



Comparative Law Review

2019

ISSN:2983 - 8993

COMPARATIVE LAW REVIEW

The Comparative Law Review is a biannual journal published by the
I. A. C. L. under the auspices and the hosting of the University of Perugia Department of Law.

Office address and contact details:

Department of Law - University of Perugia
Via Pascoli, 33 - 06123 Perugia (PG) - Telephone 075.5852437
Email: complawreview@gmail.com

EDITORS

Giuseppe Franco Ferrari
Tommaso Edoardo Frosini
Pier Giuseppe Monateri
Giovanni Marini
Salvatore Sica
Alessandro Somma

EDITORIAL STAFF

Fausto Caggia
Giacomo Capuzzo
Cristina Costantini
Virgilio D'Antonio
Sonja Haberl
Edmondo Mostacci
Valentina Pera
Giacomo Rojas Elgueta
Tommaso Amico di Meane

REFEREES

Salvatore Andò
Elvira Autorino
Ermanno Calzolaio
Diego Corapi
Giuseppe De Vergottini
Tommaso Edoardo Frosini
Fulco Lanchester
Maria Rosaria Marella
Antonello Miranda
Elisabetta Palici di Suni
Giovanni Pascuzzi
Maria Donata Panforti
Roberto Pardolesi
Giulio Ponzanelli
Andrea Zoppini
Mauro Grondona

SCIENTIFIC ADVISORY BOARD

Christian von Bar (Osnabrück)
Thomas Duve (Frankfurt am Main)
Erik Jayme (Heidelberg)
Duncan Kennedy (Harvard)
Christoph Paulus (Berlin)
Carlos Petit (Huelva)
Thomas Wilhelmsson (Helsinki)

Comparative Law Review is registered at the Courthouse of Monza (Italy) - Nr. 1988 - May, 10th 2010.

COMPARATIVE
LAW
REVIEW
FALL VOL. 10 /2

6

DIEGO LÓPEZ MEDINA

El proceso judicial europeo en América Latina: crítica de la Historia oficial y propuesta de nuevas perspectivas

80

CLÁUDIA TOLEDO

Judicial activism or judicial review:
distinction criteria and comparative analysis: Brazil, Argentina, Mexico,
Germany

106

TOMASZ GIARO

A Matter of Pure Conscience? Franz Wieacker And his “Conceptual
Change”

125

GIACOMO CAPUZZO

A Comparative Study on Algorithmic Discrimination between Europe and
North-America

158

ANNA PARRILLI

Religious Education in public schools and minority rights: a comparative
study of Italy and Turkey

174

ANDREA SEBASTIANI

Anatomy and terminology of business organizations: a brief comparison
and assimilation between Chinese and Italian companies

A COMPARATIVE STUDY ON ALGORITHMIC DISCRIMINATION BETWEEN EUROPE AND NORTH-AMERICA

Giacomo Capuzzo

TABLE OF CONTENTS:

SOME INTRODUCTORY NOTES - I.1 DISCRIMINATION AND ALGORITHMS: A CULTURAL CONTEXT AND TWO LEGAL APPROACHES - I.2 TRANSPARENCY AND RESPONSIBILITY BETWEEN LAW AND IT - II.1 ANTI-DISCRIMINATION TESTED BY ALGORITHMIC DECISION-MAKING PROCESSES - II.2 THE DISCRIMINATORY POTENTIAL OF ALGORITHMS: A PRACTICAL GUIDE - II.3 MULTILEVEL ANTI-DISCRIMINATION PROTECTION - II.4 FIGHTING ALGORITHMIC DISCRIMINATION THROUGH THE LEGISLATION ON THE PROCESSING OF PERSONAL DATA - II.5 OTHER EXAMPLES OF REGULATIONS AT EUROPEAN LEVEL - II.6 THE ANTI-DISCRIMINATION PROTECTION IN THE UNITED STATES OF AMERICA IN RELATION TO THE USE OF ARTIFICIAL INTELLIGENCE - III. ALGORITHMIC DISCRIMINATION IN PRACTICE: A SERIES OF SUBSTANTIVE CASES - BY WAY OF CONCLUSION.

This study aims at describing the legal framework that relates discrimination to the world of artificial intelligence, with particular reference to the use of algorithms. The author analyzes the operation of these automatic machines within the decision-making processes of private and public actors, emphasizing the potential discriminatory effects resulting from such use. The survey of the regulation on the subject has suggested a dichotomy of approaches, one ex post grounded on general principles that can be applied at discretion to different cases; the other ex ante based on sets of rules set up for specific areas. The paper explores the various aspects of discriminatory protection between the European and US multilevel legal systems, deepening the regulatory disciplines and dealing with some practical cases to provide an introductory map to the issue.

SOME INTRODUCTORY NOTES

The growing use of artificial intelligence within the productive sectors and public administration marked the first decades of the new century. Technological development has allowed the increasing use of new technologies capable of performing support tasks and sometimes in place of human activity.¹

¹ On law and artificial intelligence generally, see S. Rodotà, *Elaboratori elettronici e controllo sociale*, Bologna, 1973; F. Pasquale, *The black Box Society: The Secret Algorithms That Control Money And Information*, Cambridge 2015, 5 Ff.; S. Barocas, A. D. Selbst, *Big Data's Disparate Impact*, 104 Calif. L. Rev. 671, 674 N.10 (2016); D. Poole, A. Mackworth, R. Goebel, *Computational Intelligence: A Logical Approach*, New York 1998; I. Ajunwa, *Algorithms At Work: Productivity Monitoring Platforms And Wearable Technology As The New Data-Centric Research Agenda For Employment And Labor Law*, 63 St. Louis U. L.J. 2019; I. Ajunwa, *Genetic Testing Meets Big Data: Tort And Contract Law issues*, 75 Ohio St. L.J. 1225 (2014); I. Ajunwa, K. Crawford & Jason Schultz, *Limitless Worker Surveillance*, 105 Calif. L. Rev. 735 (2017); D. K. Citron, F. Pasquale, *The Scored Society: Due Process For Automated Predictions*, 89 Wash. L. Rev. 1 (2014); K. Crawford & Jason Schultz, *Big Data And Due Process: Toward A Framework To Redress Predictive Privacy Harms*, 55 B.C. L. Rev. 93 (2014); Lillian Edwards & Michael Veale, *Slave To The Algorithm? Why A 'Right to an explanation' Is Probably Not The Remedy You Are Looking For*, 16 Duke L. & Tech. Rev. 18 (2017); G. Resta, V. Zeno Zencovich. (a cura di), *La protezione transnazionale dei dati personali*, Roma, 2016; G. Resta, *Diritti esclusivi e nuovi beni immateriali*, Torino, 2011; G. Pitruzzella, *Big Data, Competition and Privacy: A Look from the Antitrust Perspective*, in *Concorrenza e Mercato*, 2016, 15 ff.; F. Pizzetti, *Privacy e diritto europeo nella protezione dei dati*

A fundamental step concerned the possibility of collecting and processing masses of data in order to elaborate information capable of allowing intelligent operation by the machines themselves. Through this path it is possible to conceive an automatic decision-making process entrusted to machine which, based on the data entered, allows the identification of a specific solution to a given problem.²

These new applications have favored the introduction of artificial intelligence in many areas of production activities and public administration. In particular, the use of algorithms, sequences of well-defined computer instructions that are used to solve a series of problems or to perform a specific calculation, has increased.³

This type of processing uses a finite set of operations that are carried out by a computer on the basis of a set of data.⁴ The algorithm is therefore one of several applications of artificial intelligence and includes a wide range of tools, from optimization to research, which in turn influence a variety of operations. It is a type of decision that is made billions of times a year in various sectors such as credit, insolvency or diagnostics in the health sector. In the field of recruitment, for example, one or more subjects are selected from a larger pool of candidates to optimize some results.⁵

The challenge is that the result is not intelligible to the individuals involved at the time the decision is made. Which candidate will be more productive if they are hired? If we refer instead to the credit sector, which borrowers will be in default in the event of a possible stipulation of a loan agreement? This type of decisions create a risk, the best solution must be found through a decision-making process based on the information in possession. Such

personali, Torino, 2016; G. Pascuzzi, *Il diritto nell'era digitale. Tecnologie informatiche e regole privatistiche*, Il Mulino, Bologna, 2002. P. T. Kim, E. Hanson, *People Analytics And The Regulation Of Information Under The Fair Credit Reporting Act*, 61 *St. Louis U. L.J.* 17 (2016); Neil M. Richards & Jonathan H. King, *Big Data Ethics*, 49 *Wake Forest L. Rev.* 393 (2014); Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal Of Explainable Machines*, 87 *Fordham L. Rev.* 2018.

² B. Marr, *What Is The Difference Between Artificial Intelligence And Machine Learning?*, *Forbes* (6 Dec., 2016, 2:24 Am), <https://www.forbes.com/sites/bernardmarr/2016/12/06/what-is-the-difference-between-artificial-intelligence-and-machine-learning/3/#157219c52bfc>

³ S. Barocas, S. Hood & M. Ziewitz, *Governing Algorithms: A Provocation Piece*, *Governing Algorithms* ¶ 9 (Mar. 29, 2013), <http://governingalgorithms.org/resources/provocation-piece/> [https://perma.cc/M8RY-73YX].

⁴ I. Bogost, *The Cathedral of Computation*, *THE ATLANTIC* (Jan. 15, 2015), <http://www.theatlantic.com/technology/archive/2015/01/the-cathedral-of-computation/384300/> [https://perma.cc/AA6T-3FWV].

⁵ K. Gaebler, *The Future of Hiring: Human Resources, Without the Humans*, *ATLANTIC* (Feb. 3, 2012), <https://www.theatlantic.com/business/archive/2012/02/the-future-of-hiring-human-resources-without-the-humans/252518/>; Aki Ito, *Hiring in the Age of Big Data*, *BLOOMBERG PAULINE T. KIM, Auditing Algorithms for Discrimination*, 166 *U. Pa. L. Rev. Online* 189 (2017); P. T. Kim, *Data-Driven Discrimination At Work*, 58 *Wm. & Mary L. Rev.* 857 (2017); P. Kim, S. Scott, *Discrimination In Online Employment Recruiting*, 63 *St. Louis U. L.J.* (Forthcoming 2019); C. A. Sullivan, *Employing Ai* (Feb. 18, 2018) (Seton Hall Public Law Research Paper), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3125738 ; J. A. Kroll Et Al., *Accountable Algorithms*, 165 *U. Pa. L. Rev.* 633 (2017).

predictive processing does not exhaust the range of potential uses of algorithms within these decision-making processes, but they are among the most important uses of these software.⁶

This type of artificial intelligence is developed using machine learning methods. These are types of algorithms that build prediction functions on the basis of a series of data entered in order to make such predictions; once this function is consolidated, the algorithm receives a particular "input" (such as the characteristics of a candidate for a particular job position for example) and it predicts some results (such as his performance in specific areas).⁷

The usefulness of these machines is such that the areas in which it is deployed are different and its social implications are completely evident. Many of the decisions that are made are based on or influenced by the results of these algorithmic elaborations. For these reasons, jurists have begun to take an interest in the repercussions of these new technologies in the legal world.⁸

Their use is relevant to the law⁹, both because the rules of the legal system extend to the functioning of the algorithm when it is used in relation to particular human activities, and because law itself has begun to employ algorithms in the context of its application, such as in procurement contracts or in the calculation of the maintenance allowance. In particular, the main areas of interest so far have concerned the protection of privacy and the processing of personal data, issues relating to the correctness of these automated decision-making processes and above all in the field of anti-discrimination protection. In particular, the latter aspect has recently attracted the attention of scholars because the use of

⁶ A. Mantelero, *I Big Data nel quadro della disciplina europea della tutela dei dati personali*, in *Il Corriere giuridico - Speciali Digitali 2018*, 2018, 46 ff.; V. Morabito, *Big Data and Analytics. Strategic and Organizational Impacts*, Springer New York, 2015, 23 ss

⁷ C. Angelopoulos, et al. 'Study of fundamental rights limitations for online enforcement through self regulation, report IViR Institute for Information Law, University of Amsterdam' (2016) <https://www.ivir.nl/publicaties/download/1796>; Angwin J et al., 'Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks'. *ProPublica*, May 23, 2016 <https://www.ProPublica.org/article/machinebias-risk-assessments-in-criminal-sentencing>; Bambauer JR and Zarsky T, 'The algorithm game', 7 March 2018, *Notre Dame Law Review* 2018 <https://ssrn.com/abstract=3135949>.

