, including discrimination. Doneda & Mendes summarize it this way:

> Four pillars of the Brazilian consumer protection system explain how it could promote and enforce data protection standards: a) specific regulations for consumer databases that address the rectification and notice process; b) a broad clause governing damage claims (overall liability); c) a public consumer redress structure, which includes both an administrative and a judicial system of redress (small claims courts); and d) a broad conceptualization of who are consumers.[151]

The central provision regarding personal data in the CDC is Article 43, which provides for specific rights and safeguards regarding personal information stored in databases, namely: (a) a right of access to all of such personal information; (b) the principle of data quality, according to which all stored data must be objective, accurate and presented in a comprehensible language; (c) a right to written notification before any negative personal information is stored; (d) a right to rectification of any inaccurate data stored and (e) a term for storage limits –a maximum of five years – of negative personal information.

Regarding discrimination, the CDC has one further provision that must be noted, Art. 6, II. This provision affirming the consumers' rights to freedom of choice and to equality in hiring services or purchasing goods was the basis for the first case against algorithmic price discrimination in Brazil. The case was brought forth by the National Consumer Secretariat (SENACON for its Portuguese acronym) against the travel website Decolar.com. The Department for Consumer Protection and Defense concluded that the economic reasoning behind the practices of geopricing and geoblocking,[152] which the website used to differentiate its pricing among Brazilian and foreigner consumers, was unconvincing and that the practices were abusive in terms of Article 39 of the CDC.[153] In their words,

---

[151] DONEDA, D. and MENDES, L. S. **Data Protection in Brazil: New Developments and Current Challenges**. In: GURWIRTH, S., LEENES, R. and HERT, P. De. (Eds). *Reloading Data Protection: Multidisciplinary Insights and Contemporary Challenges*. Springer, 2014, p. 3-20.

[152] Geoblocking happens when a given functionality or offer is blocked due to the user's or consumer's location. Geopricing, for its turn, consists in differentiated pricing based on the location of the user.

[153] The article identifies practices considered abusive by the legislation. It is worth noting that SENACON's decision was made under the purview of its administrative agency prerogatives.

> the expression 'fair' [in Article 39, X] refers to impartiality, righteousness, conformity to reason, and it is certain that justice must be achieved through the application of material equality, meaning unequals should be treated unequally, in the measure of their inequality.[154]

The Secretariat claimed it could not identify valid reasons for differentiation, and that those mentioned by Decolar.com – currency exchange rates and the differences between the legal systems which govern Brazilian and foreign consumers – did not constitute valid reasons for discrimination.

The case is the first of its kind and no final word has been issued by a court of law yet to provide us with a better idea on how these rights will be interpreted in light of algorithmic discrimination. Yet there is also a class action lawsuit against Decolar.com brought forward by the Prosecutor's Office in Rio de Janeiro that may precipitate such a decision.

Some aspects of the Prosecutor's arguments deserve careful consideration. Much like the Secretariat, the argument against the company's practices relies heavily on the CDC, and more specifically on Art. 6, II, but the allegations go much further. The application initiating the proceedings expressly claims that geo-discrimination is a practice that must be fought, *just as any other form of discrimination*.[155] The Prosecutor's Office further states this is a conclusion drawn not only from the CDC, but also (and primarily) from the principle of equality as stated in the Brazilian Constitution.[156]

---

[154] MINISTÉRIO DA JUSTIÇA. **Nota Técnica n. 92**. 2018, §39. Available at: <http://www.cmlagoasanta.mg.gov.br/abrir_arquivo.aspx/PRATICAS_ABUSIVAS_DECOLARCOM?cdLocal=2&arquivo=%7BBCA8E2AD-DBCA-866A-C8AA-BDC2BDEC3DAD%7D.pdf>. Access: January 09, 2019.

[155] Original in Portuguese reads: "Ambas as práticas impugnadas constituem meios de diferenciar arbitrária e injustificadamente os consumidores, o que é vedado pela legislação consumerista. Trata-se, aqui, da geodiscriminação (geo-discrimination), prática que deve ser combatida, tal como qualquer outra forma de discriminação."
5ª PROMOTORIA DE JUSTIÇA DE TUTELA COLETIVA DE DEFESA DO CONSUMIDOR E DO CONTRIBUINTE DA CAPITAL. **Petição inicial, Inquérito Civil n. 347/5ª PJDC/2016**, §29.

[156] Ibid, §30.

3.2.3.2 The Credit Information Act

Although the CDC and the broad understanding it confers to consumer relations in Brazilian law are relevant[157], the CDC does not fully apply to all scenarios. One instances of potential algorithmic discrimination, the collection of so-called "positive information" – data which results from the processing of borrowers' payment histories – is a prime example. The legislation that addresses the issue is the Credit Information Act (Law 12,414/2011). The Act furnishes detailed regulation concerning credit information databases and establishes a legal framework that simultaneously encourages data flow and protects users' personal data.

The Credit Information Act lays out a variety of rules ranging from the creation of a payment history to the establishment of responsibilities in case of damages. It determines, for example, when a person's payment history can be generated (Art. 4), what information can be stored (Art. 3, §2 and §3), what the rights of the data subject are (Art. 5), what the duties of the data processor are (Art. 6), who supervises the databases (Art. 17) and who is liable in case of damages (Art. 16).[158]

Consumer consent is the cornerstone of the current Credit Information Act, as provided by Article 4.[159] On the basis of this principle, the law confers the consumer the prerogative over the creation, transfer and cancellation of her credit history. Moreover, according to Article 5, consumers shall obtain the cancellation of their records upon request and, as determined by Article 9, the sharing of information is allowed only if expressly authorized by the consumer. Similar to the CDC, the Credit Information Act prescribes the principle of quality or accuracy of personal data (Art. 3, §1) and the rights to access, rectification and cancellation of data (Art. 5, II and III). In addition, it guarantees consumer access to the main criteria used in the credit rating process; that is, the consumer has the right to know which criteria are used to calculate

[157] A consumer can sue for damages from the firm with which she has a contract, as well as exercise the rights to correction and disclosure against the party responsible for a database. For this reason, the data protection norms of the CDC are much more broadly applied beyond contractual consumer relations.

[158] Many of its norms correspond to the principles provided in Convention 108 of the Council of Europe and in the European Directive 95/46/EC, but it can also be said that the Credit Information Act resemble typical U.S. regulations on credit reporting.

[159] The original in Portuguese reads: "A abertura de cadastro requer autorização prévia do potencial cadastrado mediante consentimento informado por meio de assinatura em instrumento específico ou em cláusula apartada."

credit risk (Art. 5, IV). Concerning risk assessment, the law ensures the right to review of any decision made exclusively by automated means (Art. 5, VI).[160]

The Credit Information Act also offers an explicit legal basis for the purpose limitation principle, a principle that was only implicit under the CDC. The principle of purpose, which in the GDPA also gained prominence, is extended through the entire credit information system. First, the Act defines the strict scope of its own application, which solely covers the risk assessment databases in credit and commercial transactions (Art. 2, I). Second, it establishes the right of the data subject to limit the processing of personal information to the original purposes of collection (Art. 5, VII). Third, Article 7 defines the legitimate purposes for the data collected under the Act: for risk analysis or for assistance making decisions to grant credit or engage in other commercial transactions that involve financial risk. In other words, the information gathered in these databases cannot be used for marketing or any other activity not explicitly provided for in the law.[161]

As stated in section 2.3.3, the Credit Information Act has a further prohibition against the storage or use of sensitive and excessive information, as provided by Article 3, §3. Pursuant to this norm, excessive information is defined as information unrelated to the credit risk analysis. Sensitive information, for its part, is defined as information that relates to social or ethnic origin, health, genetic information, sexual orientation or political, religious and philosophical beliefs. The prohibition on storing or using it certainly contributes to preventing some types of information from being used for profiling, discrimination or the violation of the principle of equality.[162]

The limits and the reach of the Credit Information Act have been analyzed by the Superior Court of Justice (STJ for its Portuguese acronym) on a few occasions. One such ruling,

---

[160] This is very much in line with Article 20 of the GDPA.

[161] In this context, another similarity to the European Directive can be found, particularly Article 6, 1, b, which determines that personal data should be "collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes."

[162] Another parallel to the European Directive can be drawn here, namely in reference to Article 8, which concerns the processing of special data categories. In the GDPA, as mentioned, the provision on Art. 5, II defines sensitive data.

REsp n. 1.419.697, is particularly relevant for the purposes of this dissertation. In this case, credit scoring was deemed compatible with Brazilian law as a licit commercial conduct under Articles 5, IV and 7, I, of the Credit Information Act. However, the decision specified that the limits imposed by the CDC regarding privacy and transparency of contractual relations must be respected by the credit risk evaluation. It further established that the use of excessive or sensitive information is cause for damages (Art. 3, I and II) and that refusing credit based on incorrect data will result in strict solidary liability between the service supplier and the person responsible for the database (Art. 16).

Another relevant aspect of this decision is the section dedicated to privacy protection for and transparency in consumer information, as provided by the Credit Information Act (Art. 3) and the CDC (Art. 43). The limitations imposed by these norms were expressed as five duties imposed upon the service supplier: veracity, clarity, objectivity, prohibition of excessive information, and prohibition of sensitive information. These duties, though not directed towards discrimination, will likely prevent much discriminatory behavior, and may also provide useful guidelines to which the courts could turn when tackling this issue – to see which criteria, for instance, were considered objective to identify potentially discriminatory behavior. They might also assist the court to determine which practices should be considered subjective.

Five theses summarizing the ruling were suggested by the Reporting Justice and subsequently unanimously approved by the Second Section of the Court, namely: 1) “The credit scoring system is a method developed to evaluate the credit concession risk, based upon statistic models that consider diverse variables, with the attribution of a score to the evaluated consumer (credit risk score)”; 2) “This commercial practice is licit, being authorized by Art. 5, IV and Art. 7, I, of the Law n. 12.414/2011 (Credit Information Act)”; 3) “In the credit risk evaluation, the limits imposed by the consumer protection system in respect to privacy safeguards and the maximal transparency of contractual relations must be observed, in accordance with the Consumer Protection Code provisions and Law n. 12.414/2011”; 4) “Despite consumer’s consent being unnecessary, clarifications must be given, if requested, about the considered data source (credit history), as well as about the personal information evaluated”; 5) “Failure to

observe these legal limits in using the credit score system configures an abusive exercise of rights (Art. 187 of the Civil Code), and may give rise to strict solidary liability between the service supplier and the responsible for the database, the source and the consulting (Art. 16 of Law n. 12,414/2011), for the occurrence of damages in the use of excessive or sensitive information (Art. 3, I and II of Law n. 12,414/2011) or in situations where credit is refused based on outdated or incorrect data".

It is worth noting that the Credit Information Act is currently under reform in the Brazilian Congress. Bill 441/2017 would change the current rules and make three central modifications. The first one concerns consumers' consent as a requirement to open a consumer file and process positive information – it changes the system from an opt-in to an opt-out model, meaning consumers no longer have to declare they want to be a part of the database to be included, rather they must declare they do not want it and have their information removed. The second modification is the suppression of the agents' liability clause. The third modification affects the Bank Secrecy Law, which is meant to facilitate the flow of financial information between databases controlled by different agents.

The main objective is undoubtedly to increase the number of (positive) consumer entries in the credit reporting system, since there are currently only five million entries in a potential universe of one hundred million consumers.[163] In the Temer government view, one subscribed by the financial institutions as well, low adherence to the system is due to the high number of bureaucratic requirements which must be met to create a credit report and to consumer inertia, since the current model provides for an opt-in system, in which the express consent of the consumer is necessary to open a consumer file.[164]

[163] MELLO, J. M. P., MENDES, M. and KANCZUK, F. **Cadastro Positivo e democratização do crédito**. Folha de São Paulo, March, 2018. Available at: <https://www1.folha.uol.com.br/opiniao/2018/03/joao-manoel-pinho-de-mello-marcos-mendes-e-fabio-kanczuk-cadastro-positivo-e-democratizacao-do-credito.shtml>. Access: January 05, 2019.
[164] Ibidem.

3.2.3.3 The Public Information Access Act

The Public Information Access Act (Law n. 12,527/2011 or LAI, for its Portuguese Acronym) is a federal statute binding on all public agents, including the judiciary, public prosecutor's offices, public defender's offices and so forth. Non-profit organizations that finance their activities using public funds are also submitted to the same rules in relation to the activities carried out using those public funds (Decree n. 7,724/2012 regulates the offices responsible for providing information in such cases).[165]

The legislation makes public transparency of data the default and secrecy the exception (Art. 3, I), and it establishes a comprehensive list of citizen rights. It understands public information broadly such that it includes data about individuals – including third parties – and data held by public agents that is not necessarily collected or treated by these public actors. Article 7 also explicitly affirms that information related to the implementation, monitoring and results of public programs, projects, and actions, as well as goals and indexes all fall within the category of information to which individuals should have access.

As emphasized by the Office of the Comptroller General, the Act takes transparency as its ground rule and demands it be respected by public agents in both its active and passive facets: "Active transparency is understood as the proactive and spontaneous delivery of information by the State,"[166] whereas "passive transparency depends on a citizen's request. It takes place, therefore, by means of a request for information."[167]

The LAI also protects individuals by establishing that the information request need not be accompanied by an explanation of why the data is required or for what purposes it will be used (Art. 10, §3). It provides a predetermined appeal structure in Articles 15 till 20 to ensure

[165] For more detailed information on the LAI, see the document regarding the applicability of the Act in the Federal Public Administration by the Ministry of Transparency, Monitoring, and Office of the Comptroller General. In: MINISTÉRIO DA TRANSPARÊNCIA, FISCALIZAÇÃO E CONTROLADORIA-GERAL DA UNIÃO. **Aplicação da Lei de Acesso à Informação na Administração Pública Federal**. 2a. ed. rev., atu. e amp. Brasília. 2016. Available at: <http://www.acessoainformacao.gov.br/central-de-conteudo/publicacoes/arquivos/aplicacao_lai_2edicao.pdf>. Access: January 05, 2019.
[166] Ibid, p. 52.
[167] Ibid, p. 54.

that citizens may challenge a specific decision and ask for its review by the authorities and requires that personal data should be preserved and protected by the authorities in accordance with intimacy and privacy rules (Article 31).

3.2.3.4 The Brazilian Internet Framework

The Brazilian Internet Framework (Law n. 12,965/2014 or MCI for its Portuguese acronym), for its part, establishes plurality and diversity among its goals (Art. 2, III), and also places privacy and personal data protection as part of its founding principles (Art. 3, II and III). It mostly speaks of discrimination, however, when establishing the principle of net neutrality.

Net neutrality is covered in detail in Article 9 of the MCI, which emphasizes that it is the duty of the party responsible for the transmission of data (in its broadest sense, including switching, routing and so forth) to treat any and all data packages equally, "without differentiation regarding content, origin or destiny, service, terminal, or application."[168] Discrimination is only permissible when due to indispensable technical requirements needed for the services and applications to be provided or to the prioritization of emergency services.

Decree 8,771/2016 provided clearer rules in this regard. Article 5, §1 states the technical requirements mentioned in the MCI and gives the National Communications Agency power to oversee potential infractions. Article 9, for its turn, determines the commercial practices banned by the instrument.

3.2.3.5 The General Data Protection Act

The most recent development in this landscape that more directly attacks algorithmic discrimination is the GDPA. Though the Act will only come into force in 2020, and even when it does there will remain issues that will likely only be resolved with effective enforcement, it establishes a direct non-discrimination provision in Article 6, IX, which forbids any treatment of personal data with "illicit or abusive discriminatory goals."[169] The GDPA has two other

[168] BRAZIL. **Lei no 12.956**. April 23, 2014, art. 9, *caput*. Available at: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm>. Access: January 05, 2019.

[169] The original in Portuguese reads: "As atividades de tratamento de dados pessoais deverão observar a boa-fé e os seguintes princípios: IX - não-discriminação: impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos."

provisions, in many ways similar to the European regulations, that also reflect a concern for discriminatory practices and provide data subjects with mechanisms to enforce their rights, Articles 20 and 21:

> Art. 20. The data subject has the right to request review, by a natural person, of decisions taken solely on the bases of automated processing of personal data that affects her/his interests, including decisions intended to define her/his personal, professional, consumer or credit profiles or aspects of her/his personality.
> §1 Whenever requested to do so, the controller shall provide clear and adequate information regarding the criteria and procedures used for an automated decision, subject to commercial and industrial secrecy.
> §2 If there is no offer of information as provided in §1 of this article, based on commercial and industrial secrecy, the national authority may carry out an audit to verify discriminatory aspects in automated processing of personal data.
> Art. 21. Personal data concerning the regular exercise of rights by the data subject cannot be used to her/his detriment.[170]

No clarity on the enforcement of these provisions is likely to come before the instrument is implemented, but some issues can already be brought forward. First and foremost, the extent to which goals shall be considered abusive or illicit as per the principle of non-discrimination. As the debate in chapter 2 clarified, often times discrimination takes place despite developers' intent; in other words, algorithmic discrimination is particularly prone to occur without it being the objective of those developing or applying the algorithmic system, which questions the effectiveness of the principle in such scenarios. To ensure applicability, a well-grounded understanding on why goals may be abusive even without developers' or operators' intent must be proposed – and in that sense sanctions must also be applied mindfully, to place responsibility where responsibility is in fact due, i.e. an operator may be unaware of the full extent of a given algorithm's biases, and likewise a developer may not have complete control of an algorithm's response to a scenario it was not originally intended to be applied to. Ensuring sanctions are not disproportional is paramount for any public policy to succeed and for the sanctions themselves to achieve their true objective: incentivizing compliance.

