UDK 342.7:004.8
Original scientific paper

ALGORITHMIC DISCRIMINATION: A BLUEPRINT FOR A LEGAL ANALYSIS*

Patricia Živković, PhD, Lecturer
University of Aberdeen, School of Law
High Street, Aberdeen, AB24 3UB, United Kingdom
patricia.zivkovic@abdn.ac.uk

Rossana Ducato, PhD, Senior Lecturer
University of Aberdeen, School of Law
High Street, Aberdeen, AB24 3UB, United Kingdom
rossana.ducato@abdn.ac.uk

ABSTRACT

The paper aims at providing an overview of the issues raised by algorithmic discrimination, and the key contributions proposed in the literature to address them. It is intended to be used as a starting point for those interested in approaching the topic for the first time or as a syllabus for the students taking the Erasmus+ Strategic Partnership MOOC "Time to Become Digital in Law".

First, the contribution will outline what algorithms are and what we consider algorithmic bias and what are its causes. Second, it will investigate the ethical and social implications of algorithmic bias. Then, the paper will focus on how existing laws and regulations can be applied to algorithmic discrimination. This contribution will focus in particular on the two branches of law that have been identified in the literature as the most relevant in this context: anti-discrimination law and data protection law. The work will outline their potentialities and limitations, presenting some proposals advanced in the literature to fill the new and emerging gaps of protection.

***Keywords*:** *Artificial Intelligence, Bias, Algorithmic discrimination, Anti-discrimination law, Data Protection Law*

* This paper is co-funded by the Erasmus+ Programme of the European Union. The paper reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

1. INTRODUCTION

Algorithmic bias is more and more commonly discussed in academic circles, and it necessitates perspectives from different disciplines, such as computing science, psychology, ethics, sociology, law, and others. It is a phenomenon that is by its nature multidisciplinary and interdisciplinary, and for that reason, it is also difficult to properly understand and regulate. At the heart of our paper lies the general concern of the possibility to replicate biased attitudes held by humans into machines and new discriminatory machine-generated practices. To provide a holistic view of the topic, one needs to understand the sources of automation bias, the ethical and social implications of such bias, and the current protection offered by the existing legal framework.

This is what we provided in the Massive Open Online Courses (MOOC), "Algorithmic discrimination: a blue-print for a legal analysis", that we designed for the project "Time to Become Digital in Law" (https://www.pravos.unios.hr/digin-law/), co-funded by the Erasmus Plus Programme of the European Union. This paper follows the structure of the MOOC, and it serves as a basic introduction to key legal issues raised by algorithmic discriminatory practices and the ways to counteract them. It is intended to be used as a starting point for those interested in approaching the topic for the first time or as a syllabus for the students taking the MOOC.

The paper starts with an explanation of what algorithms are and what we consider algorithmic bias and what are its causes (Section 2). This is a fundamental point to understand before we investigate the ethical and social implications of algorithmic bias (Section 3). We will stress in this part the difficult role of law to capture these implications timely and to follow rapid technological development. The paper will then focus on how existing laws and regulations can be applied to algorithmic discrimination. This contribution will focus in particular on the two branches of law that have been identified in the literature as the most relevant in this context: anti-discrimination law (Section 4) and data protection law (Section 5). We will outline their potentialities and limitations, presenting some proposals advanced in the literature to fill the gaps of protection.

2. ALGORITHMS AND AUTOMATION BIAS EXPLAINED

An algorithm is an abstract, formalised description of a computational procedure that can be used, *inter alias*, for automated decision-making.[1] Such a decision-

[1] Zuiderveen Borgesius, F., *Discrimination, Artificial Intelligence and Algorithmic Decision-Making*, Council of Europe, 2018, Strasbourg, p. 11.

making process can be fully automated or partly automated.[2] These forms of decision-making will depend on whether there is a human in the loop and to what extent: in the case of a fully automated algorithm, the decision is made entirely by an algorithm; whereas with partly automated algorithms humans are making the final decision in the end.[3] However, both partly and fully automatic decision-making may lead to discrimination. Hence, in this paper, we refer to "algorithmic discrimination" whether the discriminatory practice is performed via solely automated decision-making or support systems.

In general terms, algorithmic discrimination usually results from the lack of time, context, skills, and knowledge to assess the adequacy of automatically made decisions.[4] In the past two decades, this phenomenon has attracted the attention of academics and practitioners in law, computing science, psychology, and other disciplines and both state and corporate use of these machines has been flagged as an issue to be approached with caution and proper investigation.[5] To provide a broader understanding of this phenomenon, we will explore in this Section the discrimination risks involved in algorithmic decision-making and the fields most affected by those risks.

Algorithmic discrimination is complex and sensitive topic when it comes to machine learning. Machine learning systems are the most well-known artificial intelligence ("AI") systems.[6] These systems, instead of being given predetermined sets of solutions, are set a task and provided with training data, based on which they make decisions.[7] They will be at the heart of this paper, and the notions of "AI-based systems" and "machine learning" will for that purpose be used interchangeably.

AI-based systems have already been widely integrated into today's society and are used by all of us. The newly proposed legislations and regulations are, among

---

2 *Ibid.*

3 *Ibid.*

4 *Ibid.*

5 Kearns, M.; Aaron Roth, A., *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*, Oxford University Press, Oxford, 2020; Broussard, M., *Artificial Unintelligence: How Computers Misunderstand the World*, First MIT Press paperback edition, The MIT Press, 2019; Zuiderveen Borgesius, *op. cit.*, note 1; O'Neil, C., *Weapons of Math Destruction How Big Data Increases Inequality and Threatens Democracy*, Penguin Books, 2018; Zuboff, S., *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, First Trade Paperback Edition, Public Affairs, 2020; Webb, A., *The Big Nine*, Ingram Publisher Services US, 2019.The MIT Press 2019

6 Zuiderveen Borgesius, *op. cit.*, note 1, p. 13.

7 *Ibid.*

other reasons, based on the need to scrutinise the systems and prevent any integration of discriminatory practices and or results in their use.

One may naturally start with a question: How can AI lead to discrimination? To answer that, we need to get acquainted with the term "black box". The black box phenomenon in relation to AI means that it is often unclear to human beings how the AI system makes decisions, and this makes it difficult to assess whether there is any discrimination.[8]

Barocas and Selbst list six technical examples, where the sources of potential discrimination are discernible, although with some effort, and hence, the discriminatory practices and results stemming from the use of an AI-based system can be understood from their roots:

1. Defining the "target variable" and "class labels",
2. Training data: labelling examples,
3. Training data: data collection,
4. Feature selection,
5. Proxies, and
6. Intentional discrimination.[9]

The first example of an AI system leading to discrimination relates to the notions and defining process of target variables and class labels. The target variable is an outcome of interest, or in other words, the outcome the user wishes to achieve by using the system.[10] Class labels represent values relevant to the target variable which are mutually exclusive.[11] To showcase how defining the target variable and class labels can lead to discrimination, we can use an example of almost any performance assessment AI system. If we want to assess the performance of employees, we would need to define what a "good" or "desirable" employee is, and what a "bad" employee is, and these would be class labels.[12] A desirable employee could then be defined as an employee who is rarely or never late, and an undesirable or bad employee could be defined as someone who is often late.[13] The potential for discrimination lies in these definitions as the reason for being late to work can stem from the social context. For example, people who are on average poorer may

---

8 *Ibid.*, p. 15.