⁸ R. Avraham, 'Discrimination and Insurance' in K. Lippert -Rasmussen (a cura di), *The Routledge handbook*

of the ethics of discrimination, 2017; K. Charles, J. Guryan, *Prejudice and wages: An empirical assessment of Becker's the Economics of Discrimination*. *J. Polit. Econ.* 116, 773–809 (2008); M. Turner et al., "All other things being equal: A paired testing study of mortgage lending institutions—final report" (Tech. Rep., US Department of Housing and Urban Development Office of Policy Development and Research, Washington, DC, 2002). M. Bertrand, S. Mullainathan, *Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination*. *Am. Econ. Rev.* 94, 991–1013 (2004); V. Zeno-Zencovich, *Dati, grandi dati, dati granulari e la nuova epistemologia del giurista*, in *Rivista di diritto dei media*, 2018, II, 5-6.

⁹ On this see the pioneering work of S. Rodotà, *Elaboratori elettronici e controllo sociale*, cit.; *Id. Tecnologie e diritti*, Bologna, 1995.

algorithms in the context of different decision-making processes, both in private and public sector, has led them to ask if there could be possible discrimination associated with the deployment of this type of artificial intelligence.¹⁰

These findings are based on two series of considerations that both originate in the discourses and elaborations that accompanied the use of these tools, according to the first, although algorithms make it possible to reduce the discretion linked to decision-making processes managed by human beings, it reckons is highly probable that algorithms reproduce the same inequalities that already exist. With respect to the second line of argument, the idea that algorithmic calculations are perceived as neutral, in the belief that technology is always apolitical, without little or any impact on society and that therefore the functioning of these automatic learning machines cannot be reconstructed or altered from the outside (black box), effectively making the sector immune to the regulation by law.¹¹

If the main issues relating to the use of algorithms are shared by the prevailing doctrine in Western legal systems being in line with what is stated by the decisions of the courts of the respective systems, as regards to the solutions, the opinions are quite different and not only due to the presence of regulations and different protections between the two sides of the Atlantic in terms of discrimination. In particular, two different approaches to anti-discrimination protection were highlighted in the context of algorithmic decision-making processes. The first prefers an *ex ante* protection that takes into account the processing of the algorithm through the technical procedures used to create it and provides a precise legal regulation with respect to the different possible concrete situations due to the functioning of the algorithmic decision-making process.¹²

The second proposes an *ex post* protection through the extension of the current legal approach to anti-discrimination to cases relating to the use of algorithms. This approach

¹⁰ S. Barocas, A. Selbst, Big data's disparate impact. *Calif. Law Rev.* 104, 671–732 (2016); S. Bornstein, Antidiscriminatory algorithms. *Ala. Law Rev.* 70, 519–572 (2018); J. Kleinberg, J. Ludwig, S. Mullainathan, C. Sunstein, Discrimination in the age of algorithms. *J. Legal Anal.* 10, 113–174 (2018); L. Sweeney, Discrimination in online ad delivery. *Queue* 11, 10–29 (2013).

¹¹ On this point see F. Pasquale, *The Black Box Society: The Secret Algorithms that control Money and Information*, 34-35; I. Bogost, *The Cathedral of Computation*, cit.; K. Fink, 'Opening the government's black boxes: freedom of information and algorithmic accountability' (2018) 21(10) *Information, Communication & Society* 1453.

¹² Su di un approccio all'antidiscriminazione che si fonda su alcuni principi generali si veda, JH Gerards, 'Discrimination grounds', in M. Bell and D. Schiek (a cura di), *Ius commune case books for a common law of Europe – Non-discrimination* (Hart 2007), 33-184. Federal Trade Commission. 'Big data: A tool for inclusion or exclusion? Understanding the issues' (Gennaio 2016) <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf> ; C. Dwork et al., 'Fairness through awareness' (Proceedings of the 3rd Innovations in Theoretical Computer Science Conference ACM, 2012) 214.

aims to apply the principles of transparency and accountability in order to limit or reduce the possibility of producing a discriminatory outcome.¹³

This contribution will therefore try to shed light on these two approaches and to propose an alternative path of protection. To do this, the text was divided into three parts. Part I provides the political-legal context that has been created around the growing use of algorithms in post-industrial society. Part II summarizes the existing state of the art on the issue, including the main regulations in the field, identifying the two main theories addressed in current science. Part III examines some decisions on discrimination to highlight the different types of algorithmic discrimination and to discuss some existing arguments about the limits of the current legal protection. It is important to focus on the risks of technology-assisted decision making, but this must not obscure the potential benefits: Algorithms can incorporate structural bias, but suppress human bias. Artificial intelligence can transform itself from an implacable enemy to a potential ally, the correct anti-discrimination approach can facilitate this technological transition.¹⁴

I.1 DISCRIMINATION AND ALGORITHMS: A CULTURAL CONTEXT AND TWO LEGAL APPROACHES

It is not possible to fully understand the importance of new technologies and the role of artificial intelligence within society if we do not pay attention to the cultural and political-legal context in which this transformation is taking place. The fourth industrial revolution entailed a greater dependence on electronic computers of industrial production and public administration, this trend was accompanied by a vision of these machines as tools capable of ensuring a more efficient management of many aspects of social life. Such a change in social organization cannot occur without introducing new knowledge, modifying the

¹³ T. Khaitan, *A theory of discrimination law*, Oxford 2015; Hacker P, 'Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law' (2018) 55 *Common Market Law Review*, Issue 4, p 1143–1185; M. Hardt, 'How big data is unfair. Understanding sources of unfairness in data driven decision making' (2014) <https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de>; R. Gellert, K. De Vries, P. De Hert, S. Gutwirth, 'A comparative analysis of anti-discrimination and data protection legislations', in *Discrimination and privacy in the information society* (pp. 61-89), Berlin, Heidelberg 2013.

¹⁴ On the risks and problems connected with the massive deployment of AI, see Ferguson AG, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, New York, 2017; A. Danna A, OH Gandy Jr., 'All that glitters is not gold: Digging beneath the surface of data mining' (2002) 40(4) *J Bus Ethics* 373; J. Burrell, 'How the machine 'thinks': understanding opacity in machine learning algorithms', *Big Data & Society* (2016) 3(1). 1-12; D. M. Boyd, K. Crawford, 'Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon' (2012) 15(5) *Information, Communication & Society* 662

interactions between individuals, influencing the power relations between the social categories to which they belong.¹⁵

In this sense, the introduction of artificial intelligence, like that of the Internet, has brought with it a series of concepts and modes of thinking that affect not only IT operators, but anyone who comes into contact with these new technologies. This discourse is aimed at assuming a hegemonic position with respect to the use of these tools, it is therefore functional to the control of the means of production by those social categories that have an interest in their management.¹⁶

We have already had occasion to mention how the main concept around which the discourse relating to artificial intelligence revolves is its supposed neutrality, the conception that technological tools have no impact within society, limiting themselves to simplifying the development of some human activities. Technology is therefore conceived, according to this idea, as something objective, immutable and inherently positive.¹⁷ This approach is connected to the very idea of progress and modernization of which artificial intelligence represents the substance. A further corollary is represented by the idea that limiting human discretion as far as possible within decision-making processes allows greater efficiency and guarantees a solution that is actually objective according to the goals for which it is used.¹⁸ This way of thinking is fundamental for the spread of these new technologies because it creates demand for them. A product whose impact, whose social costs are zeroed becomes

¹⁵ On the ties between means of knowledge production and hegemonic discourse, see A. Gramsci, per un approfondimento di questi temi nell'ambito delle nuove tecnologie, si veda O. H. Gandy Jr., *The Panopticon Sort: A Political Economy of Personal Information*, Nashville, 1993; M. Hildebrandt, *Criminal Law and Technology in a Data-Driven Society*, in M. Dubber, T. Hörnle, *The Oxford Handbook of Criminal Law*, Oxford 2014, 174-197; P. Nemitz, 'Constitutional democracy and technology in the age of artificial intelligence', *Phil. Trans. R. Soc. A* 376.2133 (2018): 20180089, <http://rsta.royalsocietypublishing.org/content/376/2133/20180089> ; C. O'Neil, *Weapons of math destruction: How big data increases inequality and threatens democracy* New York, 2016.

¹⁶ B.J., Koops, 'Should ICT regulation be technology-neutral?' in Koops, Bert-Jaap et al. (eds), *Starting Points for ICT Regulation - Deconstructing Prevalent Policy One-liners* (Information Technology and Law Series, Asser 2006). <https://ssrn.com/abstract=918746> ;

¹⁷ F. Puppe, *Systematic introduction to expert systems: Knowledge representations and problem-solving methods*, Heidelberg, Berlino 1993.