[170] On December 28th, 2018, the Brazilian government issued Executive Order 869/2018 (known as Medida Provisória in Portuguese, or MP, a type of act issued by the Presidency that comes into force immediately but must be confirmed by Congress before officially becoming law, otherwise its effects are reversed). This MP changes the GDPA, including Article 20. More specifically, it excludes paragraphs 1 and 2, and also no longer requires that review of automated decisions must be carried out by a natural person. I will further debate the impacts of the proposed changes in 4.2 below.

Second, the provisions of Articles 20 bring to light the debate about what precisely is a "totally automated" decision. The idea of the law seems to be that whenever a person is entangled in the decision-making process, the risk presented by automation is mitigated, and thus the article should not apply. The difficulty lies in identifying when a human being is part of the process in any meaningful way that indeed merits such process to not present the potential of harm article 20 is aimed at attacking.

A third and connected issue is clarifying what the review by a natural person shall entail. It is not evident what the process of such review must be for the requirement to be considered fulfilled, nor is it clear what level of transparency the reviewer must comply with – for example, will an authority oversee the process, or is the declaration that a review has been carried out enough?

Lastly, what the trigger of "protected interests" will be is crucial for effective application. The provision opens up a wide range for interpretation, mentioning personal, professional, consumer and credit profiles, as well as aspects of one's personality. It remains to be seen, however, what authorities will understand such profiles to encompass, and to what extent an automated decision will be considered to "affect" this profiling.

Implementation will be complicated by the last-minute removal from the bill of the provision that established a centralized data protection authority, leaving no certainty as to how enforcement will be carried out. The issue was further difficulted when Executive Order n. 869/2018 was issued, establishing a new authority in a model that experts have described as far less independent than that originally envisioned. Still, the order is pending approval by Congress, and can be struck down. In that case, it is unclear if only the judiciary will be responsible for enforcement, or if other already established agencies such as SENACON will handle the administrative proceedings. In any case, this is an issue of fundamental importance, for inefficient, or loose enforcement will indisputably affect individuals and the private sector as a whole.[171]

[171] More on the GDPA and the creation of a centralized authority in section 4.2.

3.3 Case Studies

This last item in chapter 3 is a more thorough investigation of the hurdles in enforcement. These hurdles are illustrated using concrete instances of algorithmic use in unemployment services and credit scoring. The cases were chosen for they illustrate two facets of the debate: algorithmic discrimination by public actors, on the part of unemployment, and by private agents, on the part of credit scoring. They will enable us to scrutinize the Brazilian legislation and our incursion into legislative instruments from other jurisdictions will remind us how the problems are tackled elsewhere.[172] Here I will expressly refer to the typology presented in chapter 2 and question whether the problems identified in that section can be solved in these concrete cases using current legal instruments.

3.3.1 Labor and unemployment in Poland

In 2014, Poland decided to change its public unemployment policy. Among other modifications,[173] algorithmic profiling of the unemployed was adopted in order to determine the level of support each beneficiary would receive, with the explicit goal of "counteract[ing] unemployment more effectively, increase[ing] the efficiency of labor offices and guarantee[ing] public services of a higher quality."[174] The Ministry of Labor and Social Policy (MLSP) also announced that the introduction of profiling was aimed at adjusting the policies to each beneficiary by individualizing assistance.[175] The Panoptykon Foundation,[176] an organization established by a group of lawyers with the express goal of protecting fundamental rights and freedoms, examined several aspects of the new policy.

---

[172] I will focus on Brazilian legislation though one of the cases took place in Poland because there are to date no known cases of public algorithmic discrimination in Brazil which led to relevant impact on individuals – an investigation on whether such cases exist is beyond the scope of this dissertation but would certainly render relevant academic contributions. I am aware of the limitations of this approach, but given that the discussion on algorithmic discrimination in Brazil is still at its infancy, the solution was necessary for the objectives of this dissertation.

[173] The labor agencies were somewhat modified and new forms of assistance were made available to the unemployed, among other changes. In: NIKLAS, J. SZTANDAR-SZTANDERSKA, K. and SZYMIELEWICZ, K. **Profiling the Unemployed in Poland: social and Political Implications of Algorithmic Decision Making.** Fundacja Panoptykon. Warsaw, 2015.

[174] Ibid, p. 7.

[175] Ibid.

[176] In its own words, "The Panoptykon Foundation was established in April 2009 upon the initiative of a group of engaged lawyers, to express their opposition to surveillance. Our mission is to protect fundamental rights and freedoms in the context of fast-changing technologies and growing surveillance." For more see: <https://en.panoptykon.org/about>. Access: January 08, 2019.

Panoptykon emphasizes that as a result of the new policy three categories of unemployment status were established to which beneficiaries were assigned on the basis of a questionnaire.[177] In it, a total of 24 questions were put to the beneficiary: 12 to determine the individual's "distance from the labor market," 11 to assess one's "readiness to enter or return to the labor market", and one question that addressed both issues. Based on the answers, the unemployed is either categorized as pertaining to Profile I – the category of people who do not have serious life problems and who are adequately qualified for the job market; or Profile II – individuals with some professional skills but who either worked for a very long time at a single company and therefore are not entirely confident in their capacity to find a new job, or possess skills that are not currently needed in the labor market; or lastly to Profile III – people with serious life problems, often classified as passive individuals who either lack basic education or prefer unemployment.[178]

Access to social assistance policies was determined by the profile to which the beneficiary was assigned, meaning a person in Profile I had access to different services people in Profile II or III. The difference is services is quite significant:

> According to legal provisions those qualified to the third profile may be granted 10 types of forms of assistance – including being assigned to the Program of Activization and Integration (PAI) or a special program, or being directed to work in a social cooperative. (…) Nevertheless, these forms of support are costly and difficult to be organized, and in effect, labor offices unwillingly launch them. This is confirmed by the statistics according to which as many as 38% of labor offices do not organize any of these programs. In such a situation persons belonging to Profile III actually may not be offered any attractive form of assistance.[179]

Panoptykon's research led it to conclude that this policy was problematic and, instead of reaching the objectives set out by the government, it actually worsened the situation of the

[177] As made clear in the study by Panoptykon, there are problems and discussions that warrant attention with regards individual consent for inclusion in such processes. This dissertation will not go deeper into these aspects, but further debate can be found at NIKLAS, J. SZTANDAR-SZTANDERSKA, K. and SZYMIELEWICZ, K. **Profiling the Unemployed in Poland: social and Political Implications of Algorithmic Decision Making.** Fundacja Panoptykon. Warsaw, 2015, p. 12.

[178] These are the words of the MLSP. Panoptykon obtained a handbook which serves as guidelines for the labor officers and pages 22 till 24 cover what is understood to be Profile III.

[179] NIKLAS, J. SZTANDAR-SZTANDERSKA, K. and SZYMIELEWICZ, K. **Profiling the Unemployed in Poland: social and Political Implications of Algorithmic Decision Making.** Fundacja Panoptykon. Warsaw, 2015, p. 13.

unemployed and had discriminatory outcomes, both directly and indirectly.[180] The foundation also made clear that the practices of the MLSP had led to an illusion of standardization that bore no resemblance to reality, for the officers who carried out the interviews did so in drastically different manners:

> This supposedly standardized process in practice is carried out in a very different way, when it comes to such basic features as the way of posing questions and interpreting the unemployed person' replies. Contrary to the handbook [from MLSP], some counselors show the unemployed standardized responses during the interview, read some of them in the case of more ambiguous questions or at least suggest possible answers, while others simply select certain options in the computer system according to their own assessment, without verifying whether the unemployed has understood the question and whether the selected option fully reflects what s/he meant while answering the question.[181]

Without neglecting the particular circumstanced of each case, virtually all of the types of algorithmic discrimination described in chapter 2[182] are likely to occur in this scenario. First, the danger of discrimination by faulty collection or design is present, as Panoptykon emphasizes, because of the varying manners by which public officials collect the information. The variation in collection methods increases the chances that the data is unreliable. Second, the risk of discrimination by reproduction was also verified. The algorithm developed by the Polish government was not equally representative of all populations, and tended to reproduce socially established biases; to give one specific example, the notion that single mothers are unfit for the job market.[183] Third, discrimination by correlation is also a real possibility. It is quite clear that the algorithm works off specific characteristics as proxies but is not particularly concerned with the possibility that these characteristics mean different things to different individuals. For example, when a beneficiary identified themselves as a single parent, the algorithm only asked if the individual had access to childcare, ignoring the many different circumstances surrounding

---

[180] Ibid p. 21: "The former [direct discrimination] means that a person is treated in a worse manner than another person in a comparable situation only because e.g. she is a woman. On the other hand, indirect discrimination is understood as applying seemingly neutral criteria which in fact lead to the creation of a situation unfavorable for a given person due to e.g. disability or age."
[181] Ibid, p. 27.
[182] Section 2.3.3.
[183] NIKLAS, J. SZTANDAR-SZTANDERSKA, K. and SZYMIELEWICZ, K. **Profiling the Unemployed in Poland: social and Political Implications of Algorithmic Decision Making.** Fundacja Panoptykon. Warsaw, 2015, p. 37.

the question. Perhaps, for example, the beneficiary has no one who could take care of the child at the moment, but would be able to find such an individual if necessary for employment. Or perhaps they could find someone who could take care of the child part-time, and as such allow for a part-time job, or one of many other potential scenarios the algorithm neglects.

Fourth, a significant risk of discrimination by use of sensitive information is also present. It goes without saying that much of the information requested by the Polish government was sensitive. Panoptykon points out that the unemployed are left little choice over whether or not to deliver the requested information to the authorities, because the consequence of non-consent is denial of unemployment assistance. Fifth, discrimination by association with the fulfillment of rights is equally feasible. Again, although the algorithm was originally designed to provide people with individualized service, the risk of impaired fulfilment of rights is clear. Further examination would be needed to determine with certainty whether the classification presented was dependent on endogenous characteristics, and whether historically discriminated-against groups are affected. Still, the last question can be answered affirmatively with some confidence as several historically discriminated against groups that suffer from the algorithmic system's classification: single parents, handicapped, and the illiterate, among others.[184] The first question also begs a positive response as endogenous characteristics are indeed relevant for the algorithm – i.e., physical distance from the job market is considered a relevant trait, but it is a trait that is directly reflected in that person's lack of a job. Because the individual does not have a job, she may well have to live in peripheral zones where housing is cheaper. The algorithm treats this information as reinforcing the beneficiary's inability to obtain employment and feeds it back into the system, magnifying the problem.[185]

It should be remembered that this is a case of public algorithmic discrimination, meaning the agent carrying out the potentially discriminatory practice is an officer of the State. It would

[184] Ibid.

[185] Annex I of Panoptykon's report clarifies this point by stating what the question and the answer regarding this point are. The officer shall inquire about the applicant's place of residence in terms of distance from potential workplaces, and the answers range from "urban agglomeration" to "village or settlement significantly distant from the labor market." See in: NIKLAS, J. SZTANDAR-SZTANDERSKA, K. and SZYMIELEWICZ, K. **Profiling the Unemployed in Poland: social and Political Implications of Algorithmic Decision Making.** Fundacja Panoptykon. Warsaw, 2015, p. 44.

therefore be possible to claim that the fundamental right to equality is violated – and from a Brazilian standpoint also the right to work from the Brazilian Constitution, stated in Article 6, *caput*. Yet, such claims would likely require a detailed explanation of the violation of the rights in question, and, because of the novelty of the debate, it is difficult to predict how Brazilian courts would react.

The use of ordinary legislation to address the problems present in the Polish scenario might prove equally challenging in Brazil. The CDC would not apply, for not even the very broad Brazilian understanding of what comprises a consumer relationship would encompass this kind of public service.[186] Also, it is quite clear that this is unrelated to credit scoring, and thus the Credit Information Act would not be applicable. We would be left with the LAI, which could be of use to require transparency for the decisions taken by the algorithm, and with the GDPA, particularly its principled-norm of non-discrimination and its Articles 20 and 21.[187]

Art. 3, II of the LAI establishes that information of public interest should be made available, whereas Article 7, §3 states that public access to the grounds for decisions issued by public actors is always paramount. In that light, the statute could be used by individuals affected by public algorithmic decision-making to obtain access to the data held by public authorities and the reasons for its decisions – meaning in this case access to the questionnaire and to the classification carried out by the system. Notwithstanding, the State could claim confidentiality for information pertaining to the way the algorithm reaches a decision – but to do so it would have to prove that disclosure would harm the public good, which it would likely attempt to do by claiming that the unemployment public policy would be compromised by such disclosure, which is just what the Polish authorities believed.[188]

---

[186] Brazilian law recognizes a consumer relationship when, for example, the State provides a service such as water supply.

[187] The GDPA applies equally to public and private parties, though public agents enjoy different (and often less restricting) obligations.

[188] "Summarizing this problem, frontline staff in labor offices do not seem to believe that expectation of transparency or a right to information in the process of profiling is justified. They generally agree with the argumentation of the Ministry of Labor and Social Policy that a profiling interview is not something that the unemployed persons should be aware of in advance. As one of the managers put it: 'There were these ideas of the unemployed like please give me it on paper and I will prepare myself to these questions. So we explained, there is no such form of preparation to these questions. It must result from his, sort of, answers and not that he will match [responses] later, because if he gets all questions and knows what is going to be in which profile, then the answer

The principle of non-discrimination, much like the constitutional provision on equality, would certainly apply, but it is unclear how the applicability would be elaborated, for the norm is overarching and provides no specific dispositions for enforcement – given its prominence in the Brazilian legal tradition, some form of balancing would likely be exercised.[189] Another interesting question would involve the applicability of Article 20. The provision states that individuals affected by totally automated decisions can request the decision's review by a natural person. Strictly speaking, the decisions here are not totally automated. Panoptykon makes it very clear that the public officials conducting the questionnaire are able to review and change the classification rendered by the algorithm. It also notes, however, that such revisions are made in "only 0.58% of all cases of profiling."[190] This begs the question about the boundaries of totally automated decisions – should decisions that are theoretically subjected to human review but in practice are largely taken solely by machines be considered fully automated?

The Data Protection Working Party of the European Union issued guidelines[191] on this matter that shed light on the issue. The working party clarifies that, first, "the controller cannot avoid the Article 22 provisions by fabricating human involvement,"[192] and, additionally, that "to qualify as human intervention, the controller must ensure that any oversight of the decision is meaningful, rather than just a token gesture. It should be carried out by someone who has the authority and competence to change the decision."[193] In Brazil, such questions remain open to discretionary interpretation.[194]

---

might be biased.' The assumption behind this reasoning is that transparency of a decision-making process does not constitute a civil right. At the same time, it is believed that knowing what the questions are and how answers are scored would result in some sort of manipulation and abuse." NIKLAS, J. SZTANDAR-SZTANDERSKA, K. and SZYMIELEWICZ, K. **Profiling the Unemployed in Poland: social and Political Implications of Algorithmic Decision Making.** Fundacja Panoptykon. Warsaw, 2015, p. 31.

[189] SILVA, V. A. da. **O proporcional e o razoável**. Revista dos Tribunais, 2002, p. 798.

[190] NIKLAS, J. SZTANDAR-SZTANDERSKA, K. and SZYMIELEWICZ, K. **Profiling the Unemployed in Poland: social and Political Implications of Algorithmic Decision Making.** Fundacja Panoptykon. Warsaw, 2015, p. 28.

[191] JUSTICE AND CONSUMERS. **Guidelines on Automated Individual decision-making and Profiling for the purposes of Regulation 2016/679**. Adopted on October 03, 2017. Available at: <https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053>. Access: January 05, 2019.

[192] Ibid, p. 10.

[193] Ibidem.

[194] Again, this topic will be further discussed in section 4.2.

3.3.2 Credit scoring in Brazil

ITS Rio, a Brazilian non-profit organization dedicated to technology matters, carried out a study on two of the country's main credit bureaus, Serasa-Expedia (Mosaic) and Boa Vista.[195] The primary objective of the study was scrutinizing these bureaus' collection of personal data and the impact of the data's use on vulnerable groups.[196] Credit bureaus are primarily concerned with profiling consumers to determine creditworthiness. As former FTC Commissioner Julie Brill expressed during the NYU Conference on Algorithms and Accountability in 2015,

> credit reports are the grease that keeps the consumer economic wheel turning. Prior to the advent of credit reports, consumers obtained loans if they knew their local banker, or had a social reputation that preceded them into his office.[197]

In other words, the companies were originally established to minimize the information asymmetry between credit seekers and lenders, and as such allow for better risk assessment, which for its part has led to much more affordable credit. The bureaus have grown in importance, and many have diversified their activities. For the purposes of this work, however, the focus will be on profiling for creditworthiness, also known as credit scoring. Considered by some as the original algorithmic black box, credit scoring has existed since the 19th century, and using automated methods for calculating it has been around since the 1960s.