9 Barocas, S.; Selbst, A. D., *Big Data's Disparate Impact*, California Law Review, Vol. 104, No. 3, 2016, p. 671; as reported in Zuiderveen Borgesius, *op. cit.*, note 1, pp. 15–23.

10 Barocas; Selbst, *op. cit.*, note 9, p. 678; Zuiderveen Borgesius, *op. cit.*, note 1, p. 16.

11 Zuiderveen Borgesius, *op. cit.*, note 1. p. 16.

12 Barocas; Selbst, *op. cit.*, note 9, p. 678; Zuiderveen Borgesius, *op. cit.*, note 1, pp. 16–17.

13 Zuiderveen Borgesius, *op. cit.*, note 1, pp. 16–17.

live farther from their work, and this social circumstance makes them more likely to be late.[14] Hence, the system would potentially discriminate against such employees when assessing their work.

The next two examples of an AI system that may lead to discrimination are when the system learns from discriminatory training data. The AI system can either be trained on biased data, or it can learn from a biased sample.[15] In other words, we can say that the old principle "bias in, bias out" is visible in these situations.[16]

Algorithmic bias can result from the use of biased training data in those situations when the training data is collected in the past and does not reflect today's ethical and moral values, which are transposed to anti-discrimination law. For example, if appointment to positions or jobs was previously not allowed to women, or they were discriminated against in the past, and we train the AI system based on historical data, the discriminatory effect will be replicated.[17]

Similarly to the previous example, an AI system may lead to discrimination when the system learns from training data that is collected through a biased sampling procedure.[18] For example, to train a system that is set to predict crime, the data was collected by the police who focused their attention on certain ethnic groups and certain neighbourhoods.[19] Depending on who lives in those neighbourhoods, the AI system will provide biased results against those groups of people.

The fourth example of an AI system that can lead to discrimination relates to feature selection by the user of the system. Namely, users of the AI system may be required to set the features they want to be captured through processing and lead to the target variable, and these need to be simplified for the system to capture them in data.[20] The features, i.e. categories of data, to be analysed by the system do not need to be directly discriminatory, and usually, they are not. However, that does not mean they will not produce discriminatory results. For example, if an AI system is handling many job applications and is tasked to shortlist applicants who have a degree from one of the highest-ranked universities, this could lead to

14 *Ibid,* p. 17.

15 Barocas; Selbst, *op. cit.*, note 9, p. 681; Zuiderveen Borgesius, *op. cit.*, note 1, pp. 17–19.

16 Selmi, M., *Algorithms, Discrimination and the Law*, Ohio State Law Journal, Vol. 82, No. 4, 2021.

17 Similarly in Barocas; Selbst, *op. cit.*, note 9, p. 682.

18 Zuiderveen Borgesius, *op. cit.*, note 1, pp. 18–19.

19 *Ibid.*, p. 19.

20 Barocas; Selbst, *op. cit.*, note 9, p. 688; Zuiderveen Borgesius, *op. cit.*, note 1, p. 20.

discriminatory effects against those that had no economic means to access such education.[21]

The fifth example of an AI-based system that leads to discrimination relates to proxies, which are criteria that are genuinely relevant in marking rational and non-discriminatory decisions, but they indirectly link to biased attitudes.[22] For example, if one uses an AI-based application for approving loan applications, the target variable is to approve loans to those people who will not likely default, and through training, the machine learns that people from certain postcodes default more.[23] Despite being non-discriminatory at its face value, this criterium may lead to a discriminatory effect as it may act as a proxy for racial origin.[24] In other words, a protected characteristic may be encoded in other data, as in this case racial origin is encoded in a postcode.

Finally, the last example of an AI system according to Barocas and Selbst that can lead to discrimination encompasses a situation in which the discrimination is intentional (despite being masked as one of the above examples).[25] For example, if the users of the system have set the task for the system to identify women based on shopping behaviour to market other products to them and adjust the prices.[26]

Based on the outline of the issues presented in this section, one can conclude that it is important to understand that discrimination risks can be hidden and reproduced in different ways and that education on discrimination and a proper understanding of machine learning is crucial for the prevention of such results. There are fields in which AI brings the most discrimination risks and these require special attention in order to provide an adequate legal framework. An observation of not only the technical causes of algorithmic discrimination, but also the ethical and social implications of such applications is a step towards such a legal framework. Such implications are addressed in the next section of this paper.

3. ETHICAL AND SOCIAL IMPLICATIONS OF AI APPLICATIONS

This section will be divided into three parts. First, the components of trustworthy AI will be briefly introduced, followed by a presentation of the definition and

---

21 Barocas; Selbst, *op. cit.*, note 9, p. 688; Zuiderveen Borgesius, *op. cit.*, note 1, p. 20.

22 Barocas; Selbst, *op. cit.*, note 9, p. 691; Zuiderveen Borgesius, *op. cit.*, note 1, p. 21.

23 Zuiderveen Borgesius, *op. cit.*, note 1, p. 21.

24 *Ibid.*

25 Barocas; Selbst, *op. cit.*, note 9, p. 692; Zuiderveen Borgesius, *op. cit.*, note 1, p. 22.

26 Zuiderveen Borgesius, *op. cit.*, note 1, pp. 22–23.

scope of ethical AI and robust AI and the social implications of the lack of such attributes (Section A). After that, the paper will address the phenomenon of ethics washing, which is developed to create and preserve an image of ethical behaviour in the corporate field of Big Tech (Section B). Finally, from the sociological stance, a parallel is drawn between the (unconscious) biases held by programmers and biased programmes as a result (Section C).

A. Trustworthy AI and its social implications

EU guidelines defined three components of Trustworthy AI in 2019.[27] Ethics plays a crucial role in this definition.

Trustworthy AI has three components, which should be met throughout the system's entire life cycle:

1. it should be lawful, complying with all applicable laws and regulations;
2. it should be ethical, ensuring adherence to ethical principles and values; and
3. it should be robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm.[28]

Whereas the lawfulness of AI will be the topic of Sections 4 and 5 in this paper from the stance of anti-discrimination and data protection laws, we will focus on exploring ethical and robust AI in this part.

Achieving Trustworthy AI requires not only compliance with the law; as a matter of fact, laws are not always up to speed with technological developments.[29] Trustworthy AI inevitably requires also compliance with ethical principles and values, and a warranty of the robustness of such a system to prevent any harm to citizens.

