FACULTY OF LAW
Lund University

Axel Johnson

Machine-learning and Discrimination: Procedural Challenges of Algorithmic Decision-making

LAGM01 Graduate Thesis

Graduate Thesis, Master of Laws program
30 higher education credits

Supervisor: Niklas Selberg

Semester of graduation: Period 1 spring semester graduation 2021

Table of Contents

Abstract

The emergence of artificial intelligence, especially machine-learning methods, challenges the set of legal guarantees put in place in Europe to combat discrimination and ensure equal treatment. This paper will focus on cases of algorithmic discrimination in the context of recruitment as a business practice. Particular in the field of recruitment of workers, which has always been a field where EU non-discrimination law has hardened and evolved, the use of machine-learning algorithms in recruitment processes has triggered a debate on the application of the non-discrimination principles in EU. Beyond the discussion about the applicability of current non-discrimination law in algorithmic discrimination cases, it also important to shed light on the procedural challenges of such cases. Algorithms challenge two principles in the system of evidence in EU-non discrimination law. The first is effectiveness, given that due to the natural opacity of algorithms, the parties do not have easy and unrestricted access to information enabling them to support their claims. The second is fairness, which is an enormous task due to the algorithmic opacity, placing unrealistic burdens of proof on claimants as well as on respondents. Discrimination in such cases seems impossible to prove and, consequently, falls outside the scope of EU non-discrimination law.

However, through an examination of current principles and case-law of Union law, two possible remedies are proposed in this paper. Regarding effectiveness, a joint reading of EU non-discrimination law and the GDPR, could recognize a right to access evidence in favour of victims of algorithmic discrimination. Regarding fairness, a more proportionate way to allocate the burden off proof is suggested by extending the grounds for defence of respondents. This is done by allowing a respondent to establish that biases were autonomously developed by an algorithm. All in all, this paper shows that from a legal point of view, many problems posed by algorithmic discrimination reinforce weaknesses and shortcomings that already exists in the legal framework. Nevertheless, changes and adaptations such as those suggested might help mend the gaps and differences between those that wish for a rapid and broad development of AI, leaving legal protection in the wake, and those that wish for a careful and steady development that gives regulation a chance in the game.

Acknowledgements

The author would like to thank Niklas Selberg for excellent and patient guidance during the my work with the thesis, especially with helping me to find a suitable topic.

The author would also like to thank the whole department of law working at Lund University for providing a great learning experience, despite the different circumstances during the pandemic.

Finally, the author would like to thank his family and friends for their support. Thank you.

List of Abbreviations

AI	Artificial Intelligence
ADM	Automated Decision-Making
CFR	Charter of Fundamental Rights of the European Union
CJEU	Court of Justice of the European Union
DL	Deep Learning
EC	European Commission
EU	European Union
GC	General Court
GDPR	Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data,
TEU	Treaty on the European Union
TFEU	The Treaty on the Functioning of the European Union

1. Introduction

1.1 Background

Algorithms and extensive data sets are increasingly involved in decisions that do not only have trivial consequences for people, but also influence their way of living and personality development. Algorithms generate conclusions and outcomes that are used by human decision-makers as an information basis for their decisions or the decision-making is completely delegated to algorithms or the computer systems containing them.

Apart from early analyses of bias an unequal treatment through the use of computer systems the risks of discrimination in connection with information and communication technologies have only been thoroughly addressed in this decade. Particularly with the emergence of big data developments and the increased usage of algorithms, researchers and policymakers have pointed out the associated risks of discrimination.[1]

In recent decades, the amount of personal data and data relating to an identifiable person that is generated as a product or by-product of computerisation (digitalization), not only by and between organisations but also in public and private spheres of life, has grown rapidly. The collection of usage, location and movement data from mobile devices, the recording of the diverse uses of the internet, such as communication in social media, search engines, website visits and evaluation of browser histories and electronic financial transactions and payment systems are all significant sources of personal data.