¹⁸ On this see E. Siegel, *Predictive analytics: The power to predict who will click, buy, lie, or die*, Hoboken 2013; D. Reisman, J. Schultz J, K. Crawford, M. Whittaker, 'Algorithmic impact assessments: A practical framework for public agency accountability', AI Now Institute 2018; Oswald M, 'Algorithm-assisted decision-making in the public sector: framing the Issues using administrative law rules governing discretionary power' (2018) 376 *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 20170359; [.com/abstract=3090360](https://royalsocietypublishing.org/doi/10.1098/rsta.20170359) accessed 26 September 2018. D. Greene, A.L. Hoffman, L. Stark, 'Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning', 2018 <http://dmgreene.net/wp-content/uploads/2018/09/Greene-Hoffman-Stark-Better-Nicer-Clearer-Fairer-HICSS-Final-Submission.pdf>

more attractive to buyers, who will be willing to change their habits and embrace these innovations based on the expectations that have been created.¹⁹

The concept of neutrality therefore ensures that artificial intelligence continues to be a requested product and that the main producers remain in control of the means of production of these goods. This idea works even if errors and malfunctions in the use of these technologies are contemplated, in these cases the intervention of computer scientists and lawyers is aimed at ensuring that the automated decision-making systems do not involve unwanted solutions or in any case not in line with the regulations put in place.²⁰

I.2 TRANSPARENCY AND RESPONSIBILITY BETWEEN LAW AND IT

It is interesting to note that in both these areas of interest, the principles mentioned are the same: transparency and accountability, even if in practice these references have different meanings and functions. At the base, however, there is a profound, undeclared and hidden conception that applies both to law, to information technology and to a large part of society in general; the idea is that the manufacturer of the machine will know exactly how it was built and what the different behaviors of the machine would be when a particular lever or a specific button is pushed.²¹

The machine in question has been designed for a certain purpose and will continue to do so as long as you continue to operate it. The relative presumption is that any other person in possession of all the information relating to this operation and given the perfect mechanism of the same, perhaps with the help of an expert, may be able to understand how the machine works. The reasoning behind is that seeing the internal components of a

¹⁹ G. Malgieri, 'Trade secrets v personal data: a possible solution for balancing rights' (2016) 6 International Data Privacy Law 2, 103; M. Jordan, 'Artificial intelligence -The revolution hasn't happened yet' (Aprile 2018) <https://medium.com/@mijordan3/artificial-intelligence-the-revolution-hasnt-happened-yet-5e1d5812e1e7>; WJ Frawley, G. Piatetsky-Shapiro, C.J. Matheus, 'Knowledge discovery in databases: An overview' (1992) 13(3) AI magazine 57.

²⁰ M. Ananny, K. Crawford, 'Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability' (2018) 20(3) New Media & Society 973; P. Schwartz, *Data Processing and Government Administration: The Failure of the American Legal Response to the Computer*, 43 Hastings L.J. 1321, 1343–74 (1992).

²¹ D. Levine, *Secrecy and Accountability: Trade Secrets in Our Public Infrastructure*, 59 FLA. L. REV. 135, 180 (2007); DJ. Weitzner et al., *Information Accountability*, 51 Comm. Acm, June 2008, 82; A. Haeberlen, P. Kuznetsov, P. Druschel, *Peer Review: Practical Accountability for Distributed Systems*, 41 Acm Sigops Operating Sys. Rev. 175, 175 (2007). M. Veale, *Logics and Practices of Transparency and Opacity in Real-World Applications of Public Sector Machine Learning*, 4 Workshop On Fairness Accountability & Transparency Machine Learning, 2 (2017), <https://arxiv.org/pdf/1706.09249.pdf>[<https://perma.cc/8X2L-RXZE>]

system leads to an understanding of the same and the consequences associated with its operation.²²

In the legal context, transparency is therefore the possibility of carrying out a verification of a particular algorithm, through access to information concerning it and to the patterns of its processing, to identify defects, intentional errors and perhaps find unwanted and possibly unintentional results such as discrimination. This approach therefore allows a strict control of all the components of the algorithm in order to exclude any possible result that may lead to discrimination or other outcomes not in line with the regulations in force in the field of application of the algorithm.²³

From this operation and its results derives the ability to hold the creator or operator of the system responsible. In this sense, if you allow access to all the information regarding a given algorithm and its functioning, only the identification of negative behaviors and results can lead to the identification of a person responsible for them in the juridical-political sense of the term. The combination of transparency and responsibility is now a classic of the neoliberal paradigm as a binomial that dominates legal-political relations in both the public and private spheres. Transparency must inform the relationships between the contractors, as well as the relationship between government and individual citizens, so that individuals have all the information necessary to be able to self-determine in any situation, both in the private context, deciding to complete or not a given transaction, and in the public one, verifying what their representatives are doing and participating in the political debate, or expressing their opinions to politicians and institutions.²⁴

The problem with this approach is connected to the fact that in the case of algorithms, the control cannot take place only on the machine (the software) that allows automatic data processing, but must be extended to all the steps of the algorithmic decision-making process that concern that particular operation. The same protection of fundamental rights - such as the dignity of the individuals involved in the decision-making process - has been elaborated by the regulations and decisions in the European context cannot be implemented without the knowledge of all the aspects and reasoning that have informed the algorithmic decision-making process. from its beginning to its end.²⁵

²² On this approach to Machine learning see D. R. Desai, J. A. Kroll, *Trust but Verify: A guide to Algorithms and the Law*, in 31 *Harv. J. L. Tech.* 1(2018) 9-11.

²³ P. T. Kim, *Data-Driven Discrimination at Work*, 58 *Wm. & Mary L. Rev.* 857 (2017); D. J. Weitzner et al., *Information Accountability*, 86.

²⁴ On the public sector, see M. Veale, *Logics and Practices of Transparency and Opacity in Real-World Applications of Public Sector Machine Learning*, 2 ff.; D. K. Citron, 'Technological due process' (2007) 85 *Wash.UL Rev.* 1249. On the private sector, D. R. Desai JA. Kroll, *Trust but Verify: A guide to Algorithms and the Law*, 32 ff.

²⁵ D.R. Desai JA. Kroll, *Trust but Verify: A guide to Algorithms and the Law*, *Ibidem*;

Such a condition can hardly occur with respect to algorithms, if we refer, for example, to those cases that have aroused greater perplexity in the field of anti-discrimination protection such as in the case of online advertisements or prices displayed in the e-commerce sector, it is not possible to map all possible variables intervened within the algorithmic decision-making process underlying that particular function.²⁶

The same open access accessibility of the algorithm source code often does not allow to prove that the disclosed software has been used in any discriminatory decision, unless the algorithmic decision-making process that led to that result cannot be perfectly reproduced. This procedure is really only possible under certain conditions. For these reasons, it is important to understand how the principles of transparency and responsibility are implemented, in order to anchor these legal concepts to a substantial application that allows effective protection of the underlying interests, on the other hand without a technical declination able to guarantee the collection of evidence that allows the supervision and verification that the algorithm is operating within agreed rules, transparency and responsibility become mere formal principles that do not allow effective anti-discrimination protection.²⁷

In the IT context, the two concepts under consideration are conceived with respect to the ability to elaborate the algorithm so that it produces a report of all the operations carried out during the decision-making process. Each single step provides a series of data that certify its correct performance or any deviations from it. A system of this type cannot, in itself, allow for the implementation of the principles of transparency and responsibility in the legal field, the IT meaning of the two concepts only allows for the provision of procedures aimed at finding evidence of what happens to the within the algorithmic

²⁶ On this see generally M. Kaminski, 'Binary governance: A two-part approach to accountable algorithms' (2018). 92 S. Calif. L. Rev. 2019. Per gli esempi, A. Datta et al., 'Discrimination in online advertising: A multidisciplinary inquiry' (Conference on Fairness, Accountability and Transparency 2018) 20; A. Datta, M.C. Tschantz, 'Automated experiments on ad privacy settings', 1 Proceedings on Privacy Enhancing Technologies (2015) 92; J. Dastin, 'Amazon scraps secret AI recruiting tool that showed bias against women', 10 October 2018. Reuters, <https://www.reuters.com/article/us-amazon-com-jobs-automation-in...-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>; D.J. Dalenberg, 'Preventing discrimination in the automated targeting of job advertisements' (2018) 34(3) Computer Law & Security Review 615.

²⁷ On this critically, J. Buolamwini, T. Gebru, 'Gender shades: Intersectional accuracy disparities in commercial gender classification' (Conference on Fairness, Accountability and Transparency 2018) <http://proceedings.mlr.press/v81/buolamwini18a.html>, 77. Christian Sandvig et al., Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Plat-forms, (22 maggio 2014), <http://www-personal.umich.edu/~csandvig/research/Auditing%20Algorithms%20--%20Sandvig%20--%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf>; F. Di Porto, La regolazione degli obblighi informativi. Le sfide delle scienze cognitive e dei big data, Napoli, 2017, p. 162 ff.

decision-making process, it is subsequently up to the legal plan to identify those responsible for any discriminatory decisions.²⁸

In other words, the IT declination of the liability principle is a necessary step to allow the application of its legal correspondent. Because it is only after verifiable evidence, relevant to the ongoing investigation, that one can have the possibility of holding public or private actors deploying these algorithmic decision-making processes responsible for their actions. This cannot mean that the use of such IT systems cannot be regulated, but it requires that it is clear what these algorithms are based on, how they are developed and above all what it is possible to do or not do when it comes to regulating or monitoring the use of these technologies.²⁹

The main problem, with respect to the extensive use of the principle of transparency (and accountability) as an *ex post* method to regulate automation, is that it does not address these issues, and therefore is unable to provide the type of legal answer necessary in these cases. On the contrary, the increasingly important use of automatic learning systems, which tend to separate the creation of algorithms and the rules that govern them from the design and implementation of humans, must push society to catalog the types of damage that can be connected to such use, to deny those outcomes that may result in discrimination, to prohibit inappropriate uses in general and to require that the software be built according to certain specifications that can be tested or/and controlled.³⁰

II.1 ANTI-DISCRIMINATION TESTED BY ALGORITHMIC DECISION-MAKING PROCESSES

Based on the foregoing, it is therefore necessary to focus on the questions associated with an *ex ante* approach to anti-discrimination protection with respect to the use of algorithms. The outlines of an *ex post* protection such as that based on the application of the neoliberal principles of transparency and accountability have been discussed in the previous paragraph, revealing a concept that, although it allows general protection for all cases in which a decision emerges at the end of a algorithmic decision-making process that can be

²⁸ D.R. Desai, J.A. Kroll, *Trust but Verify: A guide to Algorithms and the Law*, 9 ff.

²⁹ On machine learning technically, see E. Bayamloğlu, R. Leenes, 'The "rule of law" implications of data-driven decision-making: a techno-regulatory perspective', in *Law, Innovation and Technology* (2018); B. Bodo et al., 'Tackling the algorithmic control crisis—the technical, legal, and ethical challenges of research into algorithmic agents' (2017) 19 *Yale JL & Tech.* 133; J. Burrell, 'How the machine "thinks": understanding opacity in machine learning algorithms', *Big Data & Society* (2016) 3(1), 1-12; J. Kleinberg, J. Ludwig, S. Mullainathan, C. Sunstein, *Discrimination in the age of algorithms*, 3.