There is some overlap between legal and ethical standards. The fundamental rights families are particularly suitable to cover AI systems among the broad range of indivisible rights outlined in international human rights legislation, the EU Treaties, and the EU Charter, making these rights legally enforceable.[30] However, even after adherence to fundamental rights is made legally enforceable, considering ethical norms can help us comprehend how the creation, application, and use of AI systems may conflict with these rights and the values that underpin them.[31] Also, as it

---

27 High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, European Commission, Brussels, 2019.

28 *Ibid.*, p. 5.

29 *Ibid.*

30 *Ibid.*, p. 10.

31 *Ibid.*

will be shown in the next two sections, ethical consideration can impact the policymakers' and legislators thinking when it comes to the regulation of technology.

Concerns about human dignity (or whether AI systems should unjustifiably subordinate, coerce, deceive, manipulate, condition, or herd humans), the principle of prevention to harm (where special attention is being paid to situations where AI systems can cause or aggravate negative impacts due to asymmetries of power or information), the principle of fairness (which is ensuring equal and just distribution of resources), and the principle of explicability (which means that processes need to be transparent and explainable) are the most crucial ethical considerations to be made.[32]

It is important to remember that implementing all the Trustworthy AI principles must be done throughout the system's life cycle. This requires a constant reassessment of such implementation and redesign of the legal framework when needed.

B. Ethics washing

The development of advisory boards, in-house moral philosophers, a focus on human design, and sponsoring "fair" machine learning are just a few of the corporations' attempts to create ethical products that have been made during the past few decades by major tech companies.[33] These initiatives can sometimes be used as a tool for ethics washing because they are not put in place for a good motive, and allow businesses to cite ethics as a legitimate pretext to explain deregulation, self-regulation, or market-based governance.[34]

Bietti warns that, in practice, these advisory councils or in-house moral philosophers have little power to shape internal company policies and that the corporations overstep the focus on human design – e.g. nudging users to reduce time spent on apps – instead of tackling the risks inherent in the existence of the products themselves.[35] What is important to notice is that the use of ethical language *per se* is not ethical washing, however, the misuse and instrumentalization of it for self-regulation and profit is.[36]

At least three possible arguments can be raised against initiatives that use ethical language and self-regulation for internal purposes. The proper application of mor-

---

32 *Ibid.*, pp. 12–13.

33 Elettra, B., *From Ethics Washing to Ethics Bashing: A View on Tech Ethics from Within Moral Philosophy*, 2021, Available at SSRN: [https://ssrn.com/abstract=3914119].

34 *Ibid.*

35 *Ibid.*

36 *Ibid.*

al philosophy can help resolve these. First, choices made by corporate AI ethical committees are constrained internally, subject to high management approval, and reliant on company funding.[37] As a result, when correctly implemented, moral philosophy can guide internal AI ethics committees toward advancing society.[38] Second, if practising moral philosophy is done for financial gain, employer satisfaction, or to earn recognition, it no longer maintains its intrinsic moral significance.[39] As a result, the right application of moral philosophy can guide the pursuit of justice and trust as well as the welfare of society.[40] Third, ethics rhetoric may encourage and support a constrained view of the potential for regulatory reform and stifle discussion.[41] Thus, by empowering activists and fostering social dialogue, the right application of moral philosophy improves society.[42]

Besides the better attempts at implementing the moral philosophy within corporations, and in that way indirectly into programmers' actions, another approach is also to implement these views through training the programmers directly. This is what the next Section will address.

C. Biased programmers or biased data

Two main theories explain most cases of bias in AI systems: the biased training data theory and the biased programmers theory.[43] It is sometimes difficult to distinguish the most contributing source to bias in AI systems.

As explained in the previous section, machine learning applications are often developed using historical data about outcomes, data coming from it would reflect and perpetuate any bias in the real world. The very fact that these were the datasets commonly used, makes it very hard to quantify the extent of this problem.

The second theory emphasizes another factor: biased programmers.[44] The community of programmers developing algorithms is highly non-representative and may exhibit biases that are passed onto the algorithms they write.[45] Some studies, however, found little effect of altering programmer demographics or from program-

37 *Ibid.*

38 *Ibid.*

39 *Ibid.*

40 *Ibid.*

41 *Ibid.*

42 *Ibid.*

43 Cowgill, B. et al, *Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics*, in: Proceedings of the 21st ACM Conference on Economics and Computation, 2020.

44 *Ibid.*

45 *Ibid.*

mers who score worse on psychology measures of implicit bias.[46] This strongly suggests that organizations should strive to ensure data (e)quality, i.e. exert efforts to increase their data reliability and inclusivity. Issuing regular non-technical reminders to programmers about biases would also address the issue at the personal level, just as regular technical education on how to eliminate these biases in the development would do the same at the professional level.[47]

What was discussed so far deals with the preventive methods for algorithmic bias, but until we reach the stage of the utopian seamless prevention of discrimination, we need to look at the available legal framework for the resolution of these issues in practice. That is what Sections 4 and 5 will explain.

4. ALGORITHMIC DISCRIMINATION AND THE ANTI-DISCRIMINATION LEGAL FRAMEWORK

Non-discrimination and data protection law are among the legal areas, identified in the literature, that can offer the most comprehensive set of tools to address the risks to fundamental rights and freedoms caused by algorithmic discrimination.

While these frameworks can respond to some of the challenges outlined in Section 2, several issues remain open and need to be addressed from a *de lege lata* and *de lege ferenda* perspective. To this end, the paper will outline some of the key proposals that have been advanced by scholars to improve the *status quo*.

In this Section, we will deal with the anti-discrimination legal framework and move to data protection law in Section 5.

Non-discrimination is one of the fundamental principles in the European legal context, and it is recognised in several legal instruments at the national (constitutions and national laws), international,[48] and European levels.[49] The principle of

---

46 *Ibid.*

47 *Ibid.*

48 See, for instance, Art. 14 of the Convention for the protection of human rights and fundamental freedoms (ECHR), which prohibits discrimination based on any ground "such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status".

49 At the EU level the principle of non-discrimination is enshrined in both primary (e.g. Art. 2 TEU; Arts. 10, 18, and 45 TFUE; Arts. 10 and 21 of the Charter of fundamental rights of the EU) and secondary law (Council Directive 2000/78/EC establishing a general framework for equal treatment in employment and occupation [2000] OJ L303/2000, Council Directive 2000/43/EC implementing the principle of equal treatment between persons irrespective of racial or ethnic origin [2000] OJ L180/2000, Council Directive 2004/113/EC implementing the principle of equal treatment between men and women in the access to and supply of goods and services [2004] OJ L373/2004, and Di-

non-discrimination essentially entails that everyone shall have equal chances to access to opportunities in society.[50]

The existing legal framework can protect us against various forms of discrimination. For instance, a rule or a practice cannot lead to treating a person in a less favourable way than others in a similar situation because of a characteristic they possess (direct discrimination); neither a neutral provision – virtually applicable to all – can lead to disadvantage a protected person or group in practice (indirect discrimination).