Society is increasingly driven by intelligent systems and the automatic processing of vast amounts of data. The mediation of life through computation means that predictions, classifications and decisions can be made about people, on the basis of algorithmic models trained on large datasets of historical trends. Where borrowers were once evaluated for financial loans on a narrow range of historical and qualitative factors, they may now face

[1] E.g. Gerards, Janneke & Xenidis, Raphaele, *"Algorithmic Discrimination Europe: Challenges and opportunities for gender equality and non-discrimination law*, (2021) Publications Office of the European Union.

opaque assessments based on a wide range of seemingly unrelated attributes.[2] For instance, online lenders observe behaviours that have been found to correlate with creditworthiness, such as speed with which potential borrowers scroll through their website, or whether they use capital letters correctly when filling loan applications.[3]

The rapidly growing use of machine-learning (ML) and artificial intelligence (AI) techniques expand the boundaries in terms of accuracy, efficiency and rapidity in everyday life application.[4] Algorithms therefore influence how opportunities and possibilities open up for, and are accessed by, individuals and entire groups. But at the same time as these techniques can potentially provide society at large with a more equal and broader access to a wide array of goods and services, they also entail unprecedented risks of discrimination.[5] AI and ML systems are from hereby referred to as algorithmic systems, if not specifically specified otherwise.

There are many concrete examples relayed by the media that show how grave, multi-facetted and pervasive the problem of algorithmic discrimination can be. For an example, Microsoft launched a chatbot in 2016, Tay.Ai, which had to be turned down after only 24 hours because it had transformed into a racist and sexist online hate speech machine.[6] Another example is Amazon's AI recruiting tool, were the algorithms were trained to vet applicants by observing patterns in resumes submitted to the company over a ten year period.[7] However, the system taught itself that male candidates were preferable. It penalized resumes that included the word "women's", as in "women's chess club captain". Furthermore, it downgraded graduates of two all-women's colleges. Although Amazon edited the system to make it neutral to these

---

[2] Joe Decille, "Leaky Data: How Wonga Makes Lending Decisions." (Charisma: Consumer Market Studies, 20 May 2013), https://www.charisma-network.net/finance/leaky-data-how-wonga-makes-lending-decisions, <accessed 24 May 2021>.

[3] Lobosco, Katie, "Facebook Friends Could Change Your Credit Score", (CNNMoney, 27 August 2013), https://money.cnn.com/2013/08/26/technology/social/facebook-credit-score/index.html, <accessed 24 May 2021>.

[4]

[5] Lentz, Aurore, "Garbage in, garbage out: is AI discriminatory or simply a mirror of IRL inequalities?" (Universal Rights Group, 18 January 2021), https://www.universal-rights.org/blog/garbage-in-garbage-out-is-ai-discriminatory-or-simply-a-mirror-of-irl-inequalities/ ,<accessed 3 April 2021>.

[6] Hunt, Elle, "Tay, Microsoft's AI Chatbot, gets a crash course from Twitter", (The Guardian, 24 Mars 2016), https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter, <accessed 24 May 2021>.

[7] Dastin, Jeffrey, "Amazon scraps secret Ai recruiting tool that showed bias against women", (Reuters, 11 October 2018), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G <accessed 24 May 2021>.

specific terms, it was no guarantee that the algorithms would not devise other ways of sorting candidates that could prove discriminatory.[8]

The application of Amazon's recruiting tool, and more specifically, the discovery that it discriminated women, underlines the troubling fact that AI is capable of expressing a bias that is neither intended nor expected by its user.