³⁰ On this program, see European Group on Ethics in Science and New Technologies, 'Statement on Artificial Intelligence, Robotics and "Autonomous" Systems', Marzo 2018, https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf

considered discriminatory, remains on a purely formal level due to the difficulty of mapping the functioning of the algorithm and finding effective evidence that can attribute the result in question to one or more responsible persons.³¹

The *ex ante* approach is instead based on a work of classification and analysis of these automatic systems to understand the type of damage they can cause, the solutions that can be produced and the permitted uses of these software. To do so, however, it is necessary to understand the development of these decision-making processes, their components, how these algorithmic systems are able to discriminate and how this can be detected on a strictly legal level.³²

In this sense, one can begin to investigate these issues by specifying how, although people often refer improperly to any process that elaborates data and produces a forecast as a "algorithm", it is important to note that there are actually two separate algorithmic processes at work in screening applications of the type we are considering: 1) The screening algorithm (or screener) simply takes the characteristics of an individual (such as a candidate of work) and returns a prediction of this individual's outcome. This prediction then informs a decision (hiring). 2) The training algorithm (or trainer) is what produces the screening algorithm.³³

Building this second algorithm involves (among other things) assembling past instances to be used as training data, defining the outcome to predict, and choosing candidate predictors to consider. The screening algorithm is only the mechanical result of applying the training algorithm to a training dataset. Thus, while the former can produce distorted decisions, the moment in which discriminatory treatment is generated is often the training phase involving the latter.³⁴

³¹ D. R. Desai JA. Kroll, *Trust but Verify: A guide to Algorithms and the Law.*, 11 ff.

³² S. Wachter, 'Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR', 34 *Computer Law & Security Review*, 3, 2018, 436-449; A.D.Selbst, S. Barocas, 'The intuitive appeal of explainable machines', *Fordham Law Review*, 86 (2018). R. Swedloff, 'Risk classification's Big Data (r)evolution' (2014) 21 *Connecticut Insurance Law Journal* 339; A. Moretti, *Algoritmi e diritti fondamentali della persona. Il contributo del regolamento (UE) 2016/679.*

³³ The functioning of an algorithmic decision-making process is well explained in J. Kleinberg, J. Ludwig, S. Mullainathan, C. Sunstein, *Algorithms as discrimination detectors*, in *Proceedings of the National Academy of Sciences of the United States of America*, 28 luglio 2020, <https://www.pnas.org/content/pnas/early/2020/07/27/1912790117.full.pdf>

³⁴ J. Kleinberg, J. Ludwig, S. Mullainathan, C. Sunstein, *ibidem*.

II.2 THE DISCRIMINATORY POTENTIAL OF ALGORITHMS: A PRACTICAL GUIDE

AI-driven decision making can lead to discrimination in several ways. In a landmark article, Barocas and Selbst distinguish five ways in which algorithmic decision making may lead to discrimination. The problems concern (I) how the target variable and class labels are defined; (II) the labeling of training data; (III) the collection of training data; (IV) the selection of indicators; (V) the proxies and finally (VI) the use of algorithms for discriminatory purposes on a voluntary basis.³⁵

1) the definition of target variable and class labels.

As previously seen, the screening algorithm works on the basis of the training algorithm, both look for correlations between groups of data, the second to provide data to the first, the first to process the required solution. For example, when a company develops a spam filter, the underlying algorithm is trained by inserting a series of e-mail messages that are labeled by programmers as "spam" and "not spam". The messages labeled are the training data.

The algorithm detects which characteristics of the messages are related to being labeled as spam, the set of these identified correlations is often called the "predictive model". Messages labeled as spam can often contain certain phrases ("magic pill to lose weight" or could be sent from certain IP addresses. The algorithm is trained to study the data entered to understand which characteristics can be taken into account to obtain the results which are referred to as the target variable.

While the target variable defines what the operators are looking for, the class labels divide the required results into mutually excludable categories, in the spam filter example above, people roughly agree on the class labels: which messages are spam or not. In other situations, it is less obvious what the target variables should be.

This point refers to one of the most controversial aspects relating to the use of algorithms as well as one of the major causes of discriminatory treatments through algorithmic decision-making processes. Anyone wishing to use an algorithm in order to predict a solution to a particular question needs to simplify the framework of the attributes they want to pay attention to in order for the machine to process them. In other words, the operators, even in the face of non-discriminatory and correctly sampled data, will have to carry out an interpretative activity of the same to allow their use by the algorithm.

³⁵ S. Barocas, A.D. Selbst, 'Big Data's disparate impact' (2016) 104 Calif Law Rev 671. Le stesse ipotesi sono riprese in F. Zuiderveen Borgesius, *Discrimination, artificial intelligence, and algorithmic decision-making*, (Consiglio d'Europa) Strasburgo 2018, 10 ff.

Take, for example, the case of a company that is looking for a candidate / or to fill a position within its organization. The idea is therefore to hire a model subject and to do so it will be necessary to indicate his qualities: suppose that less well-off individuals rarely live in the city center and have to go to their work further away than others. Therefore, individuals belonging to these social categories will be more exposed to possible delays to work than others due to traffic or problems with public transportation.³⁶

The company may choose "rarely be late" as a class label for assessing whether an employee is "good". In this case, those with lower incomes who usually live further away from their work would be at a disadvantage, even if they outperform other employees in other respects.³⁷

2) labeling of training data and (3) collection of training data

These two operations refer to two rather similar types of cases, the discrimination would derive in these hypotheses from the steps related to the processing of the algorithm training data.³⁸

According to point (2), the data are discriminatory not because they introduce new elements or correlations that lead to a similar result, but because they reproduce a discriminatory behavior already present in society. The classic example in this case is that of an English healthcare company that, faced with a significant number of candidates for some positions in its structure, decides to use a program to speed up the selection procedures. The program is trained using data from previous hires.³⁹

The result was that the program discriminated against certain social categories such as women and the non-indigenous population. In this case, no discriminating element was introduced, simply in the previous procedures someone had systematically discriminated against those particular social categories, consequently the data used for the training of the algorithm were already altered from the start.⁴⁰

Point (3), on the other hand, refers to the moment preceding the one just discussed, to the passage relating to the sampling of data to be used in the training of the algorithm.⁴¹

³⁶ F. Zuiderveen Borgesius, *Discrimination, artificial intelligence and algorithmic decision-making*, 11.

³⁷ *Ibidem*.

³⁸ S. Barocas, A.D. Selbst, 'Big Data's disparate impact', 680-1; F. Zuiderveen Borgesius, *Discrimination, artificial intelligence and algorithmic decision-making*, 11.

³⁹ This example can be found in S. Lowry, G. Macpherson, 'A blot on the profession' 296(6623) *Br Med J* (1988) 657.

⁴⁰ F. Zuiderveen Borgesius, *Discrimination, artificial intelligence and algorithmic decision-making*, 12.

⁴¹ . Barocas, A.D. Selbst, 'Big Data's disparate impact', 685; D. Robinson, L. Koepke. 'Stuck in a pattern' (2016) https://www.upturn.org/static/reports/2016/stuck-in-a-pattern/files/Upturn_-_Stuck_In_a_Pattern_v.1.01.pdf

A useful example for this case concerns the Street Bump app, an app that uses users' GPS information to notify public administrations which roads need to be serviced. In this case, the sampling was distorted by the fact that the reports occurred on the streets where the number of smartphones was greater, therefore the streets of the more affluent neighborhoods received more assistance than those of the less affluent ones, where the percentage of new generation phones was inferior.⁴²

4) the selection of indicators

This operation takes up some aspects of point (II) on the creation of class labels, classifications through which the data are processed to allow the machine to elaborate them, sometimes this procedure can be too expensive or too long, for this reason particular details are chosen. characteristics that act as indicators for the algorithm, for example a particular type of education or some specialized course. However, these choices can result in discriminatory behaviors towards some social categories, if the algorithm is trained to prefer candidates from particularly expensive private universities, the less well-to-do individuals and those belonging to minorities who are statistically less present within the group will automatically be excluded from that kind of educational institutions.⁴³

5) Proxies

This point concerns the so-called proxies, those data that have been sampled to be used as training for the algorithm, but which are indirectly linked to particular social categories and can therefore lead to discriminatory behavior on the part of the machine. The data are, in this case, perfectly neutral and do not involve an upstream choice by the operators as in the previous point, these are data that can be linked to certain situations and therefore are called proxies, because the discriminatory result is not determined by effect of the data entered in the machine, but by the characteristics related to them.⁴⁴

An example can clarify this situation, imagine a bank that wants to predict the degree of solvency of its borrowers. Among the data that are entered into the algorithm, there are the postal codes, the machine will therefore tend to associate solvency percentages with respect to those who come from particular areas included in the post codes entered. If those areas

⁴² Federal Trade Commission. 'Big data: A tool for inclusion or exclusion? Understanding the issues' cit, 27. Per un approfondimento su temi delle politiche di sorveglianza al tempo dell'intelligenza artificiale, si veda A.G. Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, New York, 2017.

⁴³ S. Barocas, A.D. Selbst, 'Big Data's disparate impact', 689.

⁴⁴ S. Barocas, A.D. Selbst, 'Big Data's disparate impact', 692.

are mainly inhabited by residents with lower income or of non-indigenous origin, a discriminatory result could occur.⁴⁵

6) voluntary algorithmic discrimination

Finally, the last case contemplates the most direct hypothesis in which the operators have a positive discriminatory intent at the basis of the use of the algorithm, in these cases they usually use proxies or incorrect data samples to allow discriminatory outcomes. that they want to obtain.⁴⁶

This whole frame is rather complicated, the type of algorithm functioning and the complexity of the decision-making process that involves it greatly complicates the legal regulation of the use of this type of software. In this sense, it appears difficult to follow an *ex post* approach on the issue because the use of general principles such as those set out above would not allow effective protection in different types of cases just treated. The principles of transparency and accountability alone would not make it possible to deal with discrimination implemented through proxies (point 5) because access to the algorithm would only reveal that no step in the decision-making process has been altered, even if the correlation with data considered neutral and social categories discriminated against demonstrates the opposite.⁴⁷

Equally, the principle of responsibility would be difficult to pursue since many of the cases listed would not allow to easily trace a separate person responsible in the case of a malicious behavior of the same. To deepen these arguments, it is good to refer to an analysis of the regulatory framework on the subject and a study of the decisions made on the subject, to see how substantial anti-discrimination protection develops and how it relates to the two approaches that have been highlighted.⁴⁸

⁴⁵ F. Zuiderveen Borgesius, *Discrimination, artificial intelligence and algorithmic decision-making*, 13.

⁴⁶ On voluntary algorithmic discrimination, see P. T. Kim, 'Data-driven discrimination at work', 857 J. Bryson, 'Three very different sources of bias in AI, and how to fix them', *Adventures in NI*, 13 July 2017, <https://joanna-bryson.blogspot.com/2017/07/three-very-different-sources-of-bias-in.html> ; B. Friedman, H. Nissenbaum, 'Bias in computer systems' (1996) 14(3) *ACM Transactions on Information Systems (TOIS)* 330.

⁴⁷ A. Kroll et al., 'Accountable algorithms' 165 *University of Pennsylvania Law Review* (2016), 633-705.

⁴⁸ On this see D.R. Desai, J.A. Kroll, *Trust but Verify: A guide to Algorithms and the Law*, 9 ff.