Anti-discrimination law can also protect those persons that are discriminated against because they are associated with a protected group, even if they are not part of it (discrimination by association).[51]

Finally, the legal protection against multiple and intersectional discrimination can be particularly helpful in the context at stake, where algorithms differentiate people based on a number of characteristics and where the discrimination might not exclusively depend on one of them. Multiple discrimination occurs when someone is treated less favourably because of the sum or the sequence of different protected grounds (e.g. a lesbian might be discriminated against because she is a woman and gay).[52] There is intersectional discrimination when the interplay of different protected grounds generates a discriminatory effect that is qualitatively different from either ground taken in isolation. Friedman explains that:

"black women may experience discrimination in a way which is qualitatively different from either white women or black men. Black women share some experiences in common with both white women and black men, but they also differ in important respects. Thus while white women may be the victims of sex discrimination, they may also be the beneficiaries and even the perpetrators of racism. Conversely, black men may experience racism but be the beneficiaries and perpetrators of sexism."[53]

---

rective 2006/54/EC of the European Parliament and of the Council on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (recast) [2006] OJ L204/2006).

50 European Union Agency for Fundamental Rights, *Handbook on European Non-Discrimination Law*, Publications Office of the EU, Luxembourg, 2018, p. 42.

51 This concept can be applied both in cases of direct and indirect discrimination.

52 European Union Agency for Fundamental Rights, *op. cit.*, note 50.

53 Fredman, S., *Intersectional Discrimination in EU Gender Equality and Non-Discrimination Law*, European Commission, Brussels, 2016, p. 7.

This framework can offer a series of tools also when applied in the context of algorithmic discrimination.

For instance, the prohibition of direct discrimination can cover situations where the algorithm bases the decision on a protected ground. However, according to several authors, direct discrimination might be relatively rare in practice.[54] In most cases, an algorithm can treat an individual less favourably based on correlations with a protected ground and not based on the protected ground itself.

Indirect discrimination could offer more grip to address this latter case, but only to a certain extent. Indirect discrimination is an open-ended clause and might be challenging to prove: it has to be verified on a case-by-case basis if a neutral rule impacts a protected category, and the victim should prove, at least, *prima facie* discrimination, usually through statistical evidence.[55] Moreover, a claim of indirect discrimination can be rebutted if the perpetrator has an objective justification (i.e. the differential treatment pursues a legitimate aim and is proportionate).[56]

More generally, it has been pointed out in the literature that anti-discrimination law protection is very much sectorial and covers only a limited number of grounds.[57] If someone is treated less favourably than another one in a similar situation, but the situation cannot fall within one of the protected grounds enumerated in the law, the victim will not be protected.

This shortcoming is particularly relevant in the context of inferential analytics, where data mining activities could identify new high-risk categories or reaffirms structural inequalities that are different from the protected characteristics that the Legislator considered a few years ago. People can be treated unjustly due to low

---

54 Zuiderveen Borgesius, *op. cit.*, note 1; Wachter, S., *Affinity Profiling and Discrimination by Association in Online Behavioral Advertising*, Berkeley Tech. LJ, Vol. 35, No. 2, 2020, p. 367; Xenidis, R., *Tuning EU Equality Law to Algorithmic Discrimination: Three Pathways to Resilience*, Maastricht Journal of European and Comparative Law, Vol. 27, No. 6, 2021, p. 736. *Contra*, Adams-Prassl, J.; Binns, R.; Kelly-Lyth, A., *Directly Discriminatory Algorithms*, The Modern Law Review, Vol. 86, No. 1, 2023, p. 144. Building on the legal rationale of the distinction between direct and indirect discrimination and a through analysis of the case law, Prassl and others argue that the role of direct discrimination is more relevant than generally assumed in the legal discourse, and it could cover some cases of proxy discrimination and sampling bias.

55 Zuiderveen Borgesius, *op. cit.*, note 1.

56 As noted by Prassl and others, when the predictivity of an algorithm is high, this element can be used to support the proportionality claim. The problem, highlighted in the literature, is that an algorithm can be fed with a biased dataset, and if a predictive model is deployed, this latter can reinforce the existing stereotypes and create a risk of self-justifying feedback loops. Adams-Prassl; Binns; Kelly-Lyth, *op. cit.*, note 54.

57 Zuiderveen Borgesius, *op. cit.*, note 1; Wachter, *op. cit.*, note 54; Xenidis, *op. cit.*, note 54.

income, financial difficulties, or degree of education.[58] However, since these are not "protected grounds" under the EU legal framework - nor will it always be possible to demonstrate statistically the relation between the inference drawn by the algorithm and a protected group – the current anti-discrimination framework might be toothless.

Another open challenge of algorithmic discrimination refers to the well-known issue of the lack of transparency of such systems. Profiling is often obscure and the victims of discrimination might not necessarily be aware of how they have been classified by the algorithm and what are the consequences of the correlations made.[59] Indeed, machine learning algorithms are often "black boxes": it might be difficult to understand the logic behind the automated decision system because of the complexity of the algorithm. Algorithms can be black boxes due to legal constraints as well.[60] Many commercial providers often oppose trade secret protection to avoid the disclosure of the parameters of the algorithm. Moreover, profiling is a dynamic activity. Hence, the classification might evolve, uses other variables, and find new correlations and patterns.[61]

All these elements can have an impact in terms of access to justice. First, due to the black box problem and the dynamicity of profiling, it might be difficult for a potential victim even to find that they have been discriminated against.[62] Moreover, if the processing is opaque, the explanation unintelligible, and information cannot be disclosed, it is challenging to provide evidence of the discrimination (or the lack thereof).[63]

As for multiple and intersectional discrimination, they could overcome some of the issues raised by algorithmic discrimination. The automated decision often relies on a combination of factors and characteristics (it is well-known the case voiced by the MIT researchers, Joy Buolamwini and Timnit Gebru, who discovered that darker-skinned women are the group that facial recognition algorithms

58 Zuiderveen Borgesius, *op. cit.*, note 1; Wachter, *op. cit.*, note 54; Xenidis, *op. cit.*, note 54."

59 Zuiderveen Borgesius, *op. cit.*, note 1. *Wachter, S., The Theory of Artificial Immutability: Protecting Algorithmic Groups Under Anti-Discrimination Law*, Tulane Law Review, Vol. 97, No. 2, 2022.

60 Malgieri, G. *Trade Secrets v Personal Data: A Possible Solution for Balancing Rights*, International Data Privacy Law, Vol. 6, No. 2, 2016, p. 102; Wachter, S.; Mittelstadt, B., *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI*, Colum. Bus. L. Rev., No. 2, 2019, p. 494.