In scrutinizing the two aforementioned bureaus, ITS Rio concluded that both Mosaic, the service provided by Expedia, and Boa Vista lack transparency. Almost all information used by the systems is collected by third-parties and as such the mechanisms to ensure consent are

---

[195] The study highlights these are the only two bureaus that belong to the Brazilian National Association for Bureaus. See: ANBC. **Por que a ANBC?**. Available at: <https://www.anbc.org.br/materias.php?cd_secao=3#5&friurl=_-Empresas-associadas-_>. Access: January 05, 2019. It should be noted, however, there are other companies in the country that provide the same or similar services, such as Quod (see in: QUOD. <https://www.quod.com.br/>. Access: January 05, 2019.).

[196] ITS Rio takes its the definition of vulnerability from the report by the city of São Paulo, available in Portuguese at: CENTRO DE ESTUDOS DA MTERÓLE (CEM). **Mapa da Vulnerabilidade Social da População da Cidade de São Paulo.** Centro Brasileiro de Análise e Planejamento-CEBRAP, do Serviço Social do Comércio-SFSC e da Secretaria Municipal de Assistência Social de São Paulo, SAS-PMSP. São Paulo, 2004. Available at: <http://web.fflch.usp.br/centrodametropole/upload/arquivos/Mapa_da_Vulnerabilidade_social_da_pop_da_cidade_de_Sao_Paulo_2004.pdf>. Access: January 05, 2019.

[197] BRILL, C. J. **Scalable Approaches to Transparency and Accountability in Decision Making Algorithms, Remarks at the NYU Conference on Algorithms and Accountability,** Commissioner Julie Brill February 28, 2015.

indirect. Similarly, it is difficult to affirm the companies' compliance with the principle of purpose, for it is not possible to determine precisely from where a given piece of data comes or why it was collected. One of the differences between the systems is that Expedia has safeguards that prevent the identification of individual subjects after the data has been treated and consumers have been categorized, whereas Boa Vista allows for detailed identification.

The potential for discrimination, according to ITS, is high, and the risks for individuals are aggravated by the lack of transparency about the scope and the uses of the collected data and the lack of a channel for individuals who wish to communicate with the bureaus.[198] Here again, many types of algorithmic discrimination are plausible. Of particular relevance is that the agent carrying out the potentially discriminatory practice is not the State, which gives rise to further difficulties. Credit scoring may lead to discrimination by faulty collection or design. Turning to the LAI is not an option, although the GDPA established a principle for data quality[199] that speaks to this concern, as well as the right for any individual to correct or complete data available in datasets. There is, however, no provision to ensure that the models are statistically correct or, worse yet, to assess their accuracy.

One might argue that the reason for lack of regulation in this regard is the sufficient market incentives. In other words, no norm is needed because it is not in a company's economic interest to use statistically imprecise algorithms, simply because inaccuracy of this kind would lead to inefficient resource allocation. Misclassifying a good creditor as a bad one ultimately means less revenue. Still, it is worth noting that there is no overreaching obligation for proof that the algorithmic models used for credit reporting are statistically sound.

All the other forms of irrational discrimination are also possible precisely because the lack of transparency prevents users from checking whether they have been misclassified because of insufficient or inaccurate data (type 2) or spurious correlations (type 3). Ensuring accuracy

[198] ITS. **Transparência e governança nos algoritmos: um estudo de caso sobre o setor de birôs de crédito.** Rio de Janeiro, 2017, p. 39.
[199] The original in Portuguese reads: Artigo 6º, V: qualidade dos dados: garantia, aos titulares, de exatidão, clareza, relevância e atualização dos dados, de acordo com a necessidade e para o cumprimento da finalidade de seu tratamento;

would significantly reduce the risk of such problems. The challenges, however, are considerable. The Brazilian credit information system is based upon the premise that the responsibility to provide correct information or have it corrected lies with the consumer – she must be able to identify mistakes, bring such mistakes forward, and require the company responsible to modify the information. The notion is very much based on the model of individual complaints that reflects the history of consumer law, but it is bound to fail, namely because of the enormity of information asymmetry – put bluntly, the consumer knows far less about the database than the company and her capacity to verify whether the information collected is accurate is completely disproportionate to the capacities of the data broker.

Even if the consumer is able to spot inaccuracies in the database, however, the problem persists. According to the Credit Information Act, if the consumer identifies a mistake, it is not clear how the problem should be addressed. Article 5, III of the Credit Information Act states that the consumer can request the correction of the data, but it does not stipulate the procedure for such requests nor, more importantly, what happens in case the consumer and the database operator dispute the accuracy of the information. Who bears the burden of proof? If we take Brazilian consumer law as the governing parameter, the burden should fall upon the database operator, but if we look to the Credit Information Act, no answer is provided. In the GDPA, the right to request for data correction is made explicit in Article 18, III. The statute goes further by stating that the request for data correction must be respected by the data controller in a timely manner (Article 18, §5 – the law also states that the exact deadline for controllers shall be determined by the authorities via supplementary regulation).

Turning to the risk of discrimination through rational generalizations, discrimination by use of sensitive information is expressly prohibited in the Credit Information Act (Art. 3, §3). The problem here is of a different nature: the lack of any enforcement whatsoever. In researching this dissertation, not a single case turned up where an individual claimed that the information used by a bureau was either sensitive or excessive, which makes it impossible to say anything about how the provision will or would be interpreted by the courts. In addition, and also likely related to lack of enforcement, there exist no standards to assess proxies. The Act prohibits the

use of sensitive or excessive information, but it says nothing about information that is neither sensitive, nor necessarily excessive, but serves as a good proxy for sensitive or excessive characteristics. Again, stronger enforcement and, by consequence, a body of case law to build criteria for what precisely comprises sensitive or excessive data would mitigate this problem.

Possible discrimination by association with the fulfilment of rights is likewise to be expected, as being denied credit can affect many rights, from housing to healthcare. Yet, once again, one would require better access to the algorithm to verify whether the traits used were endogenous – and once more Article 20 may prove useful if the decision by the bureau is completely automated. What the Polish unemployment benefit system and Brazilian credit scoring clarify is that, regardless of the precise characteristics of algorithmic systems, several problematic areas remain and either Brazilian legislation is insufficient to deal with the new problems or enforcement is still wanting. The next section will delve into these points more deeply, first with a survey of the literature on algorithmic governance – and the solutions it suggests – and then with a description of the aspects of Brazilian legislation, especially the GDPA, that could provide more legal certainty.

4 Algorithmic Discrimination and the Law – The Way Forward

By now it should be clear that algorithmic systems pose challenges for policymakers, businesses and citizens alike. It is also hopefully evident that they enable efficiencies that should not be disregarded. Legislators in many jurisdictions have taken notice of these developments and attempted to formulate answers in the form of policies. However, the technology is very recent and we have yet to grasp its full impact. As such, two main questions arise, questions that are the subject of this chapter: (i) whether existing legislation is truly insufficient to handle algorithmic discrimination or can the recent regulatory efforts effectively address the challenges posed, and (ii) if improvements to existing regulations are needed, what should their focus be.

To answer these questions, I will first give an overview of the recent literature on algorithmic governance focusing on authors who study the impacts of algorithmic decision-making and aim at establishing procedures to ensure liability. Then, I will provide an assessment of the extent to which those authors' observations are accurate, to which these solutions have been implemented in Brazil, and how helpful they might be for the design of an agenda to combat algorithmic discrimination in the Brazilian legal context.

4.1 Algorithmic Governance and Policy Proposals – From Transparency to Accountability

The trade-off between regulation and innovation, discussed in many other contexts,[200] is especially evident when it comes to algorithms – the trade-off between innovation and legal

[200] In the United Kingdom, the Communications Committee opened an inquiry aimed at exploring "how the regulation of the internet should be improved, including through better self-regulation and governance, and whether a new regulatory framework for the internet is necessary or whether the general law of the UK is adequate". See in: COMMUNICATIONS COMMITTEE. **The Internet: to regulate or not to regulate? inquiry**. Parliamentary business. Available at: <https://www.parliament.uk/business/committees/committees-a-z/lords-select/communications-committee/inquiries/parliament-2017/the-internet-to-regulate-or-not-to-regulate/>. Access: January 05, 2019. Similar debates arose in many other fields, such as environmental law (see in: BERGER, M. M. **To Regulate, or Not to Regulate – Is That the Question: Reflections on the Supposed Dilemma between Environmental Protection and Private Property Rights**. 8 Loy, L. A. L. Rev. 253, 1975. Available at: <https://digitalcommons.lmu.edu/cgi/viewcontent.cgi?referer=https://scholar.google.com.br/scholar?hl=en&as_sdt=0%2C5&q=to+regulate+or+not&btnG=&httpsredir=1&article=1187&context=llr>. Access: January 05, 2019.), antitrust (see in: DEMSETZ, H. **Why Regulate Utilities?**. Journal of Law and Economics, vol. 11, no. 1, The University of Chicago Press Journals, April, 1968. Available at: <https://www.journals.uchicago.edu/doi/abs/10.1086/466643?journalCode=jle>. Access: January 05, 2019.), gender issues, finances (see in: ACHARYA, V. **Regulating wall Street: the Dodd-Frank Act and the new**

certainty is acute with regards automated systems. Apart from isolated voices who believe no regulation whatsoever should be implemented because they think it will always and (un)necessarily impede innovation, the consensus is that some level of oversight is appropriate. The debate is mostly focused on what level of scrutiny should be implemented, and here the suggestions vary widely, ranging from claims that governments should regulate first and worry about the impact on innovation later to assertions that regulation should be carefully curated and narrowly tailored to the specific situations that require it.

As such, the body of literature that examines how discrimination can be a factor in algorithmic decision-making largely grants that policy solutions must be discussed and has accordingly provided an overview of possible ones. This section will review the prominent debates in this respect, focusing on (i) the most commonly discussed policy solutions, and (ii) their applicability and special hurdles faced with regards machine learning.

The literature on algorithmic governance, though fairly recent, is extensive. Several authors have taken up the questions of when and how algorithms should be regulated. Most notably, groups of scholars have come up with principles that could govern algorithmic decision-making. The Fairness, Accountability and Transparency in Machine Learning Organization (FAT-ML) is one such institution. It has compiled a list of what it believes to be the key principles that should be observed by companies and governments when dealing with algorithms: responsibility, explainability, accuracy, auditability, and fairness.[201] In the United States, the Association for Computing Machinery (ACM) followed a similar path and devised its own principles, adding awareness, access and redress, data provenance, and validation and testing to the list.[202]

---

architecture of global finance. Interviewer: DAVIES, V. VOX - CEPR Policy Portal. October 22, 2010. Available at: <https://voxeu.org/vox-talks/regulating-wall-street-dodd-frank-act-and-new-architecture-global-finance>. Access: January 05, 2019.) and so on.

[201] The Fairness, Accountability and Transparency in Machine Learning Organization, see in: DIAKOPOULOS, N., FIEDLER, S., ARENAS, M. et al. **Principles for Accountable Algorithms and a Social Impact Statement for Algorithms**. FAT/ML. Available at: <https://www.fatml.org/resources/principles-for-accountable-algorithms>. Access: January 05, 2019.

[202] ASSOCIATION FOR COMPUTING MACHINERY US PUBLIC POLICY COUNCIL (USACM). **Statement on Algorithmic Transparency and Accountability.** January 12, 2017. Available at: <http://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf>. Access: January 05, 2019.

Responsibility, according to FAT-ML, relates to the idea that one, in designing algorithmic systems, must consider the people that will be impacted by the decision-making process and as such should to some extent provide mechanisms for redress – both at the individual and societal levels. This idea connects to the ACM's principles of *awareness* – which is mostly focused on raising the algorithm's builders and users awareness of the possible consequences of its use, especially regarding the biases that can arise from it; and of *access and redress* – which claims regulators should adopt mechanisms that allow individuals impacted by the decisions made by algorithms to question and repair potential harms.

As Doshi-Velez et al. put it, the idea of *explanation* (or explainability as the FAT-ML calls it), when applied to decision-making, refers to "the reasons or justifications for that particular outcome, rather than a description of the decision-making process in general."[203] Therefore, what they consider to be an explanation is a "human-interpretable description of the process by which a decision-maker took a particular set of inputs and reached a particular conclusion."[204] It is important to note that explanation is not identical to transparency, for being able to understand the process by which a decision was made is not the same as knowing every step taken.

The principle of *accuracy*, according to Diakopoulos and Friedler, means that the "sources of error and uncertainty throughout an algorithm and its data sources need to be identified, logged, and benchmarked."[205] Put bluntly, it is only by understanding the origins and causes of mistakes that one can hope to mitigate them. The ACM expresses a similar notion through the principle of *data provenance*, which states that "a description of the way in which the training data was collected should be maintained by the builders of the algorithms,

---

[203] DOSHI-VELEZ, F. and KORTZ, M. **Accountability of AI Under the Law: The Role of Explanation**. Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper, 2017, p. 2. Available at : <https://dash.harvard.edu/bitstream/handle/1/34372584/2017-11_aiexplainability-1.pdf?sequence=3>. Access : January 05, 2019.

[204] Ibid, p. 2-3. They go on to say that an explanation should be able to answer at least one of the three following questions: (i) What were the main factors in a decision?; (ii) Would changing a certain factor have changed the decision?; and (iii) Why did two similar-looking cases yield different decision, or vice-versa?

[205] DIAKOPOULOS, N. and FRIEDLER, S. **How to Hold Algorithms Accountable**. MIT Technology Review. November, 2016.

accompanied by an exploration of the potential biases induced by the human or algorithmic data-gathering process."[206]

The principle of *auditability* is another constant in discussions of algorithmic governance. It entails requiring third party review of the method used by the algorithm to reach its conclusions.[207] How this disclosure should be undertaken, and whether it should take place at all in certain circumstances, especially where commercial secrets are involved, is a subject of much debate.

Fairness may be the most obvious if least clear of all the principles proposed. The idea behind fairness is preventing algorithms from reaching discriminatory outcomes. As seen in chapter 2 above, however, determining what constitutes a discriminatory outcome is often challenging. The ACM, without expressly subscribing to the principle of fairness, puts forward the *validation and testing* standard, according to which "[institutions] should routinely perform tests to assess and determine whether the model generates discriminatory harm."[208]

The principle of *transparency*, although not explicitly present in either of these manifests, is also widely discussed in the literature and in policy making circles. Algorithms have been famously called "black boxes" by Frank Pasquale due to the opacity of their decision-making processes that invites distrust. Hence, some scholars consider tools aimed at identifying the elemental components of algorithms as essential for any proposed regulatory solution.[209]

---

[206]ASSOCIATION FOR COMPUTING MACHINERY US PUBLIC POLICY COUNCIL (USACM). **Statement on Algorithmic Transparency and Accountability.** January 12, 2017. Available at: <http://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf>. Access: January 05, 2019.

[207] SANDVIG, C. et al. **An Algorithm Audit.** Data and Discrimination: Collected Essays. 2014. Available at: <http://www-personal.umich.edu/~csandvig/research/An%20Algorithm%20Audit.pdf>. Access: January 07, 2019.: "Although the complexity of these algorithmic platforms makes them seem impossible to understand, audit studies can crack the code through trial and error: researchers can apply expert knowledge to the results of these audit tests. By closely monitoring these online platforms, we can discover interactions between algorithm and data. In short, auditing these algorithms demands a third party that can combine both expert and everyday evaluations, testing algorithms on the public's behalf and investigating and reporting situations where algorithms may have gone wrong."

[208]ASSOCIATION FOR COMPUTING MACHINERY US PUBLIC POLICY COUNCIL (USACM). **Statement on Algorithmic Transparency and Accountability.** January 12, 2017. Available at: <http://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf>. Access: January 05, 2019.

[209] PASQUALE, F. **The Black Box Society**. Harvard University Press. January, 2015.

There is no consensus regarding the precise, ideal combination of these proposals, as authors disagree on their relative importance and usefulness. The most visible disagreement is voiced by scholars who believe transparency - and even explainability - are inadequate or insufficient tools and argue that establishing firm accountability mechanisms would be a better option.

Pasquale and Citron are among the authors who believe transparency is a meaningful solution for algorithmic discrimination, especially when applied to credit scoring. In their words:

> We believe that each data subject should have access to all data pertaining to the data subject. Ideally, the logics of predictive scoring systems should be open to public inspection as well. There is little evidence that the inability to keep such systems secret would diminish innovation.[210]

They clearly state that the "threats to human dignity" justify requiring disclosure t the public in general of not only the dataset and the overall functioning of the system to authorities, but also of the code and modeling of algorithms.[211]

Another author who follows a similar line of thought is Zarsky, whose argument refers to the context of automated predictions in government initiatives. In short, he claims that "the most basic and popular justification for transparency is that it facilitates a check on governmental actions."[212] He argues that, in this context, although accountability and transparency are often used synonymously, they should be distinguished as fundamentally different in that accountability involves the ethic responsibility of individuals for their actions, whereas transparency is a tool – and not the only one – whose objective is facilitating accountability.