II.3 MULTILEVEL ANTI-DISCRIMINATION PROTECTION

The legislation on anti-discrimination, in the various Western legal systems, uses different tools, in particular in the legal systems that will be taken into consideration: the US and the European one. In both of these contexts there is talk of multilevel protections of subjective legal situations that are integrated in the protection of certain interests such as those relating to the substantial equality of individuals. Specifically, at the European level, two different sources of law can be distinguished, on the one hand there is the European Convention on Human Rights and on the other the Charter of Fundamental Rights, the directive on anti-discrimination and the GDPR regulation.⁴⁹

The convention governs the anti-discrimination protection in art. 14 of the Convention, which generally protects the enjoyment of rights without discrimination based on sex, race, color, language, religion, political or other opinions, national or social origin, belonging to a minority nationality, wealth, birth or any other condition. This provision is further strengthened by the provision of Protocol 12 to the Convention which reaffirms an all-encompassing protection in the anti-discrimination field, in addition to the list contained in art. 14.⁵⁰

The decisions of the Edu Court subsequently took on the task of defining the two main types of discrimination, the direct one that finds the difference in treatment in similar or similar situations based on identifiable characteristics. On the contrary, indirect discrimination arises through an action that is apparently neutral, but results in a discriminatory act against a particular social category. A study of the Court's judgments on the subject has shown that the regulation to protect indirect discrimination does not result in a simple application of a rule as in the case of direct discrimination.⁵¹

⁴⁹ ECRI Statute Resolution 2002: Council of Europe Committee of Ministers, Resolution Res(2002)8 on the statute of the European Commission against Racism and Intolerance, https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016805e255a ; European Group on Ethics in Science and New Technologies, 'Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems', March 2018, https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf ; Hacker P, 'Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law' (2018) 55 Common Market Law Review, Issue 4, pp. 1143–1185; N. Helberger , F. Zuiderveen Borgesius, A. Reyna, 'The perfect match? A closer look at the relationship between EU consumer law and data protection law', Common Market Law Review, 2017-54, 1427-1466.

⁵⁰ Several treaties and conventions provide anti-discrimination protection, the UN Universal Declaration of Human Rights (art. 7), the European Convention on Human Rights (art. 14) and the Protocol 12 to the Convention which has not yet been ratified by all members, the Fundamental Charter of Rights of the European Union (art. 21) which will be dealt with in this paragraph and the International Covenant of Civil and Political Rights (art. 26).

⁵¹ On this point, in the ECtHR decisions, see *Biao v. Denmark* (Grand Chamber), No. 38590/10, 24, 2016, par. 89; *D.H. and Others v. Czech Republic* (Grand Chamber), No. 57325/00, 2007, par. 187-188.

In the first case, reference is usually made to the concept of standard, to a more general concept which is more difficult to apply to the single substantial case. This is because it is not always easy to detect a case of indirect discrimination, even more so to be able to prove that a regulation, an act or in the case in question, an algorithmic decision-making process affects a specific social group disproportionately.⁵²

The picture is further complicated by the fact that not only is the burden of proof on the plaintiff is quite substantial, but it is also possible for the Court to balance the alleged discrimination against the existence of a policy or decision, that can be reasonably and objectively justified it.⁵³

A very similar discipline is that provided for by the EU law, the Fundamental Charter of Nice provides for a discipline that incorporates that of the ECHR, in art. 21, 1 reads "any form of discrimination based, in particular, on sex, race, skin color or ethnic or social origin, genetic characteristics, language, religion or personal beliefs, political opinions is prohibited or of any other nature, membership of a national minority, heritage, birth, disability, age or sexual orientation."⁵⁴

This prescription was subsequently positivized also in a series of directives starting from n. 43 of 2000. The established regime is essentially the mirror image of that established by the ECtHR; also in this case there is a direct discrimination of simple interpretation and an indirect one, which is distinguished by a more discretionary discipline.⁵⁵

The legislation provides, as previously seen with respect to the decisions of the Edu Court, the possibility for the defendant to free himself by demonstrating that the alleged discriminatory conduct took place in order to obtain a legitimate result and that the tools used to achieving this goal are necessary and proportionate; the principle of proportionality is also used in EU legislation as a tool to balance anti-discrimination protection with other interests.⁵⁶

⁵² On the concept of rules and standards see D. Kennedy, *Form and Substance in Private Law Adjudication*, 89 *Harvard Law Review* 8 (1976) 1685-1778; P. Schlag, *Rules and Standards*, 33 *UCLA L. Rev.* 379 (1985); C. Sunstein, 'Problems with rules' (1995) 83(4) *California Law Review* 953.

⁵³ On the balance of rights made by the ECtHR see *Biao v. Denmark* (Grand Chamber), No. 38590/10, 24, 2016, par. 91-2.

⁵⁴ Charter of Fundamental Rights of the European Union, art. 21.

⁵⁵ This distinction can be found in the European legislation on anti-discrimination, directive 2000/43 / EC which implements equal treatment between individuals regardless of race and ethnic origin, directive 2000/78 / EC which establishes a discipline on equality of treatment in the field of employment, 2004/113 / EC which regulates equal treatment in the context of access to goods and services and 2006/54 / EC which regulates equal opportunities and equal treatment between men and women in matters of employment and occupation.

⁵⁶ This point is the same in *F. Zuiderveen Borgesius*, *Discrimination, artificial intelligence and algorithmic decision-making*, 19 ff.

In such a condition it is more difficult to detect discrimination in the use of an algorithm. In this sector, the majority of discrimination is indirect, therefore it falls within the scope of greater discretion, it will therefore be more difficult to demonstrate the effectiveness of the discriminatory conduct and the responsibility of the defendants. Furthermore, the safeguard clause could allow them to free themselves by demonstrating that the algorithm, as it was developed, is configured as a necessary and proportionate tool to achieve that legitimate goal.⁵⁷

II.4 FIGHTING ALGORITHMIC DISCRIMINATION THROUGH THE LEGISLATION ON THE PROCESSING OF PERSONAL DATA

In the context of artificial intelligence, a regulation that can prove particularly useful in the context of the EU and the Council of Europe is that relating to the processing of personal data. In this field, the protection provided is less linked to general principles and more centered on individual rules that clearly govern the protection afforded by the right to privacy. The legislation promoted in the European context on the issue defines some principles such as transparency, integrity, responsibility and confidentiality, but bases its discipline on prescriptions that govern each step of the management of personal data: from collection to storage, from purpose to control the treatment.⁵⁸

This legislation can become extremely useful in the context of anti-discrimination protection precisely because it allows the various steps of the algorithmic decision-making process to be followed, on the one hand allowing operators to develop the software and train it according to a series of clear rules that can be easily applied and on the other hand, by providing the supervisory authorities with a control of the procedures that have been used. Both the EU Charter of Fundamental Rights and the Data Protection Convention 108 of the Council of Europe oblige member countries to have independent institutions

⁵⁷ On this see art.2 (2) b 2000/43/EC: "indirect discrimination shall be taken to occur where an apparently neutral provision, criterion or practice would put persons of a racial or ethnic origin at a particular disadvantage compared with other persons, unless that provision, criterion or practice is objectively justified by a legitimate aim and the means of achieving that aim are appropriate and necessary." which is restated in the ECtHR decision *Biao v. Denmark* (Grand Chamber), No. 38590/10, 24 maggio 2016, par. 91-2: "A general policy or measure that has disproportionately prejudicial effects on a particular group may be considered discriminatory even where it is not specifically aimed at that group and there is no discriminatory intent. This is only the case, however, if such policy or measure has no "objective and reasonable" justification"

⁵⁸ On the EU legislation on data protection, with particular reference to AI and Big Data, see European Data Protection Supervisor, 'Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy', March 2014, https://edps.europa.eu/sites/edp/files/publication/14-03-26_competition_law_big_data_en.pdf ; G. Buttarelli, *Towards a New Digital Ethics: Data, Dignity and Technology*, Speech before the Institute of International and European Affairs, Dublino, 2015, 1-4; F. Pizzetti (acura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, cit.

that guarantee the processing of personal data. These authorities must also be provided with all the supervisory and investigative powers that are necessary for the effective exercise of their mandate.⁵⁹

The regulation adopted by the EU in this field, the General Data Protection Regulation offers a series of provisions that go precisely in this direction, establishing that the supervisory authorities of the member countries can request information and access to the data processing system from those responsible of the processing, they can access the physical places where data management takes place and conduct audits on a particular use of artificial intelligence.⁶⁰ In terms of algorithmic decision-making processes, the GDPR has prepared a series of rules starting with art. 22, which deals with those decisions taken with the sole aid of algorithms and for which: "The data subject shall have the right not to be subject to a decision based solely on automated processing including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her "(1).⁶¹

In other words, anyone who submits to any decision, which contemplates the processing of personal data and has direct legal effects (or similar), has the right that this does not take place with the sole aid of an automatic data processing procedure.

The denial of welfare or pension benefits, or a ruling by a judicial body or a public administration that bases its decision exclusively on an algorithmic process gives rise to a right to appeal against it. It is interesting to note how this protection also applies in cases of *profiling*.⁶²

⁵⁹ H. Collins, T. Khaitan, 'Indirect discrimination law: Controversies and critical questions' in Collins, H. and T. Khaitan (eds), *Foundations of Indirect Discrimination Law*, Oxford 2018.

⁶⁰On the relation between anti-discrimination law and GDPR, see W. Schreurs, M. Hildebrandt, E Kindt, M. Vanfleteren, 'Cogitas, ergo sum. The role of data protection law and non-discrimination law in group profiling in the private sector' in *Profiling the European citizen Heidelberg*, Berlino, 2008; P. Hacker, 'Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law' 1143 ff.; F. Zuiderveen Borgesius, *Discrimination, artificial intelligence and algorithmic decision-making*, 21.

⁶¹ Art. 21 GDPR (UE) 2016/679. On this see I. Mendoza, L. A. Bygrave 'The right not to be subject to automated decisions based on profiling', University of Oslo Research paper no. 2017-20, (2017). Sia sul tema in generale, che con specifico riferimento alla discriminazione in tema di online pricing, si veda F. Zuiderveen Borgesius, J. Poort, 'Online price discrimination and EU data privacy law', *Journal of Consumer Policy*, 2017, p. 1-20. <https://ssrn.com/abstract=3009188>

⁶² Art.4(4) GDPR defines profiling as "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements;" On this see, S. Wachter 'Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR', *Computer Law & Security Review*, Volume 34, Issue 3, June 2018, 436-449; O. De Schutter, J. Ringelheim, 'Ethnic profiling: A rising challenge for European human rights law', 358 ff.; B. E. Harcourt, *Against prediction: Profiling, policing, and punishing in an actuarial age* Chicago, 2008.