61 Wachter, *op. cit.*, note 59.

62 Zuiderveen Borgesius, *op. cit.*, note 1); Wachter; Mittelstadt, *op. cit.*, note 60.

63 Zuiderveen Borgesius, *op. cit.*, note 1; Adams-Prassl; Binns; Kelly-Lyth, *op. cit.*, note 54.

most frequently misclassify).[64] However, multiple and intersectional are not expressly recognised in the law[65] or the case law of the CJEU.[66]

To fill the current gaps in protection, several scholars have suggested broadening the scope of anti-discrimination law because it does not perfectly capture the new risks posed by algorithmic decision-making systems.[67]

The expansion of the scope of anti-discrimination protection law could happen *de lege lata.* Xenidis, for instance, has argued for a purposive interpretation of key crucial concepts, such as the notion of intersectional discrimination.[68] The latter could alleviate the burden of proof of *prima facie* discrimination in the algorithmic context, but it has not been expressly recognised by the CJEU.[69] However, according to the author, such restrictive interpretation is not absolute: she reads some encouraging signs in the case law of the CJEU[70] and in the opinion of the Advocates General (in the case Parris and Léger)[71], where the concept of multiple discrimination could open the way to the recognition of intersectional discrimination as well.[72]

Along the same lines, she contends that a contextual and expansive interpretation of the protected grounds in EU anti-discrimination law is still viable. Despite the sectorial approach recognised in the Directive, the content of the grounds is not expressly defined in the law. Hence, a broad interpretation of these grounds will contribute to making EU equality law more effective because it will protect individuals against the new types of discrimination based on the patterns and correlations identified by algorithms.[73]

---

64 Buolamwini, J.; Gebru, T., *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, in: Proceedings of Machine Learning Research, No. 81, Conference on fairness, accountability and transparency (PMLR 2018), 2018. The case is presented as an example of intersectional discrimination by Xenidis, *op. cit.*, note 54.

65 There is only a brief mention to multiple discrimination in the recitals of Directive 2000/43/EC (Recital 14) and 2000/78/EC (Recital 3).

66 Xenidis, *op. cit.*, note 54.

67 Zuiderveen Borgesius, *op. cit.*, note 1; Wachter, *op. cit.*, note 54; Xenidis, *op. cit.*, note 54; Adams-Prassl; Binns; Kelly-Lyth, *op. cit.*, note 54.

68 Xenidis, *op. cit.*, note 54.

69 *Ibid.*

70 Referring to Case C-152/11 *Johann Odar v Baxter Deutschland GmbH* [2012] ECLI:EU:C:2012:772.

71 Respectively, Case C-528/13 *Geoffrey Léger v Ministre des Affaires sociales, de la Santé et des Droits des femmes and Etablissement français du sang* [2014] ECLI:EU:C:2014:2112, Opinion of AG Mengozzi, and Case C-443/15 *David L. Parris v Trinity College Dublin and Others* [2016] ECLI:EU:C:2016:493, Opinion of AG Kokott.

72 Xenidis, *op. cit.*, note 54, pp. 743-744.

73 *Ibid.*, pp. 750-751.

As we have already pointed out, algorithmic discrimination can affect individuals even not relying on traditional protected grounds. According to Xenidis, the open-ended clause of Art. 21 of the Charter of fundamental rights and the principle of non-discrimination are interesting paths to explore, as they might provide enough flexibilities to address the new situations of harm.[74]

There are, however, challenges in this kind of approach, because the CJEU has been reluctant so far to expand the list of protected grounds. Given this premise, some authors called for a different hermeneutic approach. For instance, Wachter elaborated a new theory of harm that could close the gaps in protection.[75] She demonstrated that the legal rationale and the traditional categories of anti-discrimination law do not match the logic of algorithms. For instance, people can now be discriminated against based on non-protected features (because they are dog owners or video gamers), or characteristics that cannot be meaningfully caught by an individual (e.g. pixels in a picture).

However, such groups can experience the same harm as traditionally protected categories: ultimately, they are not given an equal opportunity to exercise their rights and freedoms, as well as to access goods to further their aims in life.[76]

Wachter notes that AI creates groups with "immutable" characteristics that the individual cannot control. This is what she calls artificial immutability. Such artificial immutability relies on five conditions: opacity (individuals do not know how they have been classified, or what the consequences are of that classification), vagueness (the individual cannot make meaningful decisions because they do not have transparent information), instability (the criteria are dynamic, they can change over time, so it is very difficult to rely on them), involuntariness and invisibility (the inputs processed by the algorithm are not self-evidently meaningful to people), and lack of social concept (the characteristics used by algorithms do not always find a functional equivalent concept in human language).[77]

This proposal has the merit to address the most subtle and invisible forms of algorithmic discrimination, identifying the new "protected grounds" in the attributes that are not under our control.

---

74 *Ibid.* pp. 755-757.

75 Wachter, *op. cit.*, note 59.

76 *Ibid.*

77 Wachter, *op. cit.*, note 59, pp. 43-45.

5. ALGORITHMIC DISCRIMINATION AND THE DATA PROTECTION FRAMEWORK

The data protection framework can offer a few valid tools that can complement the protection granted by anti-discrimination law.

For instance, the fundamental principles and the procedural guarantees laid down in the European framework, such as the Modernised Convention 108[78] and the General Data Protection Regulation (GDPR)[79], can offer a net of protection against those negative/discriminatory consequences suffered by individuals, whether the decision is fully or partially automated.

The data protection fundamental principles, in particular, the principles of lawfulness, fairness, transparency and accuracy require that processing should respect the rights and fundamental freedoms of individuals. Individuals should be informed in a clear and transparent way about how their data are processed by the machine, what the risks for them are, and what the implications are.[80] The accuracy principle should protect them against profiling misclassifications.[81]

Important data subjects' rights correspond to these principles. For instance, individuals enjoy the right to be informed about the key aspects of the processing – including its risks - in a timely and meaningful way (Arts. 12-14 GDPR). This can represent an important tool to counteract the black box problem because a data subject shall be informed about the existence of the automated decision-making process and receive meaningful information about the logic involved.[82]

---

78 Council of Europe, *Modernised Convention for the Protection of Individuals with Regard to the Automatic Processing of Individual Data*, 18 May 2018 (Modernised Convention 108).

79 Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/2016.