---

[210] Pasquale and Citron also say: "The FTC's expert technologists could test scoring systems for bias, arbitrariness, and unfair mischaracterizations. To do so, they would need to view not only the datasets mined by scoring systems but also the source code and programmers' notes describing the variables, correlations, and inferences embedded in the scoring systems' algorithms.". See in: CITRON, D. K. and PASQUALE, F. **The Scored Society: Due Process For Automated Predictions.** Washington Law Review, vol. 89:1, 2014, p. 26. Available at: <https://digital.law.washington.edu/dspace-law/bitstream/handle/1773.1/1318/89WLR0001.pdf>. Access: January 05, 2019.

[211] Ibid, p. 30-31.

[212] ZARSKY, T. Z. **Transparent Predictions.** University of Illinois Law Review, vol. 2013, no. 4, p. 1533. Available at: <https://www.illinoislawreview.org/wp-content/ilr-content/articles/2013/4/Zarsky.pdf>. Access: January 05, 2019.

Experts, however, have observed some limitations of transparency solutions. Lawrence Lessig has famously presented such a view with respect to government transparency, when he claimed that turning the panopticon to focus on the authorities, thus creating civic omniscience, was problematic. He built his argument upon the ideas expressed by Brandeis in *Other People's Money*, namely the argument that full disclosure of information would help the public judge quality and as such allow the people to regulate markets. As Lessig warns, "not all data satisfies the simple requirement that they be information that consumers can use, presented in a way they can use it."[213] Although the subject of his paper is not algorithmic discrimination, many of his observations are applicable to it.

For their part, Ananny & Crawford state the ideal of transparency rests on the belief that:

> the more facts revealed, the more the truth can be known through a logic of accumulation. Observation is understood as a diagnostic for ethical action, as observers with more access to the facts describing a system will be better able to judge whether a system is working as intended and what changes are required.[214]

As the authors emphasize, however, this assumption only holds true if one assumes that "knowing is possible by seeing,"[215] an affirmation that they contest on ten different fronts: (1) Transparency can be disconnected from power, meaning that transparency as a means of accountability will only work inasmuch as those subjected to it are somewhat vulnerable to its consequences, a condition that does not always hold. (2) Transparency may expose information about individuals or groups without any clear benefit, damaging privacy. (3) If transparency is made an overarching obligation, actors subjected to it may decide to reveal information strategically; in other words, they may do so in a way that hinders rather than facilitates understanding. (4) Transparency requirements may create "false binaries,"[216] as well as the false perception that the only options available are full disclosure or total secrecy, which is not true.

---

[213] LESSIG, L. **Against Transparency.** The New Republic, 2009. Available at: <https://newrepublic.com/article/70097/against-transparency>. Access: January 05, 2019.
[214] ANANNY, M. and CRAWFORD, K. **Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability**. SAGE journals. December 13, 20016, p. 2. Available at: <https://journals.sagepub.com/doi/abs/10.1177/1461444816676645?journalCode=nmsa>. Access: January 05, 2019.
[215] Ibid, p. 5.
[216] Ibid, p. 7.

(5) The ideal of transparency rests upon other assumptions, such as perfect information and fully rational decision-making – the premise being that once individuals are able to examine a system, they will be fully capable of understanding it, and, more importantly, of making completely rational decisions based on the information provided. Ananny & Crawford emphasize "persistent fiction"[217] of these assumptions. (6) Transparency does not always build trust. (7) Transparency usually involves some level of professional expertise, in the sense that "professionals have a history of policing their boundaries […] It may be impossible to really see professional practices without understanding that they are situated within contexts."[218] (8) The call for transparency assumes that to see is to known, something educational observation over time has proven untrue.[219] (9) Transparency requirements are sometimes made infeasible or technically cumbersome by advances or developments in computer science technology where as will be seen in section 4.1.1 below – machine learning poses additional challenges. (10) The timing of disclosure of algorithmic systems can affect results, in tat revealing the inner working of a system before, during or after the system becomes operational has distinct consequences, which is compounded by the fact that the system itself is likely to change over time.

These objections find resonance with Schauer's statement related earlier on the limitations of individual observation as the most relevant basis for decision-making. He argues that what we observe and what we are able to infer from our observations often yields an incomplete picture.[220] Echoing these concerns, Kroll et al. advance four connected arguments sustaining that transparency is not a sufficient policy proposal[221]: (1) Transparency may simply be unattainable – there will either exist well-grounded public reasons that trump the right to disclosure, such as national security or preventing strategic behavior aimed at gaming the

[217] Ibid, p.8.
[218] Ibidem.
[219] "Learning about complex systems means not simply being able to look inside systems or take them apart. Rather, it means dynamically interacting with them in order to under- stand how they behave in relation to their environments (Resnick et al., 2000). This kind of complex learning intertwines epistemological claim-making with material design, social contexts, and self-reflexivity—making sure that a system is not only *visible* but also debated and changeable by observers who are able to consider how they know what they know about it.". In: Ibid, p. 9.
[220] Ibid, p. 10: "To ask to 'look inside the black box' is perhaps too limited a demand and ultimately an ill-fitting metaphor for the complexities of contemporary algorithmic systems. It side-steps the material and ideological complexities and effects of seeing and suggests a kind of easy certainty that *knowing* comes from looking."
[221] KROLL, J. et al. **Accountable Algorithms**. 165 U. PA. L. Rev. 633, 2017. Available at: <https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3/>. Access: January 05, 2019.

system,[222] or reasons that affect the individuals under scrutiny. For example, when the data collected is highly sensitive, full transparency of the database may not be in the individual's best interest. (2) Transparency might also be insufficient – even if a rule is public, "[the] methods are often insufficient to verify properties of software systems, if these systems have not been designed with the future evaluation and accountability in mind,"[223] as is often the case. (3) Whenever the algorithm incorporates randomness – arguably a fundamental function of computerized systems – transparency mechanisms lose efficacy.[224] (4) "Intelligent" systems that change over time and adapt to their environment, such as learner algorithms, cannot be properly comprehended through transparency mechanisms. Following this line of thought, the authors argue that accountability, implemented in the form of procedural regularity, is a better policy proposal that should be studied in more depth, for it may provide the desired outcome without compromising confidentiality.

Desai & Kroll likewise claim that even though there is a place for transparency measures, they remain limited, for they may inadvertently hide discrimination by providing a false sense of clarity:

> Many of the current calls for transparency as a way to regulate automation do no address such limits [of the algorithmic systems], and so they may come up short on providing the sort of legal-political accountability they desire, and which we also support. Instead, as software (and especially machine learning systems, which separate the creation of algorithms and rules from human design and implementation) continues to grow in importance, society may find, and we argue, that identifying harms, prohibiting outcomes, and banning undesirable uses is a more promising path.[225]

Similarly, Edwards & Veale underscore the limitations of transparency and explainability, primarily as they relate to the GDPR in Europe. They claim transparency is useless both because enforcement is unfeasible and because even achieving it does not serve

---

[222] The authors use the example of tax evasion. If tax evaders knew precisely how the government identifies possible fraud scenarios, the algorithm in place would be useless.
[223] Ibid, p. 633.
[224] In their words, "a simply lottery provides an excellent example: a perfectly transparent algorithm – use a random number generator to assign a number to each participant and have the participants with the lowest numbers win – yields results that cannot be reproduced or verified because the random number generator will produce new random numbers when it is called upon later". In: Ibid.
[225] LESSIG, L. **Against Transparency.** The New Republic, 2009, p. 7. Available at: <https://newrepublic.com/article/70097/against-transparency>. Access: January 05, 2019.

users' needs – for it is unable to redress substantive injustice – and suggest a framework focused on building better algorithmic solutions from the start (solutions that are effectively concerned with policy objectives) and giving agencies and institutions the power to oversee algorithmic integrity.[226]

The Center for Data Innovation follows an even stricter line of thought. It heavily criticizes the GDPR, as it believes the new regulation will do little to protect data subjects and will put the European Union at a severe disadvantage when compared to the United States and Chine in terms of AI development.[227] The Center further claims that most authors' call for algorithmic regulation are misplaced, for "proposed solutions are typically ineffective, counterproductive, or harmful to innovation,"[228] and classifies major proposals for algorithmic governance as follows, indicating their most recurrent flaws:

Table 1. Proposals for Algorithmic Governance and Their Flaws.

Proposal	Flaws
Algorithmic transparency or explainability mandates	• Holds algorithmic decisions to a standard that does not exist for human decisions • Incentivizes organizations to not use algorithms, thus sacrificing productivity • Fails to address the root cause of potential harms • Assumes the public and regulators could interpret source code for complex algorithms even developers themselves cannot always understand

[226] "As the history of industries like finance and credit shows, rights to transparency do not necessarily secure substantive justice or effective remedies. We are in danger of creating a "meaningless transparency" paradigm to match the already well known "meaningless consent" trope.". In: EDWARDS, L. and VEALE, L. **Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For**. 16 Duke Law & Technology Review 18, 2017, p. 22-23. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2972855&download=yes>. Access: January 05, 2019.

[227] WALLACE, N. and CASTRO, D. **The Impact of the EU's New Data Protection Regulation on AI**. Center for Data Innovation, March 26, 2018. Available at: <https://www.datainnovation.org/2018/03/the-impact-of-the-eus-new-data-protection-regulation-on-ai/>. Access: January 05, 2019.

[228] NEW, J. and CASTRO, D. **How Policymakers can foster algorithmic accountability**. Center for Data Innovation, May 2018, p. 7. Available at: <http://www2.datainnovation.org/2018-algorithmic-accountability.pdf>. Access: January 05, 2019.

	• Undermines closed-source software, reducing incentives for innovation • Makes it easy for bad actors to "game the system" • Creates incentives for the use of less-effective AI, as there can be trade-offs between explainability and accuracy for complex AI • Is useful in select contexts but ineffective or harmful in others
Regulatory bodies to oversee all algorithmic decision-making	• Ignore the need for regulators to have context-specific expertise • Low-risk decisions should not be subject to regulatory oversight
Generalized regulatory proposals	• Provide no specifics on how to operationalize proposals • Rely heavily on platitudes that do not translate to effective governance
Do nothing	• Does not recognize that, for some use cases-particularly certain government applications-algorithms are less subject to market forces that would minimize potential harms • Fails to prevent algorithms from causing harm in certain contexts where such harms are not obvious or do not break the law

Source: Wallace and Castro (2018).

Following these observations, the majority of experts take the view that placing too much faith in transparency measures would indeed be misguided, although not because the goal of transparency is misguided, but rather because such measures may prove insufficient for attaining it. Algorithmic accountability has thus emerged as one of the prevailing alternatives to transparency, even though the precise implications of this concept remain ambiguous. Lodge & Stirton state that "[i]n modern parlance, accountability more commonly signifies the obligation of officials to account for their behavior"[229] whereas "'transparency' offers 'visibility', such as

[229] LODGE, M. and STIRTON, L. **Accountability in the Regulatory State.** In: BALDWIN, R., CAVE, M. and LODGE, M. (Eds). The Oxford Handbook of Regulation, Chapter 15, September 2010, p. 351.

the publication of all procurement contracts on the Internet and such like."[230] The authors recognize, however, that the terms are often times used interchangeably, for what unites them is "a concern with the use of discretionary (private and public) authority."[231] Primarily in the United States, the idea of accountability has been conceived to mean rationality. As Lodge & Striton emphasize, cost-benefit analysis has been suggested as a means for fostering accountability, as "a procedural way of enhancing the 'rationality' of rules […] and informing decision-makers as to what is the appropriate (rational) option."[232]

The authors point out that the new context of the regulatory state requires that the traditional understanding be reviewed. The privatization of public services, the rise of supposedly independent regulatory agencies, the growth of self or co-regulation, linked with the emergence of the international debate on this topic (and therefore the creation of international standards), and social pressure for more accountability have all contributed to a discussion that "echoes traditional concerns with administrative bodies" but also:

> takes place under the conditions of polycentricity (in both the vertical and horizontal senses), whether this is the distribution of authority (i.e. to international organizations and non-state organizations) or the transnational nature of corporate power in areas that traditionally were reserved for national states (especially in the area of utilities, such as telecommunications).[233]

The goal of this dissertation is not to engage in a deep debate of accountability, but rather to establish the complexity of the concept as well as its usefulness for the algorithmic discrimination discussion. With that in mind, here is the attempt of the World Wide Web Foundation to provide a definition of accountability:

> Accountability is usually referred to as the duty governments and other authorities have to present themselves before those whose interest they represent or are otherwise bound to, and justify how power was exercised, and resources were used. (…) Although we are at a stage in which the definition of algorithmic accountability is still being agreed upon, experts and practitioners have been putting forward general principles to be debated.[234]

---

[230] Ibid, p. 352.
[231] Ibid. p 352.
[232] Ibid. p. 353.
[233] Ibid. p. 356.
[234] WORLD WIDE WEB FOUNDATION. **Algorithmic Governance: Applying the Concept to Different Country Contexts**. July, 2017, p. 10-11.

Helen Nissenbaum, in what is perhaps one of the first articles written on accountability and computerized systems, makes a cunning and very pragmatic observation. She claims that even though most of the risks presented by these systems must be mitigated through careful design – since one intends to prevent failure from the outset – the need for accountability still increases, "because those who are answerable for harms or risks are the most driven to prevent them. In this way, accountability serves as a powerful tool for bringing about better practices, and consequently more reliable and trustworthy systems."[235] Nissenbaum also calls attention to four characteristics of computation that impede accountability, namely: (i) "the problem of many hands,"[236] or collective responsibility, which is common because computer systems are usually built by groups rather than individuals, and assigning blame to a group has historically been a challenge[237]; (ii) bugs, or software errors in general, which are characterized as endemic to any complex computerized system, and as such compound the assessment of responsibility; (iii) treating the machine as a scapegoat in order to remove any human responsibility or error; and (iv) the controversy over software ownership, which if resolved could provide a clearer individual target for accountability debates.

Nissenbaum makes some recommendations to rehabilitate accountability in light of these characteristics, for she understands that "we should hold on to the assumption that someone is accountable, unless after careful investigation, we conclude that the malfunction in question is, indeed, no one's fault."[238] Her first suggestion is to separate accountability from liability, since liability usually equates to determining who should pay whom and how much, whereas accountability is centered on the action. Even in the case of collective actions, each individual involved shares in the responsibility for the offense. That she is not directly liable does not make her any less accountable for the outcome, and the author believes that preserving this capacity to identify those who were behind the offensive outcome is paramount. Her second

[235] NISSENBAUM, H. **Computing and Accountability**. In: COHEN, J. Communications of the ACM, vol. 37, issue 1, January 1994, p. 74.
[236] Ibid, p. 75.
[237] Nissenbaum uses the example of Therac-25 to illustrate the problem more precisely. Therac-25, built by the Atomic Energy of Canada Limited, was a medical device designed to destroy cancerous cells in patients through radiation. Patients received overdoses, however, in at least six instances, leading to death and irreversible injury. In: Ibid, p. 75-76.
[238] Ibid, p. 79.

observation addresses the need for a "standard-of-care," or, in other words, "a call for simpler design, a modular approach to system building, formal analysis of modules as well as the whole, meaningful quality assurance, independent auditing, built-in redundancy, and excellent documentation."[239] Her view is that this approach would both incentivize better system design and setting high standards for system engineers while simultaneously differentiating between preventable and unpreventable bugs. Lastly, Nissenbaum calls for strict liability for consumer-oriented or large-scale software, which would shift the burden of proof to producers, and require extraordinary measures on their part whenever the system under construction is developed for widespread use.

Following a different path, the Center for Data Innovation suggests a model in which three goals are central in attaining algorithmic accountability: (i) promoting desirable outcomes, (ii) protecting against harmful outcomes, and (iii) ensuring that laws applicable to human decision-making also apply to algorithms.[240] They draw a stark distinction between operators and developers, claiming that only algorithms that are indeed applied to decision-making should be of concern to regulators,[241] and only some algorithms, those whose applicability "poses potential harms significant enough to warrant regulatory scrutiny"[242] should be the focus of public policy.

Goodman suggests yet another separate but connected approach: algorithm auditing. He qualifies auditing as safety engineering: "auditing identifies key process risks, evaluates whether adequate safeguards are in place and, where gaps are found, provides guidance on risk prevention going forward."[243] He sustains that it could represent a means of accountability,

---

[239] Ibid, p. 79.

[240] NEW, J. and CASTRO, D. **How Policymakers can foster algorithmic accountability**. Center for Data Innovation, May 2018, p. 21. Available at: <http://www2.datainnovation.org/2018-algorithmic-accountability.pdf>. Access: January 05, 2019.

[241] Ibid, p. 21-22: "simply creating an algorithm that exhibits some kind of demographic bias, for example, does not cause others harm and should be of no concern to regulators unless an operator applies it in a way that could cause harm, just as it is not illegal for a person to hold biases, but it is against the law for them to base certain decisions on these biases, such as deciding whom to hire."