From an EU-law perspective, there has been a debate on the future application of non-discrimination law. In conclusion, current legal framework and principles are blunt tools in the face of the challenges stemming from algorithmic discrimination.[9] Although there are several issues, all embedded with their own challenges, presented in the debate, two issues in particular seem to cause serious concern – agency and opacity. The non-human nature of algorithmic discrimination upsets the traditional legal assumption that non-human systems cannot be held accountable for discriminatory or any other kind of harmful conduct.[10] Regarding opacity, the main challenge lies in the fact that the inner processes of machine learning algorithms cannot be fully known. In addition, given that such systems are capable of analysing a large number of variables with incredible speed, it seems almost impossible to fully understand the process by which they arrive at a prediction or decision.[11]

In this context, legal researchers have focused on which ways algorithmic discrimination can challenge the law conceptually.[12] Others have underlined the importance of regulation, suggesting that algorithmic discrimination should be prevented through a series of principles and rules aimed at increasing fairness in programming stages of such algorithms.[13] However, fewer studies focus on the procedural challenge of algorithmic discrimination.

---

[8] Ibid.
[9] Hacker, Philipp, (2018), "*Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law*", 55 Common Market Law Review 1143, p. 1178.
[10] Ibid., p. 1147.
[11] Gerards, Janneke & Xenidis, Raphaele, *"Algorithmic Discrimination Europe: Challenges and opportunities for gender equality and non-discrimination law*", p. 27.
[12] Ibid., p. 31.
[13] Kleinberg, J and others, "*Discrimination in the Age of Algorithms*", (2018), Journal of Legal Analysis, Vol 10, p. 154.

Given that ML systems are most likely to become primary decision-makers in a vast array of context, such as recruitment of workers, concerns are expected to follow because of the lack of effective procedural means to challenge or defend algorithmic decisions.[14]

1.2 Purpose and research question

A debate on the future application of non-discrimination law within the EU has gained momentum during the past years and much suggests that current non-discrimination doctrine does not provide a satisfactory solution for cases of algorithmic discrimination.[15] Fewer studies, however, focus on how cases of algorithmic discrimination can be procedurally challenging. It has been found that ML-systems are capable of expressing biases, the next step is to analyse how it can be challenging in front of a court. More specifically, to examine how the effects of such biases could be established and rebutted under current EU non-discrimination law. This refers to the rules governing the establishment and rebuttal of the presumption of discrimination in the non-discrimination directives.

Since ML systems are predicted to become leading decision-makers in fields of recruitment, there are not only concerns about the current legal regime protection against algorithmic discrimination, but also regarding the lack of effective procedural means to challenge algorithmic decisions.

Algorithmic discrimination will be examined as the result of biases occurring in ML systems, so-called black box scenarios. These are decision-making processes that are impenetrable and are almost impossible to audit. Similar biases arising in the context of supervised or semi-supervised systems will not be considered, because in cases with human supervision the casual link between an unfair bias and the user of an algorithm can be more plausible established.

This paper aims to study the nature of algorithmic decisions in relation to evidence, with emphasis on effectiveness and fairness, within the context of recruitment. Effectiveness

[14] Gerards, Janneke & Xenidis, Raphaele, *"Algorithmic Discrimination Europe: Challenges and opportunities for gender equality and non-discrimination law"*, p. 141.
[15] Gerards, Janneke & Xenidis, Raphaele, *"Algorithmic Discrimination Europe: Challenges and opportunities for gender equality and non-discrimination law",* p. 152.

reflects the possibility of the parties to access facts and adduce evidence in support of their claims, whereas fairness reflects a proportionate allocation of the burdens of proof. Subsequent part will focus on the procedural, substantial and personal aspects of the process of proving and justifying algorithmic discrimination. The system of evidence in EU-non discrimination law relies on the presumption that non-human systems do not possess the capacity to develop and express biases in a fully autonomous way. However, if there is no direct human intervention of discrimination, the rules and principles of EU non-discrimination law might be ill-adapted to provide an efficient and fair way to challenge such sort of discrimination.

The principle of effectiveness is challenged by the opacity of algorithmic biases, insofar as it obstructs the access of victims to information revealing the presence of and discriminatory nature of an algorithm. Regarding the principle of fairness, the human agent postulate is a concern since the users of algorithms are by default presumed to be the creators of algorithms discrimination, absent means to establish that they are not directly responsible for the occurrence of an algorithmic bias.