80 Zuiderveen Borgesius, *op. cit.*, note 1.

81 Wachter, *op. cit.*, note 54.

82 Kaminski, M. E., *The Right to Explanation, Explained*, in: Sandeen, S. K.; Rademacher, C.; Ohly, A. (eds.), Research Handbook on Information Law and Governance, Edward Elgar Publishing, Cheltenham, 2021, p. 278.or AI. The EU's General Data Protection Regulation (GDPR Although the existence and precise boundaries of the "right to explanation" has been challenged in the literature. See, Goodman, B.; Flaxman, S., *EU Regulations on Algorithmic Decision-Making and a "Right to Explanation"*, AI Magazine, Vol. 38, No. 3, 2017; Wachter, S.; Mittelstadt, B.; Floridi, L., *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, International Data Privacy Law, Vol. 7, No. 2, 2017, p. 76; Malgieri, G.; Comandé, G., *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, International Data Privacy Law, Vol. 7, No. 4, 2017; Edwards, L.; Veale, M., *Slave to the Algorithm: Why a Right to an Explanation Is Probably Not the Remedy You Are Looking For*, Duke L. & Tech. Rev., Vol. 16, No. 1, 2017, p. 18; Selbst, A.; Powles, J., *Meaningful Information" and the Right to Explanation*, in: Proceedings of the 1st Conference on Fairness, Accountability and Transparency (PMLR 2018), Vol. 81, 2018.

The right to access (Art. 15 GDPR) can be used to verify whether someone is processing our data and discover if we are subject to automated decisions. Hence, it could be a tool to investigate potential cases of discrimination. If an individual has been misclassified, they can ask for the rectification of information (Art. 16 GDPR), etc.

A higher level of protection is recognized for the so-called "sensitive data". This category includes data that bear a high risk of discrimination for individuals, namely "data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation" (Art. 9 GDPR and Art. 6 Modernised Convention 108).[83]

Supervisory authorities can also play a fundamental procedural role in fighting algorithmic discrimination.[84] They are independent bodies that have the task to monitor the correct application of the GDPR. They have investigative powers, can perform an audit of the algorithm, and require the necessary documentation to see how it works in practice. Data Protection Authorities must also be consulted depending on the outcome of the Data Protection Impact Assessment (DPIA) performed by the controller.

The DPIA is a comprehensive analysis of the processing that the controller must carry out when the processing can result in a high risk to the rights and freedoms of individuals (Art. 35 GDPR). The GDPR does not explicitly define what is a high risk, but it exemplifies a few cases where a DPIA will be needed.[85] This is, in particular, the situation where the controller performs a "systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person" (Art. 35(3)(a) GDPR). Hence, many AI systems are likely to require a DPIA and the assessment will have to address the risks of discrimination posed by the technology.[86]

---

83 Zuiderveen Borgesius, *op. cit.*, note 1.; Wachter, *op. cit.*, note 54.

84 Zuiderveen Borgesius, *op. cit.*, note 1.

85 Although the Article 29 Working Party (now, European Data Protection Board) has provided some guidelines. See, WP29, *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679*, adopted on 4 April 2017 as last revised and adopted on 4 October 2017, WP 248 rev.01.

86 Zuiderveen Borgesius, *op. cit.*, note 1, p. 22.

Another potential tool to counter algorithmic discrimination is offered by the specific rules in case a fully automatised decision produces legal effects concerning the data subject or similarly significantly affects them (Art. 22 GDPR).[87] In principle, such decisions are forbidden, unless there is an exception, such as the explicit consent of the data subject.[88] And in this latter case, appropriate safeguards should be guaranteed to the individual, so that the latter can challenge the outcome of the decision.[89]

However, also the GDPR presents some loopholes when applied to the problem at stake.

The preliminary limitation is that data protection applies only if the processing concerns personal data, and not all algorithmic operations necessary process personal data.[90] For example, the GDPR might not apply to predictive models: they might be elaborated through the analysis of personal data (to find for example correlations between food preferences and creditworthiness), but the model as such uses mere statistical inferences.

The effectiveness of the principle of transparency and corresponding measures has also been questioned. The existence and the actual boundaries of the "right to explanation" have been at the centre of a lively debate and many scholars are sceptical about its effectiveness.[91] The prohibition under Art. 22 applies only to fully automated decisions, it is not always easy to define in a clear way if the decision can significantly negatively affect individuals, and, in any case, it is not an absolute prohibition.[92] It can be authorized in three important circumstances: a) the explicit consent of the individual; b) legislative provision; c) contractual necessity.[93]

Even when the right to information about the logic behind the algorithm is triggered, a series of obstacles (technical and legal) remain. As previously mentioned, algorithms can be so complex that their logic remains difficult to comprehend even for their developers. If the logic is intelligible, it might be challenging to

87 *Ibid.*; Wachter; Mittelstadt, *op. cit.*, note 60; Wachter, *op. cit.*, note 54.

88 See, Art. 22(2) GDPR.

89 Art. 22(3) GDPR.

90 Zuiderveen Borgesius, *op. cit.*, note 1, pp. 24-25.

91 Wachter; Mittelstadt; Floridi, *op. cit.*, note 82; Edwards; Veale, *op. cit.*, note 82.

92 It has to be noticed, however, that the formulation of such a right in the Modernised Convention 108 is broader as it does not refer to solely automated decisions nor to "the significant effects" for individuals. See, Zuiderveen Borgesius, *op. cit.*, note 1, p. 24.

93 Art. 22(2) GDPR.

translate that information in a meaningful way for laypeople.[94] Moreover, the explanation might be hampered in practice by trade secret protection.[95]

Another set of limitations concerns the rules on sensitive data. Their processing is subject to a higher standard of protection. However, the list of protected grounds is quite narrow and it does not include other vulnerable categories or sensitive information such as sex or socio-economic information.[96]

Moreover, the list contained in Art. 9 GDPR is a *numerus clausus*. Hence, it might be difficult to apply the GDPR protection on sensitive data to, for example, inferred data (non-sensitive as such) leading to a discriminatory outcome.[97]

However, it must be said that the list of special categories of data offers some possibilities for an extensive interpretation (Art. 9 GDPR includes expressions like "data revealing racial or ethnic origin" or "data concerning health"). This wording suggests that sensitive characteristics can be inferred directly but also indirectly.[98] For instance, as recently stated by the CJEU, the publication of information about a spouse's details can indirectly reveal the sexual orientation of the data subject. [99] Hence, it should be considered sensitive data. However, the decision refers to an inference made "following an intellectual operation involving deduction or cross-referencing."[100] It remains to be seen to what extent this reasoning could cover more complex elaborations that could nevertheless cause discrimination.

Finally, even if Data Protection Authorities can play an important role in ensuring the application of the GDPR, there is the concrete problem that many of them are usually understaffed or under-resourced, and they might not be supported by technical experts (which are crucial in a field like algorithmic discrimination).[101] Therefore, the enforcement powers recognized by the GDPR might be more difficult to be exercised in practice.

Also in the field of data protection, several proposals have been presented to improve the existing framework.