[242] Ibid, p. 22.

[243] GOODMAN, B. W. **A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection**. 29th Conference on Neural Information Processing Systems. Barcelona, Spain. 2016, p. 5.

emphasizing the need for pre-, in- and post-processing techniques in the execution of audits, explaining that while the former checks the training data for discrimination by modifying the dataset in some way, the last evaluate "a general supervised learning classification problem to become 'discrimination aware', i.e. to learn a classifier such that accuracy is high and discrimination with respect to the protected category is low."[244] Goodman does not contend that auditing will solve all algorithmic problems, but rather advocates it can reduce risk. Algorithms that pass audits may still be inefficient, or discriminate for other reasons, but the process of creating a paper trail, much like auditing for other purposes, could very well minimize risks.

Citron's proposal involves administrative law, and as such she is primarily concerned with making sure agencies are equipped with decision-making processes and guarantees that suffice in the world of automation. Her suggestions for technological due process, however, reflect in many ways what other authors claim algorithmic accountability would require in the context of governmental use of machine learning and algorithms in general. She argues that three practices should be adopted by government agencies: (i) maintaining audit trails, which would help comply with notice requirements[245]; (ii) holding hearings to clarify automated systems' fallibility and afford justification from officers for automated decisions on a case-by-case basis; (iii) ensuring transparency and accountability by, specifically (a) making systems' source code public, (b) conducting testing and monitoring by independent agents, (c) involving public participation in the building of systems as much as possible, and (d) refraining from automating policies which have not undergone formal or informal rulemaking.

By this point it should be clear that the challenges of algorithmic governance are daunting. If transparency seems insufficient, accountability is a somewhat vague and multifaceted, leaving much room for interpretation and ongoing debate. Approaching the problem with machine learning algorithms in mind remains to be done, and that will be the goal

---

[244] Ibidem.

[245] "Audit trails should include a comprehensive history of decisions made in a case, including the identity of the individuals who recorded the facts and their assessment of those facts. Audit trails should detail the actual rules applied in every mini-decision that the system makes. With audit trails, agencies would have the means to provide individuals with the reasons supporting an automated system's adjudication of their important rights." In: CITRON, D. K. **Technological Due Process**. 85 Wash. U. L. Rev. 1249, 2008, p. 1305. Available at: <http://openscholarship.wustl.edu/law_lawreview/vol85/iss6/2>. Access: January 01, 2019.

of the next section. The reason is straightforward: due to the characteristics of leaners algorithms, some governance solutions are either unattainable or tremendously difficult to implement. Experts have devoted particular attention to the problem, and in what follows the outcomes of their research will be clarified.

4.1.1 The Policy Challenges of Machine Learning

The limitations of policy are usually evident when machine learning is involved. Because of the way learner algorithms operate, they pose special challenges for transparency and accountability. In 2017, Will Knight wrote in the MIT Technology Review that:

> As the technology advances, we might soon cross some threshold beyond which using AI requires a leap of faith. (…) Artificial intelligence hasn't always been this way. From the outset, there were two schools of thought regarding how understandable, or explainable, AI ought to be. Many thought it made the most sense to build machines that reasoned according to rules and logic, making their inner workings transparent to anyone who cared to examine some code. Other felt that intelligence would more easily emerge if machines took inspiration from biology, and learned by observing and experiencing. (…) The machine-learning techniques that would later evolve into today's most powerful AI systems followed the latter path: the machine essentially programs itself.[246]

Most of today's machine learning uses deep learning and neural networks to execute decision-making, networks that are extremely difficult for non-specialists to understand. Often a system's developers themselves are unable to determine how a specific result was generated – Knight mentions the example of Deep Patient, a system developed at Mount Sinai Hospital in New York in order to determine patients' risk of certain diseases. The program was trained using a database of over 700,00 patient records and proved extremely accurate in identifying illnesses. It was unable to explain its conclusions to doctors, however, which severely impaired treatment efforts.

Another area where explainability is absolutely crucial is military use of machine learning. It is thus unsurprisingly that the U.S. Defense Advanced Research Projects Agency,

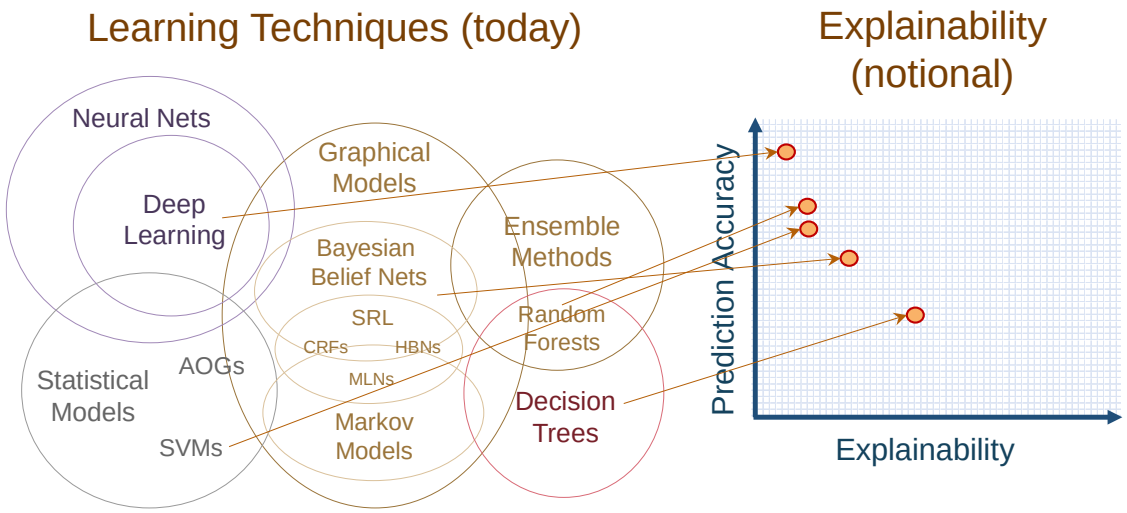
[246] KNIGHT, W. **The Dark Secret at the Heart of AI**. MIT Technology Review. April 11, 2017. Available at: <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/>. Access: January 05, 2019.

or DARPA, has undertaken studies in this area. DARPA now conducts a program entitled Explainable Artificial Intelligence (XAI), whose self-stated goal is to:

> create a suite of machine learning techniques that produce more explainable models, while maintaining a high level of learning performance (e.g. prediction accuracy) and enable human users to understand, appropriately trust, and effectively manage the emerging generation of artificial intelligent partners.[247]

DARPA's current understanding of the AI ecosystem is be summarized in the following diagram:

Figure 2 – The Artificial Intelligence Ecosystem.



Source: Gunning (2018).

That is why XAI seeks to modify this scenario by developing explainable models that retain prediction accuracy. In that light, authors have begun drafting proposals to address the challenges brought forward by machine learning and simultaneously try to ensure algorithmic governance. Gasser & Almeida, for instance, propose a layered model for AI governance, emphasizing the importance of:

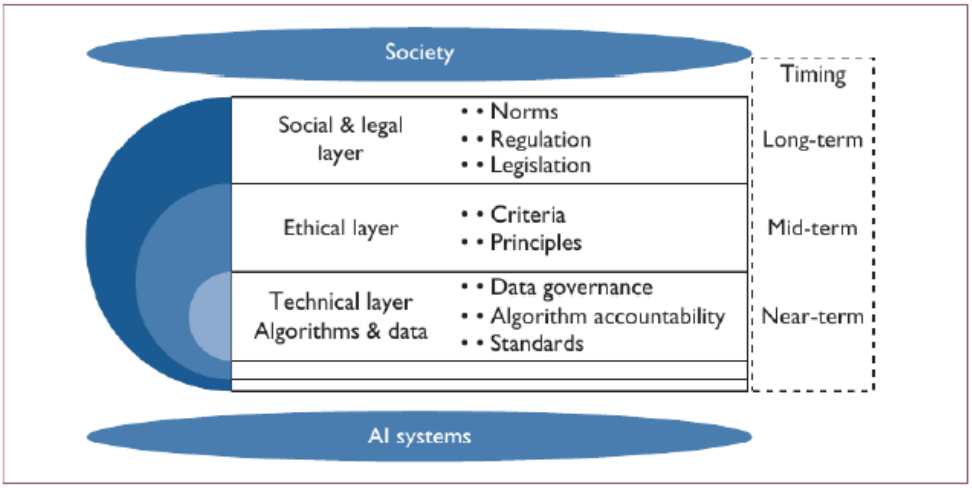
> the idea of modularity embodied in the form of layered governance, which also combines different instruments to grapple with and address the aforementioned

[247] GUNNING, D. **Explained Artificial Intelligence (XAI).** DARPA/I20. November, 2017. Available at: <https://www.darpa.mil/attachments/XAIProgramUpdate.pdf>. Access: January 05, 2019.

> substantive issues [information asymmetries, normative consensus, and government mismatches].[248]

The authors present a three-layered model: (i) the social and legal layer, (ii) the ethical layer, and (iii) the technical foundations that support the ethical and social layers, represented by figure 2.

Figure 3 – Proposal for a Lawyered Model of AI Governance.



Source: Gasser and Almeida (2017).

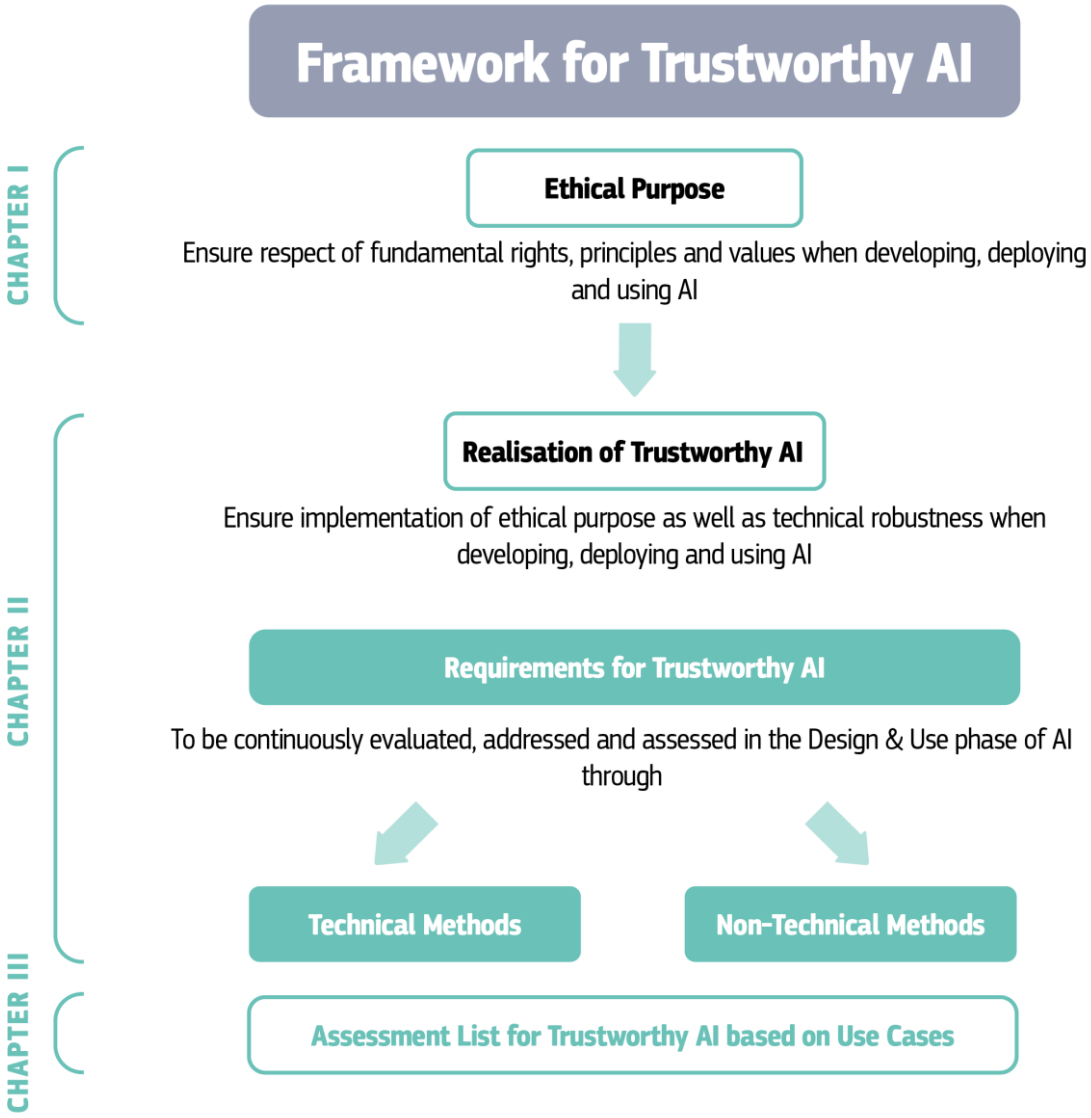
As the figure illustrates, Gasser & Almeida believe that in the short-term governments should concentrate on establishing standards and parameters for algorithmic governance, and only later take up specific regulation aimed at tackling more complex applications.

The High-Level Expert Group in Artificial Intelligence of the European Commission (AI HLEG), following the implementation of the GDPR, drafted ethics guidelines for the use and development of AI systems. In their words, "this guidance forms part of a vision embracing a human-centric approach to Artificial Intelligence, which will enable Europe to become a globally leading innovator in ethical, secure and cutting-edge AI."[249] The AI HLEG takes

[248] GASSER, U. and ALMEIDA, V. A.F. **A Layered Model for AI Governance**. IEEE Internet Computing 21 (6) (November): 2017, p. 4.
[249] EUROPEAN COMMISSION'S HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE (AI HELG). **Draft Ethics Guidelines for Trustworthy AI**. European Commission, Brussels, March, 2019 (final version), p. iii. Available at: <https://ec.europa.eu/futurium/en/node/6044>. Access: January 07, 2019.

trustworthy AI as its guiding principle, emphasizing two components of the concept: (i) respect for fundamental rights and regulations in order to maintain ethical commitments, and (ii) technical robustness and reliability. Moreover, the guidelines are structured in the form of a framework that is subdivided into three chapters: ethical purpose, realization of trustworthy AI, and assessment list and use cases, as illustrated by the figure below:

Figure 4 – A Framework for Trustworthy AI.



Source: AI HLEG (2019).

Regarding ethical purpose, the guidelines propose five principles and values to be followed.[250] With respect to the requirements for trustworthy AI, ten are put forward.[251] Yet the most meaningful contribution of the AI HLEG is represented by the methods it suggests for achieving trustworthy AI. The organization makes it explicit that the process for achieving such goal is a continuous one, and therefore the analysis, design, development and use of AI all require attention. The means to reach the objective rely on both technical and non-technical methods. Technical methods include the proposals of (i) XAI – as mentioned above, (ii) the idea of ethics and rule of law by design much in line with the widespread idea of privacy-by-design, (iii) AI architecture – "the requirements for Trustworthy AI need to be 'translated' into procedures and/or constraints on procedures, which should be anchored in an intelligent system's architecture,"[252] (iv) testing and validating – the system must be tested continuously, for "it must be ensured that the outcome of the planning process is consistent with the input, and that the decisions taken can be made plausible in a way allowing validation of the underlying process,"[253] and (v) traceability and auditability – documenting the decisions taken by AI systems and their processes allows for internal and external auditors to analyze the system and to reach conclusions about when and why a specific decision may have been mistaken.