Art. 22 establishes exceptions to the rule set out in par. 1, the prohibition does not apply if the automated decision (i) is based on the explicit consent of the interested party; (ii) it is necessary for a contract between the natural person and the data controller; or (iii) this is authorized by law. However, the first two cases trigger the application of another discipline for which the data controller is required to implement adequate measures to safeguard the rights, freedoms and legitimate interests of the data subject and at least the right to obtain intervention. by the data controller and to contest the decision (Article 22 par.3 GDPR).⁶³

This prediction becomes extremely interesting because it allows the possibility of resorting to decisions expressed exclusively through algorithmic procedures that may have compressed the rights of the subjects involved by them. Some authors have worked on the idea of an algorithmic due process that protects certain fundamental freedoms and rights of individuals with regard to such decision-making process.⁶⁴

Alongside this specific provision, the GDPR provides for further more general protection mechanisms such as the "right to lodge a complaint with a supervisory authority" (Article 77 of the GDPR), which individuals can exercise in their place of residence, workplace or place of the alleged violation. The "right to an effective judicial remedy against a supervisory authority" (Article 78 GDPR), which can be asserted against any supervisory authority that "does not handle a complaint or does not inform the data subject within three months of the progress o on the outcome of the complaint presented. "The" right to an effective judicial remedy against a data controller or processor "(Article 79 GDPR) which, in the case of multiple data processors and / or data processors, specifies that each party of the violation is responsible (see Par. 4).⁶⁵

⁶³ F. Zuiderveen Borgesius, *Discrimination, artificial intelligence and algorithmic decision-making*, 22. Sulla possibilità di regolamentare questo ambito, P. De Hert, S. Gutwirth, 'Regulating profiling in a democratic constitutional state' in M. Hildebrandt M, S. Gutwirth (a cura di), *Profiling the European Citizen*, Heidelberg, Berlino, 2008.

⁶⁴ On the concept of *algorithmic due process* see M. Kaminski, 'Binary governance: A two-part approach to accountable algorithms' (2018). 92 S. Calif. L. Rev.; Sullo stesso punto, si parla più in generale di *technological due process*, K. Citron, 'Technological due process' (2007) 85 Wash.UL Rev. 1249.

⁶⁵ On the potential of the GDPR on contrasting algorithmic discrimination, see B. Goodman, 'Discrimination, data sanitisation and auditing in the European Union's General Data Protection Regulation' (2016) 2 Eur.Data Prot.L.Rev. 493; S. Wachter, 'Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR', 436 ff.; S. Wachter, B. Mittelstadt, 'A right to reasonable inferences: Re-thinking data protection law in the age of Big Data and AI', 1 ff. Nell'ambito del Consiglio d'Europa si veda, A. Mantelero, 'Artificial Intelligence and data protection: Challenges and possible remedies', draft report for the Council of Europe's Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, 17 September 2018, T-PD(2018)09Rev <https://rm.coe.int/report-on-artificial-intelligence-artificial-intelligence-and-data-pro/16808d78c9> ;

The "right to compensation and liability" (Article 82 GDPR), which creates the obligation for both data processors and data controllers to compensate "any person who has suffered material or immaterial damage as a result of [their] violation of this regulation".⁶⁶

From these rules it can be inferred that a greater effort has been made in an attempt to regulate the use of algorithms in the field of personal data protection and that a more effective anti-discrimination protection is possible according to this legislation. Also in the context of the general principles to which it has already referred, the need was felt to decline the concept of transparency specifically for cases relating to algorithmic decision-making processes, Articles 13 paragraph 2 lett. F and Article 14 paragraph 2 letter F regulate the obligations of the data controller towards those affected by an automated decision-making process, including profiling and, among which, the communication of all significant information on the logic used, as well as the importance and expected consequences of such processing towards the interested parties themselves.⁶⁷

There still remains a topic of discussion whether these obligations give rise to a right of explanation for the subjects involved in the decision by virtue of which it is possible to request clarification with respect to any result identified with the aid of artificial intelligence. This would oblige operators to implement those steps required by IT transparency and would allow the parties involved to have technical information through which to try to understand the steps of the decision-making process, on this point the Data Protection Convention 108 of the Board of Directors is clearer. Europe of 2018 which in art. 9 provides for a right of this type: "Every individual shall have a right (...) to obtain, on request, knowledge of the reasoning underlying data processing where the results of such processing are applied to him or her".⁶⁸

The legislation on the protection of personal data offers a more articulated guarantee against decisions involving artificial intelligence, the treatment tends to be based not only on general principles, which are more difficult to apply to the various technical cases that may occur at " within an algorithmic decision-making process, but also on the work carried out by the guarantor authorities who can develop sector regulations or codes of conduct in

⁶⁶ artt. 77, 78, 79, 82 GDPR (UE) 2016/679.

⁶⁷ See F. Zuiderveen Borgesius, *Discrimination, artificial intelligence and algorithmic decision-making*, 23.

⁶⁸ Art. 9 Council of Europe Data protection Convention 108 (2018). Council of Europe Police and Personal Data Guide 2018, Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, 'Practical guide on the use of personal data in the police sector', T-PD(2018)01, 15 febbraio 2018.

specific areas that allow greater versatility of the rules prepared, which are thus able to relate substantially with the IT characteristics covered by the algorithms.⁶⁹

However, these positive aspects are mitigated by a series of objections that the doctrine on the subject has put in place with respect to anti-discrimination protection through the protection of personal data. Firstly, the discipline in question can be used only in the event that the algorithmic decision-making process manages personal data.

In the event that data classified as such are not involved, it is not possible to appeal to the data protection authority or rely on the legislation. For example, many of the data we referred to in the cases set out in the previous paragraph: the ZIP code, the course of study, but above all because the training algorithm that creates the predictive model does not use personal data, it elaborates a decision-making process on the general data base: the majority of residents in a particular zip code do not diligently fulfill loan contracts.⁷⁰

The same special protection relating to sensitive data can in some cases be an obstacle to the anti-discrimination one. The legislation on the processing of these data means that they cannot be held on the part of the operators who develop the algorithm, on the other hand, the discriminatory regulation often requires that the data be in the possession of the organization that processes the algorithmic decision-making process.⁷¹

Another remark concerns the so-called right of explanation, it is often difficult to explain the logic behind a decision, when an artificial intelligence system, analyzing large amounts of data, arrives at that decision. In some cases, it is not clear how much an explanation could help the stakeholders, especially to the extent that it places the burden of understanding the decision itself and its adequacy on them.⁷²

⁶⁹ See S. Bornstein, *Antidiscriminatory algorithms*, 522 ff.; J. Kleinberg, J. Ludwig, S. Mullainathan, C. Sunstein, *Discrimination in the age of algorithms*, 5.

⁷⁰ On the limits of the data protection regulations; S. Wachter, B. Mittelstadt, 'A right to reasonable inferences: Re-thinking data protection law in the age of Big Data and AI', 81 ss; M. Ananny M K. Crawford, 'Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability' (2018) 20(3) *New Media & Society* 973.

⁷¹ On this point see G. Malgieri, G. Comandé, 'Why a right to legibility of automated decision-making exists in the general data protection regulation' (2017) *International Data Privacy Law*; J. Ringelheim O. De Schutter, 'The processing of racial and ethnic data in antidiscrimination policies: Reconciling the promotion of equality with privacy rights' (2009) *Brussels*.

⁷² In this field see, L. Edwards, M. Veale, 'Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for' (2017) 16 *Duke L. & Tech. Rev.* 18. Id., 'Enslaving the algorithm: from a "right to an explanation" to a "right to better decisions"?' *IEEE Security & Privacy*, 2018, 16(3), 46-54; M. Kaminski, 'The right to explanation, explained' (2018). <https://osf.io/preprints/lawarxiv/rgeus/>; G. Malgieri G, 'Right to explanation and algorithm legibility in the EU Member States legislations', 17 August 2018, <https://ssrn.com/abstract=3233611>; A.D. Selbst, J Powles, 'Meaningful information and the right to explanation' (2017) 7(4) *International Data Privacy Law* 233; S. Wachter, B. Mittelstadt, L. Floridi, 'Why a right to explanation of automated decision-making does not exist in the general data protection regulation', *International Data Privacy Law* 2017-2.

II.5 OTHER EXAMPLES OF REGULATIONS AT THE EUROPEAN LEVEL

Anti-discrimination protection in the field of artificial intelligence can also make use of other regulatory areas that can mitigate the problems related to discrimination. For example, consumer law could be invoked to protect individuals from certain types of manipulative advertising managed through the use of algorithms. The discriminatory conduct of a company causes more problems when it operates in a monopoly position, competition law could therefore intervene in this context. Administrative law becomes fundamental in regulating the use of algorithmic decision-making processes within the public administration.⁷³

The use of these fields of law in the field of artificial intelligence is largely unexplored. An exhaustive discussion of these sectors is outside the scope of this text.⁷⁴

Several regulatory measures are currently being studied that could be relevant to discrimination in the field of artificial intelligence. In 2018, the European Commission published a communication on this topic and set up a group of experts, which has the task of presenting ethical guidelines on the basis of which to regulate this field. Furthermore, in 2017 the Commission proposed a regulation for the protection of internet privacy (ePrivacy), which could be relevant for artificial intelligence in general and for machine learning software in particular as it would limit the collection of certain types of data sensitive in the network.⁷⁵

The Steering Committee of the Council of Europe on the Media and Information Society has established a committee of experts on artificial intelligence: the committee of experts on the human rights dimensions of automated data processing and on the different forms

⁷³ On anti-discrimination protection through competition law, A. Ezrachi, M. E. Stucke, *Virtual Competition*, Cambridge, 2016; I. Graef, *EU Competition Law, Data Protection and online platforms: Data as essential facility* (Kluwer Law International 2016); Id., 'Algorithms and fairness: What role for competition law in targeting price discrimination towards end consumers?' (2017) <https://ssrn.com/abstract=3090360>. In the context of Italian administrative law, the recent ruling of the Council of State on the subject, no. 2270, April 8, 2019, according to which the robotic algorithmic decision-making process "requires the judge to evaluate the correctness of the automated process in all its components"; R. Ferrara, *The administrative judge and the algorithms* Extemporaneous Notes on the sidelines of a recent jurisprudential debate, *Administrative Law*, fasc. 4, 1 December 2019, 773 ff.

⁷⁴ On this see, C. Buzzacchi, *La politica europea per i big data e la logica del single market: prospettive di maggiore concorrenza?* in *Concorrenza e mercato*, 2016, 153; G. Ghidini, *Profili evolutivi del diritto industriale: innovazione, concorrenza, benessere dei consumatori, accesso alle informazioni*, Milano, 2015; I. Govaere, H. Ullrich. (a cura di), *Intellectual Property, Market Power and the Public Interest* Bruxelles, 2008.

⁷⁵ European Group on Ethics in Science and New Technologies, 'Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems', March 2018, https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf

of artificial intelligence. The panel of experts will conduct studies and provide guidance for possible future standard setting.⁷⁶

II.6 ANTI-DISCRIMINATION PROTECTION IN THE UNITED STATES OF AMERICA IN RELATION TO THE USE OF ARTIFICIAL INTELLIGENCE

Anti-discrimination in the American legal world is a particularly developed tool capable of adapting over the years to the various fields of application. At the federal constitutional level, the focus of protection revolves around the equal protection clause contained within the XIV amendment, which guarantees equal protection to all individuals by the law. This requirement was aimed at protecting the rights of individuals vis-à-vis the federated states, at the federal level, equal protection by the fifth amendment to the constitution is guaranteed by the jurisprudence of the Supreme Court. This constitutional requirement allowed the introduction of a series of regulatory instruments at both national and state level to protect the social categories most exposed to discriminatory treatment. In particular, it is good to remember the Civil Right Act of 1964 which recognizes general protection against all types of discrimination, with particular reference to the subject of this paper, Title VII provides for a compensation guarantee against all forms of indirect discrimination (disparate treatment).⁷⁷

The US legal system, despite the vanguard position of the national industry in the field of artificial intelligence, does not have a general legislation on the subject. The regulatory production mainly concerned a couple of executive orders from the Obama and Trump administrations that mainly outlined industrial plans in the field of artificial intelligence, some regulations of certain sectors such as that relating to driverless cars and a series of proposed bills presented in front of Congress and waiting to be voted on.⁷⁸

Of the latter, the most prominent examples were the Senate and House Bills for Algorithmic Accountability Act (S. 1108, HR2231) which were introduced to Congress on April 10, 2019, likely in response to recently publicized reports. on the risks of distorted results produced by the use of intelligence artificial.⁷⁹ The bill aims to require private actors

⁷⁶ Council of Europe MSI-AUT 2018, Council of Europe Committee of experts on Human Rights Dimensions of automated data processing and different forms of artificial intelligence, [https://www.coe.int/en/web/freedom-expression/msi-aut#{"32639232":\[0\]}](https://www.coe.int/en/web/freedom-expression/msi-aut#{) .