---

94 Edwards; Veale, *op. cit.*, note 82.

95 Malgieri, *op. cit.*, note 60; Wachter; Mittelstadt, *op. cit.*, note 60.

96 Zuiderveen Borgesius, *op. cit.*, note 1; Wachter, *op. cit.*, note 54.

97 Wachter, *op. cit.*, note 54.

98 See, Case C-184/20 *OT v Vyriausioji tarnybinės etikos komisija* [2022] ECLI:EU:C:2022:601, Opinion of AG Pikamäe, para 85.

99 Case C-184/20 *OT v Vyriausioji tarnybinės etikos komisija* [2022] ECLI:EU:C:2022:601.

100 *Ibid.*, para. 123.

101 Zuiderveen Borgesius, *op. cit.*, note 1.

As previously mentioned, one of the main shortcomings of data protection in the field of algorithmic discrimination is that not all harmful inferences might be classified as personal data or benefit from the stronger protection reserved for sensitive data.

To enhance this level of protection, Wachter and Mittlestadt have proposed the introduction of a new right: the right to reasonable inferences.[102] This right would address the harmful consequences of high-risk inferences. According to the authors, these latter should include inferences that: a) violate privacy or have the potential to harm someone's reputation now or in the future, or b) are based on opinions or have little possibility of verification but are still used to make crucial decisions.[103]

In order to be effective, this right would be formed of an *ex ante* justification mechanisms and an *ex post* control.[104]

The *ex ante* justification would require controllers to explain and justify "(1) why certain data are a normatively acceptable basis to draw inferences; (2) why these inferences are normatively acceptable and relevant for the chosen processing purpose or type of automated decision; and (3) whether the data and methods used to draw the inferences are accurate and statistically reliable."[105]

In addition to that, the individual would have the *ex post* right to contest the unreasonable inference and provide additional information that could lead to an alternative outcome.[106] According to the authors, this right would complement the right to contest the decision at Art. 22 GDPR.

The same authors articulate a more comprehensive set of recommendations to address high risks inferences.

First, they notice that the current scope of data protection – in order to adequately protect individuals - should be expanded to include "the assessment of the reasonableness of inferential analytics and accuracy of decision-making processes."[107]

Second, they recognise that the level of protection depending on the categorisation of personal, non-personal, and sensitive data, is not effective anymore in the

---

102 Wachter; Mittelstadt, *op. cit.*, note 60.

103 *Ibid.*, p. 580.

104 *Ibid.*

105 *Ibid.*, p. 581.

106 *Ibid.*, p. 588.

107 *Ibid.,* p. 614.

Big Data environment. Neutral data can cause the same harm as sensitive data. A predictive model based on anonymous data can be as privacy-invasive as those created from personal data. Hence, they argue that future legislative interventions or judicial interpretations should focus more on how data is used and its impact and rely less on the concept of identifiability.[108] Thus, they should recognise appropriate redress mechanisms when a predictive model is applied to individuals.

Third, to appropriately support the mechanisms of the right to reasonable inferences, future policy interventions should provide for an obligation of the controller to justify the data sources and the intended inferences, and an ability for the data subject to contest the decision. [109]

Among the other data protection tools that could be used to combat algorithmic discrimination, DPIAs figure prominently. However, in practice, DPIAs focus mainly on data security and data quality (and not other substantial aspects that could help address deeper societal issues).[110] Multi-layered models of Algorithmic Impact Assessment (AIA) have been then proposed to complement the existing system.

Mantelero, for instance, introduced the idea of the Human Rights, Social and Ethical Impact Assessment (HRESIA), a more comprehensive tool for AI developers and providers for assessing the impact of their IT solutions.[111] This tool relies on two main components: on the one hand, self-assessments, questionnaires, and risk assessment instruments; and, on the other hand, consultation with experts. According to Mantelero, the universalist dimension of the HRESIA can provide a framework for "the collective dimension of data use"[112], providing a further tool for protecting non-traditional groups created by algorithms.

Kaminski and Malgieri recognise as well that DPIAs do not work perfectly as AIA.[113] However, the GDPR's DPIA is a useful starting point for designing a solid AIA. They suggest the key elements that this model should have. For example, it should involve civil society as an additional form of oversight, the assessment should consider not only the technology in isolation but in its context of use and

---

108 *Ibid.,* pp. 615-618.

109 *Ibid.,* p. 619.

110 As noticed by Kaminski, M. E.; Malgieri, G., *Multi-Layered Explanations from Algorithmic Impact Assessments in the GDPR*, in: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 2020.

111 Mantelero, A., *AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment*, Computer Law & Security Review, Vol. 34, No. 4, 2018, p. 754.

112 *Ibid.*, p. 771.

113 Kaminski; Malgieri, *op. cit.*, note 110.

on a system-wide level to mitigate social harms, and most importantly, the system should guarantee individual, group and systemic explanations (what they call "multi-layered explanations").[114]

On a more general level, scholars have shown how the problem of the new forms of differentiations that are created by machine learning is calling for a revision of the current legal frameworks and the adoption of new forms of protection. These new interventions should be grounded on empirical evidence and research.[115]

For instance, Borgesius has pinpointed several measures that could improve the *status quo*, such as the provision of more support to Equalities Bodies and Data Protection Authorities and closer collaboration between them (given the mutual interaction between anti-discrimination and data protection law)[116], and the possibility of carving out a broader research exception to intellectual property protecting the algorithm.[117] To complement these measures, education and research remain crucial: special campaigns aimed at the general public could be launched, and Fairness, Accountability, and Transparency (FAccT) studies should be further supported.[118]

6. CONCLUSIONS

Nowadays, AI-powered decision-making systems are routinely used in many sectors and activities. However, their introduction should be carefully assessed and evaluated. As a growing study of literature demonstrates, fully or partly automated can replicate existing biases or create new and more subtle forms of discrimination.

This paper offered an overview of these risks and the ethical and legal attempts to address them.

To develop trustworthy AI, ethical guidelines can serve as a basis. However, big tech companies need to refrain from the instrumentalisation of ethical language for the purpose of profit and self-regulation. It is important to properly apply moral philosophy in development for the benefit of society at large.

From a legal point of view, the anti-discrimination and data protection frameworks provide an array of tools and remedies to combat algorithmic discrimination. This

---

114 *Ibid.*

115 Zuiderveen Borgesius, *op. cit.*, note 1.

116 *Ibid.*, p. 35.

117 *Ibid.*, p. 65.

118 *Ibid.*, p. 28.

framework, however, does not adequately cover the new situation of harm generated by algorithms. To this end, many authors have argued for introducing specific rules or functional interpretations of the current law to close the loopholes.

In this paper, we have provided a blueprint for analysing a complex and dynamic field. Technology is advancing rapidly, but keeping these tools under vigilant and critical scrutiny is crucial in a democratic society. Our legal framework should respond to these challenges in a timely and meaningful way. Hence, a broader consideration of the issues raised by algorithmic discrimination should find a place in the initiatives tabled by the European legislator, such as the proposed Artificial Intelligent Act, which is currently under discussion.