Non-technical methods consist of (i) regulation – more than introducing new rules, the AI HLEG emphasizes the need for revision and adaptation of current legislation, and mechanisms to redress harm; (ii) standardization of AI algorithms by external accreditation associations that may be helpful in managing quality; (iii) accountability mechanisms that may include creating an oversight position within the organization in question; (iv) codes of conduct – the guidelines presented by the AI HLEG might be officially adopted by organizations, or the organization could make public the rules according to which its activities will be carried out; (v) education and awareness of ethical principles – stakeholders must participate in the process

[250] They are: (i) the principle of beneficence, (ii) the principle of non-maleficence, or do no harm, (iii) the principle of autonomy, which aims to preserve human self-determination, (iv) the principle of justice, and (v) the principle of explicability.
[251] They are: (a) accountability, (b) data governance, (c) design for all, (d) governance of AI autonomy, or human oversight, (e) non-discrimination, (f) respect for (and enhancement of) human autonomy, (g) respect for privacy, (h) robustness, (i) safety, and (j) transparency.
[252] Ibid, p. 19.
[253] Ibid, p. 20.

of AI-building for it to be truly trustworthy, which includes the creators, developers, users, and the impacted groups; (vi) stakeholders and social dialogue – the process of implementing AI needs to be wide-ranging and public, and engage as many stakeholders as possible in the discussion; and (vii) diversity and inclusive design terms – it is crucial that "the teams that design, develop, test and maintain these systems reflect the diversity of users and of society in general."[254]

It should be noted that the EU's approach to AI and algorithms is challenged by some authors, primarily for the impact they believe it will have on innovation. As mentioned, the Center for Data Innovation claims that the GDPR and the obligations it imposes on controllers and developers will put EU firms at a competitive disadvantage, and "consign Europe to second-tier status in the emerging AI economy."[255] Jia et al. point in the same direction:

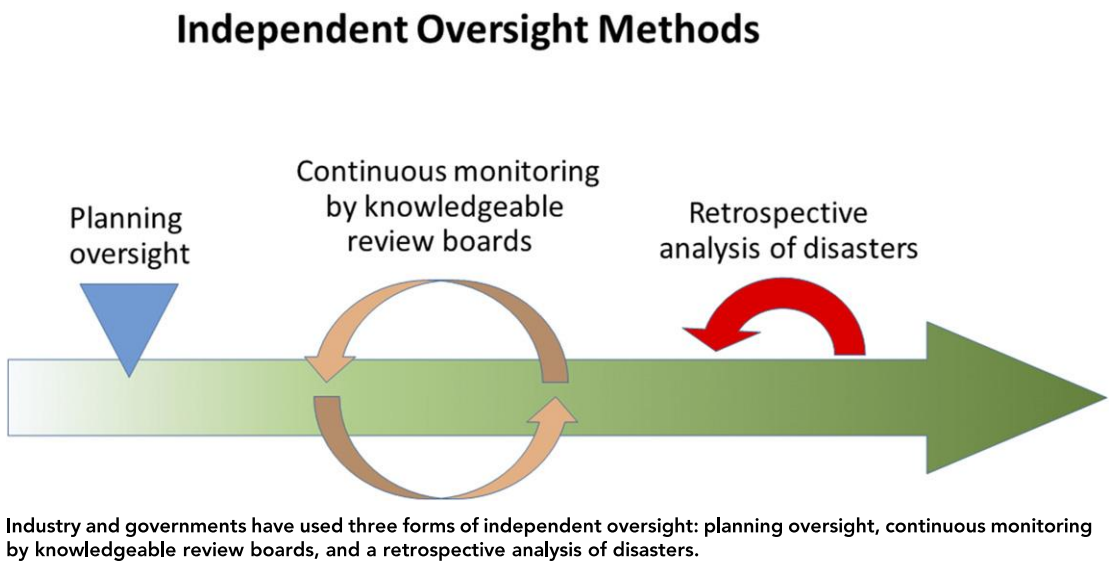
> Our findings indicate a negative differential effect on EU ventures after the rollout of GDPR relative to their US counterparts. These negative effects manifest in the overall number of financing rounds, the overall dollar amount raised across rounds, and in the dollar amount raised per individual round. Specifically, our findings suggest a $3.38 million decrease in the aggregate dollars raised by EU ventures per state per crude industry category per week, a 17.6% reduction in the number of weekly venture deals, and a 39.6% decrease in the amount raised in an average deal following the rollout of GDPR.[256]

Other authors, however, support the recommendations, and propose solutions that resemble the ones set forth in Europe. Ben Shneiderman is one such author who argues that independent oversight is essential to ensuring learner algorithms reliability. He claims that three traditional forms of oversight, common in other areas, are pertinent: (i) planning oversight, (ii) continuous monitoring, and (iii) retrospective analysis, as illustrated by the figure below:

---

[254] Ibid, p. 22.
[255] WALLACE, N. and CASTRO, D. **The Impact of the EU's New Data Protection Regulation on AI**. Center for Data Innovation, March 26, 2018, p. 2. Available at: <https://www.datainnovation.org/2018/03/the-impact-of-the-eus-new-data-protection-regulation-on-ai/>. Access: January 05, 2019.
[256] JIA, J., JIN, G. J. and WAGMAN, L. **The Short-Run Effects of GDPR On Technology Venture Investment.** NBER Working Paper no. 25248, November, 2018, p. 4. Available at: <https://www.nber.org/papers/w25248>. Access: January 07, 2019.

Figure 5 – Independent Oversight Methods.



Source: Shneiderman (2016).

The first step suggested by Shneiderman involves an impact assessment of algorithmic systems,[257] which he suggests could be similar to the environmental impact assessments required for major construction and development projects: "Algorithm impact statements would document the goals of the program, data quality control for input sources, and expected outputs so that deviations can be detected."[258] The second step is continuous monitoring by knowledgeable review boards. Shneiderman himself, during a seminar at Harvard University, recognized the high expense of this component, but insisted on its workability and relevance, emphasizing its successfully implementation in other areas.[259] He further maintains that "[v]ital systems might be under review by in-house monitors, just as Food and Drug Administration

[257] The AI Now Institute at New York University is another proponent of algorithmic impact assessment. See REISMAN, D., et al. **Algorithmic Impact Assessment: A Practical Framework for Public Agency Accountability**. AI Now, April 2018.
[258] SHNEIDERMAN, B. Opinion: **The dangers of faulty, biased, malicious algorithms requires independent oversight**. Proceedings of the National Academy of Sciences of the United States of America, November 29, 2016, p. 13539.
[259] Video of the seminar available at: SHNEIDERMAN, B. **Algorithmic Accountability: Designing for Safety.** Raddliffe Institute for Advanced Study Harvard University. March 22, 2018. Available at: <https://www.radcliffe.harvard.edu/video/algorithmic-accountability-designing-safety-ben-shneiderman>. Access: January 07, 2019.

meat and pharmaceutical inspectors continuously check on production."[260] The third step is retrospective analysis of disasters, for which the author proposes the creation of a national board in charge of algorithms safety, "[l]ike the National Transportation Safety Board, the National Algorithms Safety Board could be an independent board, outside of any government agency, with only power to investigate accidents and no authority to regulate."[261]

Goodman echoes these concerns in his suggestion that algorithmic audits be implemented, audits that would be comprised of "third party inspections of algorithmic decision-making modelled on audit studies from social sciences."[262] He further states that it is possible, although not explicit, that these audits could be requested by authorities by virtue of some of the GDPR provisions, namely: (a) the data impact assessments of Article 24, (b) the codes of conduct of Article 40, and (c) the certification of Article 42. The audits would focus on identifying procedural risks, determining whether safeguards are in place, and providing guidance for future endeavors.

These concerns gain another layer of complexity in light of the study by Kleinberg & Mullainathan. The authors examined the trade-off between explainability and fairness and came to a meaningful conclusion: more complex prediction functions are more efficient and also more equitable than simple prediction functions.[263] They very auspiciously remark that the literature on performance versus interpretability of systems is well-established and has long recognized the trade-off between them, and that the balance often times tips towards interpretability because we assume that comprehensible models are more fair and equitable. Their research revealed, however, that the opposite was true:

> in a formal sense, any attempt at simplification in fact creates inequities that a more complex model could eliminate while also improving performance.

---

[260] SHNEIDERMAN, B. Opinion: **The dangers of faulty, biased, malicious algorithms requires independent oversight**. Proceedings of the National Academy of Sciences of the United States of America, November 29, 2016, p. 13539.
[261] Ibidem.
[262] GOODMAN, B. W. **A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection**. 29th Conference on Neural Information Processing Systems. Barcelona, Spain. 2016, p. 4.
[263] KLEINBERG, J. and MULLAINATHAN, S. **Simplicity Creates Inequity: Implications for Fairness, Stereotypes, and Interpretability**. Cornell University, September 12, 2018. Available at: <https://arxiv.org/abs/1809.04578>. Access: January 05, 2019.

> Achieving interpretabilty through simplification sacrifices not only performance but also equity.[264]

What the literature herein presented makes clear is that the issue is by no means simple, and that it will not be resolved quickly. Much research is needed, and for optimal results, the private and public sectors will certainly have to cooperate closely. In the next section, I shall delve deeper into this matter looking specifically at Brazil and the policy challenges in this jurisdiction.

4.2 An Agenda for Brazil

If algorithmic discrimination as set forth in chapter 2 is a matter of concern for the data-driven economy, if the mechanisms currently in place are flawed or insufficiently grounded in jurisprudence to offer the necessary degree of protection described in chapter 3, and if the debate over algorithmic governance is gaining in urgency, as argued in 4.1, the last and crucial step of this dissertation is discussing how these factors should be absorbed and processed, notably in Brazil, so that effective solutions to the potential problems can be found. I will do so by focusing on (i) an institutional debate and (ii) a substantive perspective, which are in many ways co-dependent and complimentary, given that effective public policies require both sound institutions and sufficient legislation establishing rights and obligations.

From (i) an institutional point of view, two necessary strategies for algorithmic discrimination come to mind: (a) coherently and cohesively applying the legislation currently in place, and (b) engaging program developers and operators to build human-centered technology. Regarding (a), the actors in charge of enforcement must remain aware of the developments in science and society so that application is always connected to the real world.

---

[264] Ibid, p. 3.
The authors move on hoping to exemplify their conclusions: "A concrete example helps illustrates the equity sacrifice. Suppose that a college, to simplify its ranking of applicants, foregoes the use of college essays for all students. (This is in keeping with the type of simplification discussed above: in the representation of each applicant, the college is grouping applicants into cells by projecting out the dimension corresponding to the quality of the essay.) In doing so, it harms those disadvantaged students with excellent essays; they cannot use their essay to show their talent. Were it instead to selectively ignore essays only for the advantaged, this ends up benefiting advantaged applicants with bad essays; their disadvantaged counterparts with equally bad essays would have these essays read and factored into the decision. The machinery of our proof shows that not only are such problems endemic to simplification, they make it so that simple rules can always be modified to make them more equitable."

Moreover, clear criteria must be established, especially for those legal institutions that will apply them in real-life situations. In that light, the creation of a centralized, independent, and technically capable data protection authority would go a long way in ensuring that such requirements are met. Examples of two separate public policies help understand the Brazilian case: consumer protection and the antitrust policies. Both bear similarities to data protection inasmuch as they are transversal, meaning that the policymakers who carry out implementation are specialized in a methodology applied in several different economic sectors, not within the sectors themselves.

Whereas consumer protection in Brazil is largely decentralized, carried out simultaneously by administrative actors (such as the SENACON), the PROCONs (a foundation with offices all over the country to provide assistance and support to consumers), and the judiciary, antitrust or competition public policy is largely centralized in the hands of the Administrative Council for Economic Defense (or CADE). In a domain as technical as data protection, one that has so many implications for so many different sectors of the economy, the centralization of competences in the hands of a single body seems the best alternative – it would provide a unified forum for debates with stakeholders, would concentrate decision-making in the hands of highly specialized personnel, and facilitate coordination with other policies impacted by the provisions and actions of the authority. The Brazilian experience as illustrated by the aforementioned cases confirms this perception: competition policy as carried out by CADE has largely been considered successful, attracting both national and international recognition, and has generated a much more coherent policy than consumer protection – despite the many efforts to rationalize decision-making under the CDC.

The creation of such an authority, however, remains in the balance. The provisions that established the data protection authority in the bill that later became the LGPD were struck down by the Presidency,[265] which later enacted Executive Order n. 869/2018, creating the National

[265] Although the bill that later came to be the LGPD was proposed by the Executive, the inclusion of the data protection authority was carried out by of a member of the Legislative. The Office of the President alleged this constitutes a violation of Articles 61, § 1º, II, 'e' and 37, XIX, of the Brazilian Constitution.

Data Protection Authority (or ANPD for its Portuguese acronym) following a rather different model. The ANPD will be under the Office of the President; it will have no separate budget, nor will its directors be confirmed by the Senate. The previous institutional model was that of a so-called "autarchy," which in Brazil is a body of indirect administration, meaning that it possesses its own legal personality, a higher degree of autonomy, and a separate budget. Experts have expressed their preference for this model and their concern for the future of data protection in light of the latest developments.[266] Because the ANPD was created by Executive Order, the terms of its creation must still be confirmed by Congress, and both the Senate and the House of Representatives may introduce changes. It seems unlikely that significant modifications will be implemented, however, given that the administration that took office in 2019 has so far not put forward an agenda that addresses data protection, nor expressed any particular concern for the subject.

While well aware that this institutional issue might be settled before this dissertation enters the literature, as the Executive Order could be ratified in the coming weeks or months, I believe the observations retain significance. Policies evolve, and changes in administration or in the focus of administrators lead to modifications. We should therefore not lose sight of the institutional models that may foster better policymaking, nor cease to strive for them.

Regarding the need to engage stakeholders (b), the human-centered approach to technology has been widely discussed, especially with regards artificial intelligence. Fei-Fei Li, Stanford Professor and former Chief-Technologist at Google AI, proposes three goals for the development of machines in compliance with this ideal: first, she claims "AI needs to reflect more of the depth that characterizes our own intelligence,"[267] meaning machines must be able to incorporate context and nuances. In order to reach the first goal, computer scientists will have to collaborate with other specialists from other domains, such as psychology, neuroscience,

[266] See, for example: LEMOS, R. et al. **A criação da Autoridade Nacional de Proteção de Dados pela MP nº 869/2018.** JOTA, December 29, 2018. Available at: <https://www.jota.info/?pagename=paywall&redirect_to=//www.jota.info/opiniao-e-analise/artigos/a-criacao-da-autoridade-nacional-de-protecao-de-dados-pela-mp-no-869-2018-29122018>. Access: January 07, 2019.

[267] LI, F.-F. **How to Make A.I. That's Good for People.** The New York Times, March 7, 2018. Available at: <https://www.nytimes.com/2018/03/07/opinion/artificial-intelligence-human.html>. Access: January 11, 2018.

ethics, sociology, and so forth. Second, machines should be built to enhance humans, not to replace them. There are tasks for which machines are better suited than humans, and we should make use of their abilities in these fields to enhance our capacity in others. As Li emphasizes, "Robots may never be the ideal custodians of the elderly, but intelligent sensors are already showing promise in helping human caretakers focus more on their relationships with those they provide care for by automatically monitoring drug dosages and going through safety checklists."[268] Third, technology should always be developed with concern for the effects it might have on humans.

Regarding (ii) a substantial point of view, the first step is better educating authorities on algorithmic discrimination and its causes so that they can address it with the tools already in place. In this dissertation, that understanding is represented by the typology put forward in section 2.3.3, together with the observations in chapter 3 that show how discrimination takes place in concrete cases. The second step involves legislation and regulations to be either interpreted in order to provide answers to the challenges brought forward by algorithmic discrimination, or adapted to tackle issues that current legal doctrine is unable to address.

Here, many relevant concerns arise. One that stands out in the literature on algorithmic governance is liability. Nothing in the typology presented in chapter 2 answers rather basic legal questions: assuming discrimination of some kind did take place, and it harmed an individual, who should be held liable, and under which circumstances? Because of the characteristics of algorithmic discrimination, it makes little sense to claim that only those who somehow intended for discriminatory outcomes to occur should be held liable (though that might be an aggravating factor), and it is also difficult to determine which individual should be responsible for redressing harm for, after all, algorithms are usually joint efforts that involve engineers, designers, marketers, business people, and so on, all of whom are somehow "responsible" for the final outcome. The GDPA addresses this matter in its third section. It states that controllers and operators[269] are liable for any treatment of personal data that causes someone harm, be it in the

[268] Ibidem.
[269] As the names suggest, the operator is the person who carries out data treatment at the request of the controller. The controller is the person who makes decisions about data treatment.

form of moral damages or damages to property, individual or collective. Moreover, it establishes that controllers and operators must repair the harms, and emphasizes that solidary liability applies between controller and operator whenever both are directly involved in the treatment of data.

Still, the GDPA does not clarify is who precisely the data controller or operator is. It provides definitions but does not answer the previous question regarding the specific individual responsible for the harmful outcome. Should the legal entity be held solely liable? It seems unlikely that this is the legislation's intent, for otherwise phrasing to that effect would have been explicitly included in the law. But then who? The discussion gains relevance when we take Article 42, §4 into account, which states: "that who repairs the damage has the right of return against the other responsible actors, to the exact extent of his or her participation in the harmful event." This question will remain open-ended until rulings in concrete cases delineate a clearer landscape, but those in charge of enforcement should be aware of this difficulty, and of the challenges of individualizing liability when it comes to algorithmic discrimination. Ultimately, the decision of whom should be held liable is political, in the sense that there will be no universally right or wrong answer, only better- or worse-suited solutions for a given complex matter, solutions that may yield different results in terms of compliance.

Another topic that deserves closer inspection is discriminatory categorizing that derives from faulty collection or design. Although it is possible to state that the GDPA principle of data correction responds to the concerns raised by this type of discrimination, Brazilian legislation fails to provide a specific requirement imposing algorithm accurateness. It can be reasonably argued that the reason for the lack of regulation are the market incentives that sufficiently address the problem. In other words, no norm is needed because it is not in a company's best interests to use statistically imprecise algorithms, simply because inaccuracy of this kind would lead to inefficient resource allocation. Classifying a good creditor as bad ultimately means less revenue for the lending institution. Still, this problem is not treated in cases of public algorithmic discrimination, as no overreaching obligation exists for either private or public actors to show that their models are statistically sound. Other countries, such as Germany, do possess such a

provision – in that country, the Bundesdatenschutzgesetz (BDSG), in Section 31, states that the method used for credit scoring must follow a scientifically recognized statistical procedure.[270]

A further topic central to algorithmic discrimination is the use of proxies. As mentioned in section 2.2.1, proxies are extremely practical and very efficient, yet they pose risks. There is nothing in the Brazilian legislation that expressly speaks to these risks, which means that addressing the problem requires extremely intelligent recourse to and enforcement of the legal instruments available, or on the modification of the current body of law to insert provisions that more explicitly tackle the matter. German regulation, for example, although lacking an overall prohibition on the use of proxies, states that no score based solely on residential information shall be allowed in data processing and that, if such location data is in fact used, the data subject must be notified in advance.[271] This represents a partial solution for it deals with only one type of data – zip codes – but it could be expanded and improved to deal with other complications involving proxies.