The research question of this paper is:

- What procedural challenges entails a case of algorithmic discrimination and what tools are there to effectively remedy those challenges?

1.3 Methodology & Definitions

The main method that is being used throughout this paper is the legal dogmatic method. Under the traditional view, the legal dogmatic research entails two main parts, which are the core of the methodology, systematization and interpretation of legislation.[16] Firstly, the systematization of legal rules is made through the construction of legal concepts. The hierarchy of the sources used the method are mostly those that are used in the legal process; primarily statutes, which is supplemented by case law and by literature expanding the rule and lastly a reflection on those rules.[17] Secondly, the interpretation of the legal rules is reached through an examination of their content and their application. Therefore, the legal

[16] Nils Jareborg, 'Rättsdogmatik som rättsvetenskap' (2004) SvJT 1, 4.
[17] McCrudden, Christopher, "*Legal research and the social science*", (2006), The Law Quarterly Review, Oxford Legal Studies Research Paper No. 33/2006, page 633.

dogmatic method answers the questions by looking at accepted legal sources, such as European or national primary and secondary law.[18]

Traditionally, legal sources are divided into three groups, primary, secondary, and supplementary sources of law. Primary law includes fundamental rights of the constitutions, national laws and international treaties that are incorporated into national legislation. EU law, which is external to national law, is also considered to be a binding source that consists of the Treaty on European Union (TEU)[19] and the Treaty on the Functioning of the European Union (TFEU)[20]; Charter of Fundamental Rights of the European Union[21] and the European Convention on Human Rights.[22]

The body of law that stems from principles and objectives of the mentioned treaties is known as secondary law, which consists of regulations, directives, decisions, recommendations, and opinions listed in Article 288 TFEU. These sources of law are binding, and they shall not be disregarded.[23]

In addition to primary and secondary sources of law, this paper will also build upon reports, online journals and articles of legal scholars, for instance Philipp Hacker, Raphaele Xenidis and Fredrick Borgesius actively debating on the, sometimes, underrated risks that follows with the package that is algorithms. A few comparisons are also made to US case law, in order to see comparisons and differences.

In this paper AI refers to a class of computer programs designed to solve problems requiring inferential reasoning, decision-making based on incomplete or uncertain information, classification, optimization and perception.[24]

---

[18] Hettne, Jörgen and Eriksson, Ida, ”*EU-rättlig metod, Teori och genomslag i svensk rättstillämpning*”, (2011, 2nd edition, Nordstedts juridik), p. 40.
[19] Consolidated version of the Treaty on the European Union [2012] OJ C 326/01.
[20] Consolidated version of the Treaty on the Functioning of the European Union [2012] OJ C 326/01.
[21] Charter of Fundamental Rights of the European Union [2012] OJ C 326/391.
[22] Protocol 1 to the European Convention for the Protection of Human Rights and Fundamental Freedoms.
[23] Craig, Paul & De Búrca, Gráinne,” *EU Law: Text, Cases, and Materials*” (6th edn, Oxford University Press 2015), p. 266.
[24] Bathaee, Yavar, “*The Artificial Intelligence black box and the failure of intent and causation*”, 31 Harvard Journal of Law & Technology (2018), p. 898.

1.4 Delimitations

Due to practical constraints such as limited time and the complexity of the topic the scope and quality of this paper are restricted in several aspects. The reader should bear in mind that the aim of this thesis is limited to conducting exploratory research. AI, and algorithmic discrimination, is a complicated and inter-disciplinary field of study that requires knowledge of many areas.

It should be stressed that this paper will focus on cases of algorithmic discrimination in the context of recruitment as a business practice. Discrimination stemming from policy or collective agreements will not be examined. Although, direct discrimination will be addressed, this paper will mainly focus on indirect discrimination throughout the analysis. Such discrimination or different treatment may be justified under EU non-discrimination, which facilitates an interesting analysis.