REFERENCES

BOOKS AND ARTICLES

1. Adams-Prassl, J.; Binns, R.; Kelly-Lyth, A., *Directly Discriminatory Algorithms*, The Modern Law Review, Vol. 86, No. 1, 2023, pp. 144-175
2. Barocas, S.; and Selbst, A. D., *Big Data's Disparate Impact*, California Law Review, Vol. 104, No. 3, 2016, pp. 671-732
3. Broussard, M., *Artificial Unintelligence: How Computers Misunderstand the World*, First MIT Press paperback edition, The MIT Press, 2019
4. Buolamwini, J.; Gebru, T., *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, in: Proceedings of Machine Learning Research, No. 81, Conference on fairness, accountability and transparency (PMLR 2018), 2018, pp. 1-15
5. Cowgill, B. et al, *Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics*, in: Proceedings of the 21st ACM Conference on Economics and Computation, 2020, pp. 1-5
6. Edwards, L.; Veale, M., *Slave to the Algorithm: Why a Right to an Explanation Is Probably Not the Remedy You Are Looking For*, Duke L. & Tech. Rev., Vol. 16, No. 1, 2017, pp. 18-84
7. Elettra, B., *From Ethics Washing to Ethics Bashing: A View on Tech Ethics from Within Moral Philosophy*, 2021, pp. 1-15, Available at SSRN: [https://ssrn.com/abstract=3914119].
8. Goodman, B.; Flaxman, S., *EU Regulations on Algorithmic Decision-Making and a "Right to Explanation"*, AI Magazine, Vol. 38, No. 3, 2017, pp. 50-57
9. Kaminski, M. E., *The Right to Explanation, Explained*, in: Sandeen, S. K.; Rademacher, C.; Ohly, A. (eds.), Research Handbook on Information Law and Governance, Edward Elgar Publishing, Cheltenham, 2021, pp. 278-299
10. Kaminski, M. E.; Malgieri, G., *Multi-Layered Explanations from Algorithmic Impact Assessments in the GDPR*, in: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 2020, pp. 68-79
11. Kearns, M.; Aaron Roth, A., *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*, Oxford University Press, Oxford, 2020

12. Malgieri, G., *Trade Secrets v Personal Data: A Possible Solution for Balancing Rights*, International Data Privacy Law, Vol. 6, No. 2, 2016, pp. 102-116
13. Malgieri, G.; Comandé, G., *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, International Data Privacy Law, Vol. 7, No. 4, 2017, pp. 243-265
14. Mantelero, A., *AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment,* Computer Law & Security Review, Vol. 34, No. 4, 2018, pp. 754-772
15. O'Neil, C., *Weapons of Math Destruction How Big Data Increases Inequality and Threatens Democracy*, Penguin Books, 2018
16. Selbst, A.; Powles, J., *Meaningful Information" and the Right to Explanation*, in: Proceedings of the 1st Conference on Fairness, Accountability and Transparency (PMLR 2018), Vol. 81, 2018, p.1
17. Selmi, M., *Algorithms, Discrimination and the Law*, Ohio State Law Journal, Vol. 82, No. 4, 2021, pp. 611-651
18. European Union Agency for Fundamental Rights, *Handbook on European Non-Discrimination Law*, Publications Office of the EU, Luxembourg, 2018
19. Wachter, S., *Affinity Profiling and Discrimination by Association in Online Behavioral Advertising*, Berkeley Tech. LJ, Vol. 35, No. 2, 2020, pp. 367-430
20. Wachter, S., *The Theory of Artificial Immutability: Protecting Algorithmic Groups Under Anti-Discrimination Law*, Tulane Law Review, Vol. 97, No. 2, 2022, pp. 1-50
21. Wachter, S.; Mittelstadt, B., *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI*, Colum. Bus. L. Rev., No. 2, 2019, pp. 494-620
22. Wachter, S.; Mittelstadt, B.; Floridi, L., *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, International Data Privacy Law, Vol. 7, No. 2, 2017, pp. 76-99
23. Webb, A., *The Big Nine*, Ingram Publisher Services US, 2019
24. Xenidis, R., *Tuning EU Equality Law to Algorithmic Discrimination: Three Pathways to Resilience*, Maastricht Journal of European and Comparative Law, Vol. 27, No. 6, 2021, pp. 736-758
25. Zuboff, S., *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, First Trade Paperback Edition, Public Affairs, 2020

COURT OF JUSTICE OF THE EUROPEAN UNION

1. Case C-152/11 *Johann Odar v Baxter Deutschland GmbH* [2012] ECLI:EU:C:2012:772
2. Case C-528/13 *Geoffrey Léger v Ministre des Affaires sociales, de la Santé et des Droits des femmes and Etablissement français du sang* [2014] ECLI:EU:C:2014:2112, Opinion of AG Mengozzi
3. Case C-443/15 *David L. Parris v Trinity College Dublin and Others* [2016] ECLI:EU:C:2016:493, Opinion of AG Kokott
4. Case C-184/20 *OT v Vyriausioji tarnybinės etikos komisija* [2022] ECLI:EU:C:2022:601

5. Case C-184/20 *OT v Vyriausioji tarnybinės etikos komisija* [2022] ECLI:EU:C:2022:601, Opinion of AG Pikamäe

COUNCIL OF EUROPE

1. Convention for the Protection of Individuals with Regard to the Automatic Processing of Individual Data, 28 January 1981, ETS 108, as amended by Protocol, 18 May 2018, CETS 223
2. European Convention for the Protection of Human Rights and Fundamental Freedoms, 4 November 1950, ETS 5

EU LAW

1. Charter of Fundamental Rights of the European Union [2012] OJ C 326
2. Council Directive 2000/43/EC implementing the principle of equal treatment between persons irrespective of racial or ethnic origin [2000] OJ L180/2000
3. Council Directive 2000/78/EC establishing a general framework for equal treatment in employment and occupation [2000] OJ L303/2000
4. Council Directive 2004/113/EC implementing the principle of equal treatment between men and women in the access to and supply of goods and services [2004] OJ L373/2004
5. Directive 2006/54/EC of the European Parliament and of the Council on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (recast) [2006] OJ L204/2006
6. Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679, adopted on 4 April 2017 as last revised and adopted on 4 October 2017, WP 248 rev.01.
7. Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/2016
8. Treaty on European Union (Consolidated version) [2016] OJ C 202
9. Treaty on the Functioning of the European Union (Consolidated version) [2016] OJ C 202

REPORTS

1. Fredman, S., *Intersectional Discrimination in EU Gender Equality and Non-Discrimination Law*, European Commission, Brussels, 2016
2. High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, European Commission, Brussels, 2019
3. Zuiderveen Borgesius, F., *Discrimination, Artificial Intelligence and Algorithmic Decision-Making*, Council of Europe, Strasbourg, 2018