⁷⁷ On this point in relation to algorithmic decision-making processes see, S. Bornstein, Antidiscriminatory algorithms, cit. 525 ff.

⁷⁸ Executive Office of the President National Science and Technology Council, Committee on Technology, "Preparing for the Future of Artificial Intelligence", October 2016. This first Obama administration document was followed by an executive order from the Trump administration, Exec. Order No. 13,859, 3 C.F.R. 396 (2019).

⁷⁹ Algorithmic Accountability Act of 2019, S. 1108, H.R. 2231, 116th Cong. (2019). On this, S. Revanur, In a Historic Step Toward Safer AI, Democratic Lawmakers Propose Algorithmic Accountability Act,

and public entities to conduct impact assessments on their automated decision-making systems considered "high-risk" in order to verify how the algorithmic decision-making process relates to the general principles of accuracy, fairness, privacy and security.⁸⁰ The proposal also provides that such assessments should be conducted by external third parties, including independent auditors and technology experts.⁸¹

The bill also defines algorithmic decision making in a broad sense as "any computational procedure, including that derived from the learning machine, statistics or other data processing or artificial intelligence techniques, which make a decision or facilitate decision making. human beings, which have an impact on consumers."⁸²

The Algorithmic Accountability Act would operate in a federal setting and would not contradict any state law, which means that private operators would have to remain vigilant to keep up with any development of state law on similar topics. The bill also does not provide for the possibility of a judicial remedy by individuals, the protection in this case is implemented by the Federal Trade Commission, pursuant to sect. 5 of the Federal Trade Commission Act on competition and unfair practices, or by the attorney general of the competent state through a civil appeal.

This legislation requires organizations employing artificial intelligence to conduct impact assessments similar to those proposed by the federal bill on algorithmic decision making and information systems. This verification implies an evaluation of the system development process, including the component relating to its design and training data, must contain a detailed description of the best practices used to minimize risks and a cost-benefit analysis.⁸³

The proposal also requires organizations employing automated systems to work with independent third parties, record any bias or threat to the security of consumers' personal data that are identified through impact assessments and provide any other information requested by the Director of the Business Division. Consumers of the Department of Law and Public Safety. At the state level, there is a growing trend among state and local legislators to consider regulations capable of regulating the use of artificial intelligence by government agencies.⁸⁴ In this regard, the State of New Jersey has introduced the New

Medium (20 aprile 2019), <https://medium.com/@sneharevanur/in-a-historic-step-toward-safer-ai-democratic-lawmakers-propose-algorithmic-accountability-act-86c44ef5326d> .

⁸⁰ Algorithmic Accountability Act (2019) par. 2 c. 2 e 3 lett. (b).

⁸¹ Algorithmic Accountability Act (2019) par. 3 lett. (b)(1)(C)

⁸² The definition of algorithmic decision process is in par. 2.1.

⁸³ J. Tail et al., Proposed Algorithmic Accountability Act Targets Bias in Artificial Intelligence, JD Supra (27 giugno 2019), <https://www.jdsupra.com/legalnews/proposed-algorithmic-accountability-act-70186>

⁸⁴ . Kelly, Y. Chae, Insight: AI Regulations Aim at Eliminating Bias, Bloomberg Law (31 maggio, 2019), <https://news.bloomberglaw.com/tech-and-telecom-law/insight-ai-regulations-aim-at-eliminating-bias>

Jersey Algorithmic Accountability Act⁸⁵ which incorporates many aspects of the federal bill. Already at the beginning of 2018, the city of New York issued the first regulation on algorithmic accountability in the United States of America, a local regulation of the automated decision-making systems used by independent agencies in the metropolis. "(Int. N. 1696-2017)⁸⁶. In January 2019, the State of Washington introduced two legislations to protect against algorithmic discrimination, although these are also limited to governmental entities.⁸⁷ In February 2019, California introduced a draft of law, SB 444, by virtue of which private entities that rely on artificial intelligence for the delivery of a product to a public entity are required to disclose to this entity all the measures taken in order to reduce the possible distortions present in the internal algorithm used.⁸⁸

III. ALGORITHMIC DISCRIMINATION IN PRACTICE: A SERIES OF SUBSTANTIVE CASES

An exhaustive study of this issue cannot be achieved without an in-depth study of some substantial cases that concerned the discriminatory protection with respect to the use of algorithms within the decision-making processes developed by the public administration or by private actors. As illustrated in the previous section, the case history is quite wide and it is not possible to analyze it in an exhaustive way, for this reason, this part will focus on three categories concerning the selection of employees and students, online advertising and price discrimination. It is obvious that this examination excludes some very relevant areas such as public safety, automatic translations, the search for images on the network and the automatic completion of search engines.

Selection and evaluation of public employees

It has been showed how artificial intelligence can be used to select potential employees or students. As the example analyzed above relating to the Med School in the UK showed, an algorithmic decision-making process could result in the discrimination of particular social categories. due to the use of distorted data. A similar case involved Amazon⁸⁹ employing an artificial intelligence system to select candidates for job positions in the company, programmers had trained it to find templates in resumes sent to IT technicians in the previous ten years, most of the time. which, due to the demographics of those holding

⁸⁵ New Jersey Algorithmic Accountability Act, A.B. 5430, 218th Leg., 2019 Reg. Sess. (N.J. 2019)

⁸⁶ Int. No. 1696-2017 (N.Y.C. 2018).

⁸⁷ S.B. 5527, H.B. 1655, 66th Leg., 2019 Reg. Sess. (Wash. 2019).

⁸⁸ S.B. 444, 2019 Reg. Sess. (Cal. 2019).

⁸⁹ J. Dastin, Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women, Reuters (October 9, 2018), <https://www.reuters.com/article/us-amazon-com-jobs-automationinsight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>

those jobs, came from male candidates. Consequently the algorithm had developed a model that led him to prefer male candidates, so any sentence that included the word "woman" or its derivations resulted in the automatic exclusion of the curriculum of the candidate.⁹⁰

In the case of Amazon, the company has settled the dispute by setting aside the software in the impossibility of excluding any possible discriminatory outcome of the algorithmic decision-making process, in other cases distinguishing the extent of some data is much more difficult, consequently proving any discriminatory effects becomes complicated. In this sense, the case of *McKinzy v. Union Pacific*⁹¹, in which the plaintiff, a candidate for a job position sued the potential employer, on the basis of alleged racial discrimination by the algorithm.⁹² The proceeding established that the company had used an algorithm to evaluate the resume of the plaintiff, who claimed to have been excluded due to his ethnic origin, the defense of Union Pacific provided data relating to the work experience of the 'actor, pointing out that McKinzy was not qualified for the open position according to data that did not take into account ethnicity. The court ruled in favor of Union Pacific precisely because of its ability to justify the candidate's rejection on the basis of neutral criteria.⁹³ We have seen above how similar criteria can in any case lead to an indirect discriminatory result if they are connected (proxies) with characteristics relating to certain social categories. With respect to indirect discrimination, we have already had the opportunity to talk about how both the Edu Court and the EU directives⁹⁴ allow the exception of the pursuit of a legitimate and necessary objective. In this sense, the Supreme Court of the United States⁹⁵ had already paved the way with the theory of "business necessity", according to which, if an employer can demonstrate that a particular measure or company policy is necessary for the conduct of the business also if this involves the occurrence of indirect discrimination. With reference to cases of algorithmic discrimination, this defense could apply if the employer can demonstrate that the data an algorithm relies on is a "business necessity" or, in other words, is statistically correlated to proper management of the company.

⁹⁰ The importance of the case is well explained in S. Bornstein, *Antidiscriminatory algorithms*. 521.

⁹¹ *McKinzy v. Union Pac. R.R.*, 2010WL3700546(2010).

⁹² *Ibidem*.

⁹³ *Ibidem*.

⁹⁴ *EcHR, D.H. and Others v. Czech Republic (Grand Chamber)*, No. 57325/00, 13 November 2007, par. 187-188; art. 2 c. 2, direttiva 2000/43/EC.

⁹⁵ The theory of business necessity was elaborated in the decision *Griggs v. Duke Power Co.*, 401U.S.424 (1971). See on the question I. Ajunwa, *Automated Hiring*, 41-2. The use of statistical data to prove the business need is recalled both by Ajunwa with regard to the Supreme Court, and by F. Zuiderveen Borgeusius, *Discrimination, artificial intelligence and algorithmic decision-making*, cit., 19 with regard to the Edu Court and EU legislation.

Online advertising

Artificial intelligence is employed for targeted online advertising, it is a profitable sector for some companies, especially Facebook and Google derive most of their profits from this type of advertising. In this field, the algorithm processes the data provided by users through their interactions to propose tailor-made advertising messages. This use of artificial intelligence was reported because it was particularly exposed to discriminatory practices, already in 2013 it was shown how people who searched for names of African American descent, were exposed by the Google search engine to advertisements for recipients of criminal convictions or police history. On the contrary, if you entered names of Caucasian origin, the same search engine identified a significantly lower number of advertisements connected to the judicial records of the recipients. Presumably, the artificial intelligence system that manages Google's search engine analyzed data relating to users' online searches and derived a racial bias with respect to the results to be shown to them.⁹⁶

In 2015, a study conducted by researchers developed a simulation in which users of the Google search engine, who self-declared themselves male or female in the settings, carried out identical online searches on the platform. The researchers then analyzed the ads presented by the algorithm. Google displayed advertisements to male users from a certain consulting agency that promised high salaries significantly more frequently than those offered to female sim users with discriminatory effects. The study also noted that it is not possible to identify the reason why the simulated female users were shown fewer advertisements for higher-wage jobs, due to the opacity of the automatic system that manages the platform and processes the various data entered by users; the interactions between Google, advertisers, websites and users do not allow us to trace the functioning of the algorithm that regulates the appearance of the ads.⁹⁷

This is an example where the black box logic surrounding AI systems makes it harder to uncover discrimination and its causes. Individuals could be discriminated against without being aware of it. If an algorithm targets job ads only to men, women may not realize they are excluded from the advertising campaign.⁹⁸

Another interesting case concerned the methods of insertion of advertisements proposed by the Facebook platform. The social network has allowed advertisers to target

⁹⁶ L. Sweeney, 'Discrimination in online ad delivery' (2013) 11(3) ACM Queue 10.