The potential impacts of algorithms and the challenge that they pose to non-discrimination law is broad and complicated topic. In a longer perspective, the use and development of algorithms may create challenges that are impossible to foresee at present. With that in mind, this paper does not attempt to address the long-term effects of algorithmic discrimination. Instead, focus will be turned to the current and near future impacts.

Another delimitation derives from the nature of algorithms. Regulation on AI and algorithms is a developing, complex and far-reaching topic hence the paper cannot provide a comprehensive review of it. Furthermore, the scope of this paper is within EU law, since the European Union have adopted several strategies and initiatives to counter algorithmic discrimination, such as the General Data Protection Regulation (GDPR). Nevertheless, although the European Union have presented several official documents regarding AI such as the White Paper on AI[25], regulation draft on AI[26] and an overall strategy on AI[27], they leave a few stones unturned. Since they are of general character, they do not specifically address any

[25] White Paper, "*On Artificial Intelligence – A European approach to excellence and trust*", COM (2020)65 final.
[26] Proposal for a *REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS*, COM/2021/206 final.
[27] European Commission (2018), *Communication from the Commission to the European Parliament, The European Council, The Council, The European Economic and Social Committee and the Committee of the Regions – Artificial Intelligence for Europe*, COM(2018) 137 final.

procedural differences to cases of algorithmic discrimination, therefore they are not further presented in this paper.

1.5 Disposition

This paper consists of six chapters. Following the introduction, the second chapter sets out to create a basic understanding of algorithms, with emphasis on ML systems. The third chapter concerns EU non-discrimination law in relation to algorithms and their nature. The fourth chapter concerns what challenges a case of algorithmic discrimination in a recruitment process might entail from a procedural perspective. Firstly, it presents the core of evidence theory, building upon law and case law. Subsequently, casting the light on what hurdles there are for a claimant as well as for a respondent in a case of algorithmic discrimination. The fifth chapter concerns on how to remedy the gaps found in previous chapters and examines the findings of the paper and proposes potential solutions. The final discussion in the sixth chapter brings the previous analyses together, to fulfil the overall purpose of the paper, and adding a final conclusion.

2. Algorithmic decision-making

The use of the term "automated decision-making" has become common in both scientific and legal practices. The term addresses both the use of algorithms for decision-making support of a human decision maker and the automated execution of decisions.

For this paper, the decision-making process can be divided abstractly into several steps, ranging from the recording of the outcomes of the data analyses, the evaluation of the situation and the alternatives including the reconciliation of predefined conditions, the selection between alternatives, to the triggering of an action.[28] In a decision that has been taken in a completely automated fashion, all steps of the decision-making rules are executed by software. These are the sort of autonomous algorithmic decisions that are in focus for this paper. However, it is important to also present a brief overview of the vast array of algorithms and their decision-making processes.

How data analysis and decision-making processes relate in practice is very different and can be divided into several types.[29] One type is when automated data processing and the decision-making process are separated, and the outcomes of the data processing are virtually "manually" transferred to automated decision-making processes or programmed there as decision-making rules.[30] Another type is when the data processing is integrated into the decision-making process. For example, with data mining and machine learning methods the outcomes in the form of optimised models can be directly embedded in the decision systems as programme components as rules of differentiation.[31] The distinction between the two types is important for the identification of discrimination, since in the latter case, the outcomes are often less comprehensible.[32]

[28] Parasuraman, Raja & Riley, Victor, "Humans and Automation: Use, Misuse, Disuse, Abuse", 39 Human Factors, 1997:39(2), p. 232.
[29] Barocas, Solon & Selbst, Andrew, "Big Data's Disparate Impact", (2016), 104 Californa Law Review, p. 677.
[30] Kleinberg, J and others, "Discrimination in the Age of Algorithms", p. 115.
[31] Barocas, Solon & Selbst, Andrew, "Bid Data's Disparate Impact" (2016), p 679.
[32] Gurney, Kevin, *"*An introduction to neural networks*",* first published 1997, UCL Press, p. 13.