More importantly, perhaps, is the all-embracing concern expressed in the literature on algorithmic governance regarding the need for a human-centered approach. It is clear that transparency is insufficient for effectively preempting discrimination, and that reaching explainability is by no means easy, as was demonstrated in the previous section. For those reasons, and in light of the many different alternatives put forward by specialists, creating an environment characterized by compliance and ethical standards in automated systems is challenging, and will likely take a decade-long effort on the part of public and private institutions. As such, the engagement of both policymakers and technology developers is paramount.

---

[270] The original in German states: § 31 (1, n. 2) BDSG: "Die Verwendung eines Wahrscheinlichkeitswerts über ein bestimmtes zukünftiges Verhalten einer natürlichen Person zum Zweck der Entscheidung über die Begründung, Durchführung oder Beendigung eines Vertragsverhältnisses mit dieser Person (Scoring) ist nur zulässig, wenn [...] die zur Berechnung des Wahrscheinlichkeitswerts genutzten Daten unter Zugrundelegung eines wissenschaftlich anerkannten mathematisch-statistischen Verfahrens nachweisbar für die Berechnung der Wahrscheinlichkeit des bestimmten Verhaltens erheblich sind."

[271] § 31 (1, n. 3 and 4) BDSG.

The mindful use of provisions such as Articles 10, §3 and 38 of the GDPA will be critical. The articles states that the ANPD may request an impact assessment report from the data controller, and that such report must contain at least "the description of the type of data collected, the methodology used for collection and for ensuring information security, as well as the controller's assessment of risk mitigation mechanisms."[272] I strongly hold that the careful handling of such attributions will determine the success of the data protection public policy in general, and of the prevention of algorithmic discrimination in particular. It is also crucial that these provisions be interpreted such that innovation is not hampered, and the Brazilian legislation is astute in that regard, asserting the relevance of protections for commercial and industrial secrecy. As the previous section demonstrated, explainability is not about dissecting the algorithms for competitors to see, but rather involves providing relevant information for consumers and policymakers so that future harmful discriminatory outcomes can be prevented and present ones addressed.

For that to be possible, and once again emphasizing the complementarity of institutional gravitas and robust substantive legislation, the foundations of a solid data protection authority must be laid, and the continuous debate among specialists on ethical technology must continue. There is no shortage of examples that demonstrate the importance of coherent and active enforcement for policies to exert real change in the market, notably in Brazil where no data protection tradition exists and the public and private sectors are still very much unsure whether the current legislation will be maintained and enforced. As Cathy O'Neil puts it:

> Fairness is squishy and hard to quantify. It is a concept. And computers, for all of their advances in language and logic, still struggle mightily with concepts. They 'understand' beauty only as word associated with the Grand Canyon, ocean sunsets, and grooming tips in Vogue magazine. They try in vain to measure 'friendship' by counting likes and connection on Facebook. And the

[272] The original in Portuguese reads: Art. 38: A autoridade nacional poderá determinar ao controlador que elabore relatório de impacto à proteção de dados pessoais, inclusive de dados sensíveis, referente a suas operações de tratamento de dados, nos termos de regulamento, observados os segredos comercial e industrial.
Parágrafo único. Observado o disposto no caput deste artigo, o relatório deverá conter, no mínimo, a descrição dos tipos de dados coletados, a metodologia utilizada para a coleta e para a garantia da segurança das informações e a análise do controlador com relação a medidas, salvaguardas e mecanismos de mitigação de risco adotados.
Art. 10, § 3º: A autoridade nacional poderá solicitar ao controlador relatório de impacto à proteção de dados pessoais, quando o tratamento tiver como fundamento seu interesse legítimo, observados os segredos comercial e industrial.

> concept of fairness utterly escapes them. Programmers don't know how to code for it, and few of their bosses ask them to.[273]

It is imperative we comprehend this difficult and invest on programs that are able to capture its nuances. The Law can help, but much engagement from many specialists will be of prime importance.

[273] O'NEIL, C. **Weapons of Math Destruction – How Big Data Increases Inequality and Threatens Democracy**. New York: Crown, 2016, p. 95.

References

ACHARYA, V. **Regulating wall Street: the Dodd-Frank Act and the new architecture of global finance**. Interviewer: DAVIES, V. VOX - CEPR Policy Portal. October 22, 2010. Available at: <https://voxeu.org/vox-talks/regulating-wall-street-dodd-frank-act-and-new-architecture-global-finance>.

AGRAWAL, A., GANS, J. and GOLDFARB, A. **Prediction Machines: The Simple Economics of Artificial Intelligence**. Harvard Business Press, April 17, 2018.

ANANNY, M. and CRAWFORD, K. **Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability**. SAGE journals. December 13, 2016. Available at: <https://journals.sagepub.com/doi/abs/10.1177/1461444816676645?journalCode=nmsa>.

ANBC. **Por que a ANBC?**. Available at: <https://www.anbc.org.br/materias.php?cd_secao=3#5&friurl=_-Empresas-associadas-_>.

ANDERSON, C. **The End of the Theory: the Data Deluge Makes the Scientific Method Obsolete**. Wired. June 23, 2008. Available at: <https://www.wired.com/2008/06/pb-theory/>.

ANGWIN, Julia; LARSON, Jeff; MATTU, Surya; KIRCHNER, Lauren. **Machine Bias**. ProPublica, May 23, 2016. Available at: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

ASSOCIATION FOR COMPUTING MACHINERY US PUBLIC POLICY COUNCIL (USACM). **Statement on Algorithmic Transparency and Accountability.** January 12, 2017. Available at: <http://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf>.

AUSTRALIAN GOVERNMENT. **Automated Assistance in Administrative Decision-Making – Better Practice Guide.** February, 2007. Available at: <https://www.oaic.gov.au/images/documents/migrated/migrated/betterpracticeguide.pdf>.

BANERJEE, A. DUFLO, E. GLENNERSTER, R. and KINNAN, C. **The miracle of microfinance? Evidence from a randomized evaluation**. Northwestern University of Economics and NBER. March, 2014. Available at: <https://economics.mit.edu/files/5993>.

BAROCAS, S. and SELBST, A. D. **Big Data's Disparate Impact.** p. California Law Review, vol. 671, 2016. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899>.

BERGER, M. M. **To Regulate, or Not to Regulate – Is That the Question: Reflections on the Supposed Dilemma between Environmental Protection and Private Property Rights**.

8 Loy, L. A. L. Rev. 253, 1975. Available at: <https://digitalcommons.lmu.edu/cgi/viewcontent.cgi?referer=https://scholar.google.com.br/scholar?hl=en&as_sdt=0%2C5&q=to+regulate+or+not&btnG=&httpsredir=1&article=1187&context=llr>.

BILBAO UBILLOS, J. M. **La Eficacia de Los Derechos Fundamentales Frente a Particulares: Analisis de La Jurisprudencia del Tribunal Constitucional.** Centro de Estudios Políticos y Constitucionales, 1997.

BLASS A. and GUREVICH, Y.. **Algorithms: A Quest for Absolute Definitions**. Bulletin of European Association for Theoretical Computer Science. vol. 81, 2003. Available at: <https://web.eecs.umich.edu/~gurevich/Opera/164.pdf>.

BOYD, D. and CRAWFORD, K. **Six Provocations for Big Data.** A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, September 2011. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431>.

BRILL, C. J. **Scalable Approaches to Transparency and Accountability in Decision Making Algorithms, Remarks at the NYU Conference on Algorithms and Accountability.** FTC. February 28, 2015. Available at: <https://www.ftc.gov/public-statements/2015/02/scalable-approaches-transparency-accountability-decisionmaking-algorithms>.

BULLUNCK, M. **Histories of algorithms: Past, present and future**. Historia Mathematica, Elsevier, 2015, 43 (3).

CENTRO DE ESTUDOS DA MTERÓLE (CEM). **Mapa da Vulnerabilidade Social da População da Cidade de São Paulo.** Centro Brasileiro de Análise e Planejamento-CEBRAP, do Serviço Social do Comércio-SFSC e da Secretaria Municipal de Assistência Social de São Paulo, SAS-PMSP. São Paulo, 2004. Available at: <http://web.fflch.usp.br/centrodametropole/upload/arquivos/Mapa_da_Vulnerabilidade_social_da_pop_da_cidade_de_Sao_Paulo_2004.pdf>.

CITRON, D. K. Citron, **Technological Due Process**, 85 Wash. U. L. Rev. 1249, 2008. Available at: <http://openscholarship.wustl.edu/law_lawreview/vol85/iss6/2>.

CITRON, D. K. and PASQUALE, F. **The Scored Society: Due Process For Automated Predictions.** Washington Law Review, vol. 89:1, 2014. Available at: <https://digital.law.washington.edu/dspace-law/bitstream/handle/1773.1/1318/89WLR0001.pdf>.

COMMISSION OF THE EUROPEAN COMMUNITIES. **Council Directive on implementing the principle of equal treatment between persons irrespective of religion or belief, disability, age or sexual orientation**. Brussels, 2008. Available at: <https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52008PC0426>.

COMMUNICATIONS COMMITTEE. **The Internet: to regulate or not to regulate? inquiry**. Parliamentary business. Available at: <https://www.parliament.uk/business/committees/committees-a-z/lords-select/communications-committee/inquiries/parliament-2017/the-internet-to-regulate-or-not-to-regulate/>.

CORMEN, T. H., **Algorithms Unlocked**. MIT Press, 2013.

COUNCIL OF EUROPE PORTAL. **Algorithms and Human Rights: a new study has been published**. May 22, 2018. Available at: <https://www.coe.int/en/web/freedom-expression/-/algorithms-and-human-rights-a-new-study-has-been-published>.

CRAWFORD, H. **The Hidden Biases in Big Data**. Harvard Business Review. April 01, 2013. Available at: <https://hbr.org/2013/04/the-hidden-biases-in-big-data>.

CUMMINGS, M. L. **The Social and Ethical Impact of Decision Support Interface Design**. International Encyclopedia of Ergonomics and Human Factors. 2005. Available at: <https://pdfs.semanticscholar.org/a9b3/ec436508ebfa40f3a3f5b59231bece4f3e34.pdf>.

DEMSETZ, H. **Why Regulate Utilities?**. Journal of Law and Economics, vol. 11, no. 1, The University of Chicago Press Journals, April, 1968. Available at: <https://www.journals.uchicago.edu/doi/abs/10.1086/466643?journalCode=jle>.

DEPARTMENT FOR DIGITAL, CULTURE, MEDIA & SPORT of the UK Government. **Consultation on the Centre for Data Ethics and Innovation**. November 20, 2018. Available at: <https://www.gov.uk/government/consultations/consultation-on-the-centre-for-data-ethics-and-innovation/centre-for-data-ethics-and-innovation-consultation>.

DIAKOPOULOS, N. and FIEDLER, S. **How to Hold Algorithms Accountable**. MIT Technology Review. November, 2016.

DIAKOPOULOS, N., FIEDLER, S., ARENAS, M. et al. **Principles for Accountable Algorithms and a Social Impact Statement for Algorithms**. FAT/ML. Available at: <https://www.fatml.org/resources/principles-for-accountable-algorithms>.

DOMINGOS, P. **Master Algorithm**. Basic Books Inc. New York, 2018.

DONEDA, D. and MENDES, L. S. **Data Protection in Brazil: New Developments and Current Challenges**. In: GURWIRTH, S., LEENES, R. and HERT, P. De. (Eds). *Reloading Data Protection: Multidisciplinary Insights and Contemporary Challenges*. Springer, 2014.

DOSHI-VELEZ, F. and KORTZ, M. **Accountability of AI Under the Law: The Role of Explanation**. Berkman Klein Center Working Group on Explanation and the Law, Berkman

Klein Center for Internet & Society working paper, 2017. Available at : <https://dash.harvard.edu/bitstream/handle/1/34372584/2017-11_aiexplainability-1.pdf?sequence=3>.

DÜRIG, G. **Grundrechte und Zivilrechtsprechung.** Vom Bonner Grundgesetz zur gesamtdeutschen Verfassung : Festschrift zum 75. Geburtstag von Hans Nawiasky, 1956.

EDWARDS, L. and VEALE, L. **Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For**. 16 Duke Law & Technology Review 18, 2017. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2972855&download=yes>.

ELLIS, E. and WATSON, P. **EU Anti-Discrimination Law - Second Edition**. Oxford EU Law Library, 2012.

ELLROTT, J. TRITTMANN, R. and WERMEISTER, C. **Automated driving law passed in Germany**. June 21, 2017. Available at: <https://www.freshfields.com/en-gb/our-thinking/campaigns/digital/internet-of-things/connected-cars/automated-driving-law-passed-in-germany/>.

EUBANKS, V. **Automating Inequality: How High-Tech Tools Profile, Police, and Punish de Poor**. St. Martin's Press. January 23, 2018.

EUROPEAN COMMISSION. **Charter of Fundamental Rights: the Presidents of the Commission, European Parliament and Council sign and solemnly proclaim the Charter in Strasbourg**. Brussels, December 12, 2007. Available at: <http://europa.eu/rapid/press-release_IP-07-1916_en.htm>.

EUROPEAN COMMISSION'S HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE (AI HELG). **Draft Ethics Guidelines for Trustworthy AI**. European Commission, Brussels, March, 2019 (final version). Available at: <https://ec.europa.eu/futurium/en/node/6044>.

FEDERAL TRADE COMMISSION. **Big Data, A Tool for Inclusion or Exclusion? – Understanding the Issues**. January, 2016. Available at: <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf>.

FUSTER, G. G. **The Emergence of Personal Data Protection as a Fundamental Right of the EU**. Springer International Publishing, 2014.

GARDBAUM, S. **The 'Horizontal Effect' of Constitutional Rights**. Michigan Law Review, vol. 102, UCLA School of Law Research Paper No. 03-14, pp. 388-459, 2003. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=437440>.

GASSER, U. and ALMEIDA, V. A.F. **A Layered Model for AI Governance**. IEEE Internet Computing 21 (6) (November): 2017.

GILLESPIE, T. **Chapter 2- Algorithm**. In: PETERS, B. (Ed.). *Digital Keywords: a Vocabulary of information society and culture.* Princeton: Princeton University Press, 2016.

GOODMAN, B. W. **A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection**. 29th Conference on Neural Information Processing Systems. Barcelona, Spain, 2016.

GORZONI, P. F. A. da C. **Supremo Tribunal Federal e a Vinculação dos Direitos Fundamentais nas Relações entre Particulares**. 2007. Available at: <http://www.sbdp.org.br/publication/supremo-tribunal-federal-e-a-vinculacao-dos-direitos-fundamentais-nas-relacoes-entre-particulares/>.

GUNNING, D. **Explained Artificial Intelligence (XAI).** DARPA/I20. November, 2017. Available at: <https://www.darpa.mil/attachments/XAIProgramUpdate.pdf>.

GUREVICH, Y. **What Is an Algorithm?**. In: BIELIKOVÁ M., FRIEDRICH, G., GOTTLOB, G., KATZENBEISSER, S., TURÁN, G. (eds) *SOFSEM 2012: Theory and Practice of Computer Science*. SOFSEM 2012. Lecture Notes in Computer Science, vol. 7147. Springer, Berlin, Heidelberg, 2012. Available at: <https://www.researchgate.net/publication/221512843_What_Is_an_Algorithm>.

HARFORD, T. **Big data: are we making a big mistake?**. The Financial Times. 2014. Available at: <https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1740-9713.2014.00778.x>.

HAWKINS, J. and BLAKESLEE, S. **On Intelligence: How a New Understanding the Brain Will Lead to the of the Creation of Truly Intelligent Machines.** Times Books, 2004.

HESSE, K. **Verfassungsrecht und Privatrecht**. Müller Jur. Vlg. C. F., 1988.

HUET, E. **Server And Protect: Predictive Policing Firm PredPol Promises To Map Crime Before It Happens**. Forbes, February 11, 2015. Available at: <https://www.forbes.com/sites/ellenhuet/2015/02/11/predpol-predictive-policing/#6efe33ae4f9b>.

IAPP and THE UNITED NATIONS GLOBAL PULSE. **Building Ethics into Privacy Frameworks for Big Data and AI.** Available at: <https://iapp.org/media/pdf/resource_center/BUILDING-ETHICS-INTO-PRIVACY-FRAMEWORKS-FOR-BIG-DATA-AND-AI-UN-Global-Pulse-IAPP.pdf>.