⁹⁷ A. Datta, M. C. Tschantz, A. Datta, 'Automated experiments on ad privacy settings' (2015) 2015(1) Proceedings on Privacy Enhancing Technologies 92. <https://www.andrew.cmu.edu/user/danupam/dtd-pets15.pdf>; A. Datta et al., 'Discrimination in online advertising: A multidisciplinary inquiry' (Conference on Fairness, Accountability and Transparency 2018) 20. <http://proceedings.mlr.press/v81/datta18a.html>

⁹⁸ C. Jernigan, B. F. Mistree, 'Gaydar: Facebook friendships expose sexual orientation' (2009) 14(10) First Monday <https://firstmonday.org/article/view/2611/2302>.

advertisements to users on the basis of a series of sensitive data processed by its algorithm such as, for example, data relating to sexual preferences.⁹⁹ The Dutch data protection authority has launched an investigation in this area and has recorded a series of false profiles on the platform indicating, among the various information requested, the category "men that are interested in other men"¹⁰⁰. All of the registered profiles did not engage in further interactions on the platform and were exposed to targeted advertising campaigns for that specific category, one by a dating site, the other by an advertiser created by the Guarantor himself to verify operation. of the mechanism: "Therefore it is a fact that the Facebook group has derived the aspect 'men interested in men' from the contents of the profile of the investigation accounts. The selection by the Facebook group of these data subjects based on their sexual preferences for advertising purposes has to be qualified as processing of special categories of data ... "¹⁰¹

A further aspect relating to advertisements on the social network Facebook was the subject of a proceeding before US federal courts. The platform allowed advertisers, through some choice menus called "ethnic affinity" on the Facebook Ad interface, to exclude a series of social categories such as the population of African or Hispanic descent from displaying ads.¹⁰² The platform also allowed the exclusion of specific groups of individuals such as "women in the workforce," "moms of grade school kids," "foreigners"; "Puerto Rico Islanders", or subjects interested in "parenting," "accessibility"; "Service animal"; "Hijab Fashion"; "Hispanic Culture" or the advertising of job advertisements only to people of a certain age group.¹⁰³ A study by Spanish researchers showed that Facebook labels 73% of EU users using classifications based on sensitive data such as "Islam", "reproductive health"

⁹⁹ 'Dutch data protection authority: Facebook violates privacy law', 16 May 2017, <https://autoriteitpersoonsgegevens.nl/en/news/dutch-data-protection-authority-facebook-violates-privacy-law>

¹⁰⁰ Dutch Data Protection Authority, 'Informal English translation of the conclusions of the Dutch Data Protection Authority in its final report of findings about its investigation into the processing of personal data by the Facebook group, 23 February 2017', https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/conclusions_facebook_february_23_2017.pdf, 3

¹⁰¹ Ibidem

¹⁰² On this point J. Angwin et al., 'Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks'. ProPublica, Maggio 2016 <https://www.ProPublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. J. Angwin, A. Tobin, M. Varner, 'Facebook (still) letting housing advertisers exclude users by race', ProPublica, Novembre 2017 <https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin>.

¹⁰³ J. Angwin, N. Scheiber A. Tobin, 'Dozens of companies are using Facebook to exclude older workers from job ads', ProPublica, Dicembre 2017 <https://www.propublica.org/article/facebook-ads-age-discrimination-targeting>

and "homosexuality" in order to give advertisers more opportunities to target own advertisements.¹⁰⁴

On the basis of these findings, a series of non-profit organizations operating in the field of housing rights and anti-discrimination protection in the real estate market have sued Facebook¹⁰⁵ for violating the Fair Housing Act which prescribes that "[t]o make, print, or publish, or cause to be made, printed, or published any notice, statement, or advertisement, with respect to the sale or rental of a dwelling that indicates any preference, limitation, or discrimination based on race, color, religion, sex, handicap, familial status, or national origin, or an intention to make any such preference, limitation, or discrimination"¹⁰⁶. The appeal essentially alleged three discriminatory behaviors on the part of the platform: 1) The ability for advertisers to include or exclude Facebook users from receiving ads based on their gender or age, or based on interests, behavior or demographics. presumed to be related or associated with race, national origin, sex, age, disability or family status; 2) The ability for advertisers to define a narrow geographic area for ad audience that could presumably have a negative impact based on race or national origin; 3) the possible use of the computer tool Lookalike Audience by advertisers which would have allowed the creation of audience segments among Facebook users according to dynamics capable of having a negative impact on various groups, also based on gender, race and age.¹⁰⁷

The parties have decided to reach an out-of-court settlement¹⁰⁸ based on a series of behaviors that the defendant has agreed to adopt to amend the discriminatory practices cited in the appeal, including: a) the platform will create a separate advertising portal for the creation of classifieds for housing, employment and credit ("HEC") on Facebook, Instagram and Messenger which will have limited targeting options, to prevent discrimination; b) Facebook will develop a webpage where Facebook users can search and view all accommodation ads that have been placed by advertisers for renting and selling accommodation regardless of whether users have received such property listings on their wall ; c) the platform will require advertisers to certify compliance with Facebook's policies prohibiting discrimination and all applicable anti-discrimination laws; d) Facebook will allow plaintiffs to verify the testing of the platform to ensure that the changes established by the agreement are effectively implemented; f) Finally, Facebook will engage academics,

¹⁰⁴ This study is available online J. G.Cabañas, Á. Cuevas, R. Cuevas, A. Arrate, 'Facebook use of sensitive data for advertising in Europe' (2018) <https://arxiv.org/pdf/1907.10672.pdf>.

¹⁰⁵ NFHA v. Facebook complaint, Case 1:18-cv-02689 (June 25 2018).

¹⁰⁶ 42 U.S.C. § 3604(c).

¹⁰⁷ NFHA v. Facebook complaint, Case 1:18-cv-02689 (June 25 2018). <https://nationalfairhousing.org/wp-content/uploads/2019/03/2018-06-25-NFHA-v.-Facebook.-First-Amended-Complaint.pdf>.

¹⁰⁸ NFHA v. Facebook settlement agreement (March 18 2019), <https://nationalfairhousing.org/wp-content/uploads/2019/03/FINAL-SIGNED-NFHA-FB-Settlement-Agreement-00368652x9CCC2.pdf>.

researchers, civil society experts, civil rights / liberties and privacy advocates (including plaintiffs) to investigate the potential for unintended bias in the algorithmic modeling used by social media platforms.¹⁰⁹

An appeal was filed by the Housing and Urban Development Department of the United States of America based on the same findings as the case just reported.¹¹⁰ With a summons in the summer of 2019, the Federal Department sued the Facebook platform for the violation of the Fair Housing Act. The fate of this dispute is not known to date, the discriminatory practices referred to extend up to the first quarter of 2020.¹¹¹

Price discrimination

Artificial intelligence makes it possible to classify customers and consequently to differentiate the price for identical products based on the data collected. This practice is called price differentiation and consists in identifying the "highest value that the potential buyer intends to pay for the purchase of the product or service and thus determines an erosion of the consumer surplus which can constitute a characteristic element of the market based on data analysis"¹¹² A company can develop an algorithm which, through cookies and other data deriving from interactions with users, classifies them as sensitive (or insensitive) to the price. These two categories are then used to charge each consumer the maximum price he is willing to pay.¹¹³

Princeton Review, a US company that offers online tutoring services, has charged different prices in different areas of the United States for identical costs for providing its service throughout the United States.¹¹⁴ One study found that the algorithmic system that handled the company's price differentiation resulted in higher fares for people of Asian descent: Customers in geographic areas with a high density of Asian residents had stronger chance to receive higher prices, regardless of their income.¹¹⁵ According to the same researchers, in this case the company engaged in a completely involuntary discriminatory behavior, probably, the differentiation of prices was tested in different geographical areas and when

¹⁰⁹ Ibidem.

¹¹⁰ 42 U.S.C. § 3604(c).

¹¹¹ HUD v. Facebook complaint, si può consultare il documento all'indirizzo elettronico https://cdn.theatlantic.com/assets/media/files/hud_v_facebook.pdf

¹¹² A. Ottolia, Big Data e innovazione computazionale, Torino, 2017, 316; D.M. Kochelek, Data Mining and Antitrust, in Harv. J.L. & Tech., 2009, 22, 515 ff.

¹¹³ M. Maggiolino, Big Data e prezzi personalizzati, in Concorrenza e Mercato, 2016, 95 ff.

¹¹⁴ F. Zuiderveen Borgesius, Discrimination, artificial intelligence and algorithmic decision-making, 16.

¹¹⁵ J. Angwin, S. Mattu, J. Larson, 'The tiger mom tax: Asians are nearly twice as likely to get a higher price from Princeton Review' (2015) ProPublica. <https://www.ProPublica.org/article/asians-nearly-twice-as-likely-to-get-higher-price-from-princeton-review> .

the algorithm learned that in some of these areas, individuals were buying the same amount of services even at higher prices continued to update the price to maximize the surplus based on demand. In practice, however, this has meant that some ethnic groups have paid more for the same service than others.¹¹⁶

The study of these substantial cases has, in some way, strengthened the idea of the dualism between the *ex ante* and *ex post* approaches. The second, which is articulated on the basis of general principles that must be applied to the various concrete situations, proves to be too abstract and difficult to use in a context such as that of new technologies, where flexible rules are necessary and capable of adapting to all possible developments of the technique. The *ex ante* approach, on the other hand, has several uses because it tries to concretely regulate many of the scenarios that can occur in the use of artificial intelligence. In this sense, regulations, codes of conduct and guidelines are essential that can be useful both to programmers when developing the algorithmic system, and to the supervisory authorities, in the exercise of their powers. As it clearly emerged, almost all cases were resolved through cooperation between the public or private organization that developed the algorithm and the supervisory authority or its counterpart in the proceedings. The case of the Dutch data protection authority and housing associations demonstrates how the search for an agreement between the parties, where possible, allows for greater effectiveness of anti-discrimination protection and favoring its greater implementation.

BY WAY OF CONCLUSION

It is difficult to draw a definitive picture of a discipline that is still in full evolution. Technological development in the field of artificial intelligence is rapidly expanding and law, by its very nature, is struggling to keep up. The two approaches that have been used to present the anti-discrimination discipline with respect to the use of algorithms represent the basis on which this effort is being addressed. As mentioned, the *ex ante* approach is often preferable, especially in those contexts where it is possible to intervene with independent authorities such as the data protection authorities who are able to specify the regulatory material to adapt it to different realities. In those areas where such cooperation is more difficult, the *ex post* approach can make up for any gaps in the regulation. A synergy between standards and specific rules seems to be the best way to extend effective anti-discrimination protection to algorithmic decision-making processes.

¹¹⁶ J. Larson, S. Mattu, J. Angwin, 'Unintended consequences of geographic targeting', Technology Science 2015. <https://techscience.org/a/2015090103/>