ITS. **Transparência e governança nos algoritmos: um estudo de caso sobre o setor de birôs de crédito.** Rio de Janeiro, 2017.

JIA, J., JIN, G. J. and WAGMAN, L. **The Short-Run Effects of GDPR On Technology Venture Investment.** NBER Working Paper no. 25248, November, 2018. Available at: <https://www.nber.org/papers/w25248>.

JUSTICE AND CONSUMERS. **Guidelines on Automated Individual decision-making and Profiling for the purposes of Regulation 2016/679**. Adopted on October 03, 2017. Available at: <https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053>.

KELLSTEDT, P. M. and WHITTEN, G. **The Fundamentals of Political Research**. New York: Cambridge University Press, 2009.

KLEINBERG, J. and MULLAINATHAN, S. **Simplicity Creates Inequity: Implications for Fairness, Stereotypes, and Interpretability**. Cornell University, September 12, 2018. Available at: <https://arxiv.org/abs/1809.04578>.

KNIGHT, W. **The Dark Secret at the Heart of AI**. MIT Technology Review. April 11, 2017. Available at: <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/>.

KROLL, J. et al. **Accountable Algorithms**. 165 U. PA. L. Rev. 633, 2017. Available at: <https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3/>.

LAZER, D. et. al. **The Parable of Google Flu: Traps in Big Data Analysis**. Science 343:1203-1205, March 14, 2014. Available at: <https://gking.harvard.edu/files/gking/files/0314policyforumff.pdf>.

LEMOS, R. et al. **A criação da Autoridade Nacional de Proteção de Dados pela MP nº 869/2018.** JOTA, December 29, 2018. Available at: <https://www.jota.info/?pagename=paywall&redirect_to=//www.jota.info/opiniao-e-analise/artigos/a-criacao-da-autoridade-nacional-de-protecao-de-dados-pela-mp-no-869-2018-29122018>.

LERMAN, J. **Big Data and Its Exclusions**. Stanford Law Review Online, 2013.

LESSIG, L. **Against Transparency.** The New Republic, 2009. Available at: <https://newrepublic.com/article/70097/against-transparency>.

LEVIN, S. **New AI can guess whether you're gay or straight from a photograph.** The Guardian. September 08, 2017. Available at:

<https://www.theguardian.com/technology/2017/sep/07/new-artificial-intelligence-can-tell-whether-youre-gay-or-straight-from-a-photograph>.

LI, F.-F. **How to Make A.I. That's Good for People.** The New York Times, March 7, 2018. Available at: <https://www.nytimes.com/2018/03/07/opinion/artificial-intelligence-human.html>.

LO, H. **The Judicial Duty to Give reasons.** Legal Studies, vol. 20, issue 1, March 2000.

LODGE, M. and STIRTON, L. **Accountability in the Regulatory State.** In: BALDWIN, R., CAVE, M. and LODGE, M. (Eds). The Oxford Handbook of Regulation, Chapter 15, September 2010.

MATTIUZZO, M. **Voto Vencido, Fundamentação Diversa e Fundamentação Complementar: um estudo sobre deliberação no Supremo Tribunal Federal.** 2011. SBDP. Available at: <http://www.sbdp.org.br/publication/voto-vencido-fundamentacao-diversa-e-fundamentacao-complementar-um-estudo-sobre-deliberacao-no-supremo-tribunal-federal/>.

MAYER-SCHÖNBERGER, V. and CUKIER, K. **Big Data: A Revolution That Will Transform How We Live, Work, Think**. Houghton Mifflin Harcourt, 2013.

MAYER-SCHÖNBERGER, V. and CUKIER, K. **The Rise of Big Data: How It's Changing the Way We Think**. Foreign Affairs, vol. 92, no. 3, May/June, 2013. Available at: <https://www.jstor.org/stable/23526834?seq=1#page_scan_tab_contents>.

MCTIC. **Estratégia Brasileira para a Transformação Digital (E-Digital).** Brasília, 2018. Available at: <http://www.mctic.gov.br/mctic/export/sites/institucional/estrategiadigital.pdf>.

MELLO, J. M. P., MENDES, M. and KANCZUK, F. **Cadastro Positivo e democratização do crédito**. Folha de São Paulo, March, 2018. Available at: <https://www1.folha.uol.com.br/opiniao/2018/03/joao-manoel-pinho-de-mello-marcos-mendes-e-fabio-kanczuk-cadastro-positivo-e-democratizacao-do-credito.shtml>.

MENDES, L. S. **Habeas Data e autodeterminação informativa: os dois lados de uma mesma moeda**. In: Centro de Direito, Internet e Sociedade do Instituto Brasiliense de Direito Público (CEDIS/IDP) (Orgs). *Internet & Regulação* Saraiva. To be published.

MILLER, A. P. **Want Less-Biased Decisions? Use Algorithms.** Harvard Business Review. July 26, 2018. Available at: <https://hbr.org/2018/07/want-less-biased-decisions-use-algorithms>.

MINISTÉRIO DA TRANSPARÊNCIA, FISCALIZAÇÃO E CONTROLADORIA-GERAL DA UNIÃO. **Aplicação da Lei de Acesso à Informação na Administração Pública Federal**.

2a. ed. rev., atu. e amp. Brasília. 2016. Available at: <http://www.acessoainformacao.gov.br/central-de-conteudo/publicacoes/arquivos/aplicacao_lai_2edicao.pdf>.

NEW, J. and CASTRO, D. **How Policymakers can foster algorithmic accountability**. Center for Data Innovation, May 2018. Available at: <http://www2.datainnovation.org/2018-algorithmic-accountability.pdf>.

NIKLAS, J. SZTANDAR-SZTANDERSKA, K. and SZYMIELEWICZ, K. **Profiling the Unemployed in Poland: social and Political Implications of Algorithmic Decision Making.** Fundacja Panoptykon. Warsaw, 2015.

NISSENBAUM, H. **Computing and Accountability**. In: COHEN, J. Communications of the ACM, vol. 37, issue 1, January 1994.

OETER, S. **Fundamental Rights and Their Impact on Private Law – Doctrine and Practice Under the German Constitution**. 12 Tel Aviv U. Stud. L. 7, 1994. Available at: <https://heinonline.org/HOL/LandingPage?handle=hein.journals/telavusl12&div=4&id=&page=>.

O'NEIL, C. **Weapons of Math Destruction – How Big Data Increases Inequality and Threatens Democracy.** New York: Crown, 2016.

QUOD. Available at: <https://www.quod.com.br/>.

PANOPTYKON. Available at: <https://en.panoptykon.org/about>.

PARASURAMAN, R. and MILLER, C. A. **Trust and etiquette in high-criticality automated systems**. Communications of the ACM – Human-computer etiquette, vol. 47 issue 4, April, 2004. Available at: <https://dl.acm.org/citation.cfm?id=975844>.

PASQUALE, F. **The Black Box Society**. Harvard University Press. January, 2015.

REISMAN, D., et al. **Algorithmic Impact Assessment: A Practical Framework for Public Agency Accountability**. AI Now, April 2018.

RIOS, R. R. and SILVA, R. da. **Democracia e direito da antidiscriminação: interseccionalidade e discriminação múltipla no direito brasileiro**. Cienc. Cult., São Paulo, vol. 69, n. 1, p. 44-49, March, 2017. Available at: <http://cienciaecultura.bvs.br/scielo.php?script=sci_arttext&pid=S0009-67252017000100016&lng=en&nrm=iso>.

SANDIG, C. et al. **An Algorithm Audit.** Data and Discrimination: Collected Essays. 2014. Available at: <http://www-personal.umich.edu/~csandvig/research/An%20Algorithm%20Audit.pdf>.

SARLET, I. W. **Direitos Fundamentais e Direito Privado: algumas considerações em torno da vinculação dos particulares aos direitos fundamentais**. Revista dos Tribunais Online. Available at: <http://www.direitocontemporaneo.com/wp-content/uploads/2018/03/SARLET-Direitos-fundamentais-e-direito-privado.pdf>.

SARMENTO, D. **Direitos Fundamentais e Relações Privadas**. Rio de Janeiro: Lumen Juris, 2004.

SCIENCE AND TECHNOLOGY COMMITTEE of the House of Commons. **Algorithms in decision-making**. Fourth Report of Session 2017-19. May 23, 2018. Available at: <https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf>.

SEATTLE INFORMATION TECHNOLOGY. **About the Surveillance Ordinance**. September 1, 2017. Available at: <https://www.seattle.gov/tech/initiatives/privacy/surveillance-technologies/about-surveillance-ordinance>.

SCHAUER, F. **Profiles, Probabilities, and Stereotypes**. Belknapp Press. April, 2016.

SHNEIDERMAN, B. **Algorithmic Accountability: Designing for Safety.** Raddliffe Institute for Advanced Study Harvard University. March 22, 2018. Available at: <https://www.radcliffe.harvard.edu/video/algorithmic-accountability-designing-safety-ben-shneiderman>.

SHNEIDERMAN, B. Opinion: **The dangers of faulty, biased, malicious algorithms requires independent oversight**. Proceedings of the National Academy of Sciences of the United States of America, November 29, 2016

SILVA, V. A. da. **A Constitucionalização do Direito: os direitos fundamentais nas relações entre particulares.** São Paulo: Malheiros Editores, 2014.

SILVA, V. A. da. **O proporcional e o razoável**. Revista dos Tribunais, 2002.

SILVER, D. et al. **Mastering the game of Go with deep neural networks and tree search.** Nature, vol. 529, January 28, 2016. Available at: <https://storage.googleapis.com/deepmind-media/alphago/AlphaGoNaturePaper.pdf>.

SILVER, D. et al. **Mastering the game of Go without human knowledge**. Nature, vol. 550, January 19, 2017. Available at:

<https://www.nature.com/articles/nature24270.epdf?author_access_token=VJXbVjaSHxFoctQQ4p2k4tRgN0jAjWel9jnR3ZoTv0PVW4gB86EEpGqTRDtpIz-2rmo8-KG06gqVobU5NSCFeHILHcVFUeMsbvwS-lxjqQGg98faovwjxeTUgZAUMnRQ>.

SIMON, H. A. **Spurious Correlation: A Causal Interpretation**. Journal of the American Statistical Association, vol. 49, September 1954. Available at: <http://digitalcollections.library.cmu.edu/awweb/awarchive?type=file&item=33513>.

SUPREMO TRIBUNAL FEDERAL. **Ministra Carmen Lúcia anuncia início de funcionamento do Projeto Victor, de inteligência artificial**. Notícias STF. August 30, 2018. Available at: <http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=388443>.

THE COMMITTEE OF EXPERTS ON INTERNET INTERMEDIARIES (MSI-NET) of the Council of Europe. **ALGORITHMS AND HUMAN RIGHTS - Study on the Human Rights Dimensions of Automated Data Processing Techniques (in particular algorithms) and Possible Regulatory Implications**. Council of Europe. March, 2018. Available at: <https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5>.

THE ECONOMIST. **How an algorithm may decide your career.** June 21, 2018. Available at: <https://www.economist.com/business/2018/06/21/how-an-algorithm-may-decide-your-career>.

THE GUARDIAN. **The Cambridge Analytica Files – A year-long investigation into Facebook, data, and influencing elections in the digital age.** Available at: <https://www.theguardian.com/news/series/cambridge-analytica-files>.

THE NEW YORK CITY COUNCIL. **Automated decision systems used by agencies – Law number 2018/049**. January 11, 2018. Available at: <https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0>.

TRIBUNAL DE JUSTIÇA DO ESTADO DE MINAS GERAIS. **TJMG utiliza inteligência artificial em julgamento virtual**. November 07, 2018. Available at: <https://www.tjmg.jus.br/portal-tjmg/noticias/tjmg-utiliza-inteligencia-artificial-em-julgamento-virtual.htm#.XC1Vby2ZO1s>.

TUSHNET, M. **The issue of state action/horizontal effect in comparative constitutional law**. Oxford University Press and New York University School of Law, vol. 1, number 1, 2003.

VIGEN, Tyler. **EBay's Total Gross Merchandise Volume Correlates With Visitors to Disney Worlds Animal Kingdom.** Tylervigen, Spurious Correlations. Available at: <http://tylervigen.com/view_correlation?id=28571>.

VIGEN, T. **Spurious correlations.** Available at: <http://www.tylervigen.com/spurious-correlations>.

WALLACE, N. and CASTRO, D. **The Impact of the EU's New Data Protection Regulation on AI**. Center for Data Innovation, March 26, 2018. Available at: <https://www.datainnovation.org/2018/03/the-impact-of-the-eus-new-data-protection-regulation-on-ai/>.

WORLD WIDE WEB FOUNDATION. **Algorithmic Governance: Applying the Concept to Different Country Contexts**. July, 2017.

ZARSKY, T. Z. **Transparent Predictions.** University of Illinois Law Review, vol. 2013, no. 4. Available at: <https://www.illinoislawreview.org/wp-content/ilr-content/articles/2013/4/Zarsky.pdf>.

USA Cases:

Blum v. Yaretsky, 457 U.S. 991 (1982).

Griggs v. Duke Power Co., 401 U.S. 424 (1971).

Marsh v. Alabama, 326 U.S. 501 (1946).

Reitman v. Mulkey, 387 U.S. 369 (1967).

Shelley v. Kraemer, 334 U.S. 1 (1948).

German Cases:

BUNDESVERFASSUNGSGERICHT. **Decision in "Stadium Ban" proceedings clarifies the indirect horizontal effects of the right to equality in private law relations.** Press Release No. 29/2018, April 27, 2018. Available at: <https://www.bundesverfassungsgericht.de/SharedDocs/Pressemitteilungen/EN/2018/bvg18-029.html>.

BUNDESVERFASSUNGSGERICHT. **Headnotes to the Order of the Firste Senate of 11 April 2018 – 1 BvR 3080/09.** Available at:

<https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/EN/2018/04/rs20180411_1bvr308009en.html;jsessionid=34F6550A81C53BC9257433E87D7CF087.1_cid393>.

BVerfGE, 7, 198.

BVerfGE 65,1, Volkszählung.

Brazilian Cases:

5ª PROMOTORIA DE JUSTIÇA DE TUTELA COLETIVA DE DEFESA DO CONSUMIDOR E DO CONTRIBUINTE DA CAPITAL. Petição inicial, Inquérito Civil n. 347/5ª PJDC/2016.

MINISTÉRIO DA JUSTIÇA. Nota Técnica n. 92. 2018, §39. Available at: <http://www.cmlagoasanta.mg.gov.br/abrir_arquivo.aspx/PRATICAS_ABUSIVAS_DECOLARCOM?cdLocal=2&arquivo=%7BBCA8E2AD-DBCA-866A-C8AA-BDC2BDEC3DAD%7D.pdf>.

STF. Recurso Extraordinário 201.819-8, Rio de Janeiro, 2005. Available at: <http://redir.stf.jus.br/paginadorpub/paginador.jsp?docTP=AC&docID=388784>.

Legislation:

BRAZIL. **Constitution.** 1988. Available at: <http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm>.

BRAZIL. **Bill n. 7.716.** January 05, 1989. Available at: <http://www.planalto.gov.br/ccivil_03/LEIS/L7716.htm>.

BRAZIL. **Bill n. 8.078 (CDC)**. September 11, 1990. Available at: <http://www.planalto.gov.br/ccivil_03/Leis/L8078compilado.htm>.

BRAZIL. **Bill n. 9.029.** April 13, 1995. Available at: <http://www.planalto.gov.br/CCIVIL_03/LEIS/L9029.HTM>.

BRAZIL. **Bill n. 10.406 (Civil Code).** January 10, 2002. Available at: <http://www.planalto.gov.br/ccivil_03/leis/2002/l10406.htm>.

BRAZIL. **Bill n. 12.288..** July 20, 2010. Available at: <http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2010/Lei/L12288.htm>.

BRAZIL. **Bill n. 12.414**. June 09, 2011. Available at: <http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2011/Lei/L12414.htm>.

BRAZIL. **Bill n. 12.527 (LAI).** November 18, 2011. Available at: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm>.

BRAZIL. **Bill n. 12.711**. August 29, 2012. Available at: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12711.htm>.

BRAZIL. **Bill n. 12.965**, April, 23, 2014. Available at: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm>.

BRAZIL. **Bill n. 13.709 (LGPD)**. August 14, 2018. Available at: <http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm>.

BRAZIL. **Draft Bill n. 441.** 2017. Available at: <http://www.camara.gov.br/proposicoesWeb/fichadetramitacao?idProposicao=2160860>.

BRAZIL. **Decree n. 7.724**. May 16, 2012. Available at: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/Decreto/D7724.htm>.

BRAZIL. **Decree n. 8.771**. May 11, 2016. Available at: <http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2016/Decreto/D8771.htm>.

BRAZIL. **Executive Order n. 869**. December 27, 2018. Available at: <http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm>.

EUROPEAN UNION. **Regulation (EU) 2016/679 (General Data Protection Regulation)**. May 25, 2018. Available at: <https://gdpr-info.eu/>.

POLAND. **The Constitution of the Republic of Poland**. April 02, 1997. Available at: <https://www.sejm.gov.pl/prawo/konst/angielski/kon1.htm>.