2.1 What is an algorithm?

Algorithms can be defined as a set of instructions that, based on a series of unput data, can produce certain value or set of values as output.[33] Some algorithms can directly inform a decision, such as a decision to grant a social security benefit to a specific person.[34] Other algorithms mainly calculate probabilities, such as the probability that a certain deviation in human cells is indicative of cancer or a person with certain qualifications is well suited for a particular position. Such algorithms usually do not directly inform decisions, but they can support decisions by human beings. In the first example, a doctor may take account of the probability calculations made by the algorithm in helping her or him diagnose cancer. In conclusion, there are different types of algorithms, and they can have different functions.

These differences between algorithms may play a crucial role in the discussion of algorithmic discrimination and non-discrimination law, this chapter will briefly present a few of the differences.[35] Another section of the chapter will present in what way, depending on their characteristics, the different types of algorithms can have various functions, such as automated decision-making, pattern detection, classification, clustering, probability calculations etc.

2.2 The muncipiality of algortihms

2.2.1 Ruled-based algorithms

Most of the procedures of decision-making require a logical thought process that can be simplified to "if this, then that".[36] As an example, a legislative rule may state that driving faster than 40 km per hour outside a school is prohibited and that violation is punishable by a fine. Therefore, if a person has been found to have been driving faster than the speed limit outside a school, then the consequence must be that she or he has to pay a fine. Although

[33] Borgesius, Fredrick, "*Discrimination, Artificial Intelligence, and Algorithmic Decision-Making",* (Council of Europe, Directorate General of Democracy 2018), p. 10.
[34] Kulk S & others (2020), "*Legal aspects of algorithms that make decisions. An exploratory study*", The Hague, WODC 2020, p. 2.
[35] Orwat, Carsten, "*Risks of Discrimination through the Use of Algorithms*", (2020), Federal Anti-Discrimination Agency, p.40.
[36] Surden, Harry, "*Artificial Intelligence and Law: An Overview*" (2019), Georgia State University Law Review 1305, p. 1311.

most decision-making processes are more complex, the example shows that if rules are sufficiently clear and the variables are known, a thought process can be split into different "if this, then that" and creates a decision tree.[37]

Such processes can quite easily be translated into computer instructions or algorithms.[38] This results in rule-based algorithms that are highly predictable, since the set of rules and instructions are fixed and all possible variables and outcomes are programmed into the algorithm.[39] When fully developed, such algorithms can relatively easily replace human decision-making, since humans only need to feed the algorithm with the relevant data, and the algorithm can then automatically produce the output in the shape of a decision that in line with the process that humans have devised. The main purpose or main function of such rule-based algorithms is to, with speed, handle large volumes of decisions, ensure consistence and reduce the number of human mistakes in conducting repetitive tasks.[40]

2.2.2 Machine-learning

In general, machine-learning methods enable that mathematical models, which represent the linear or non-linear relationships between different variables, are optimised on data sets, which are characterised by a large, previously unknown set of variables.[41] A distinction is often made between the learning or training phase on the one hand and the application or productive phase on the other. [42] As an outcome of the training phase, recognised correlations or patterns are converted into models, and in the application phase, incorporated into decision-making rules and transferred to new decision situations.[43]

In simplified terms, machine-learning consists of training a model with the following elements:

- 1) collecting and compiling a data set,

[37] Géron, Aurélien, "*Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and techniques to Build Intelligent Systems"*, (2019), O'Reily Media, p.177.
[38] Surden, Harry, "*Artificial Intelligence and Law: An Overview*" (2019), p. 1319.
[39] Ibid, p. 1320.
[40] Ibid, p. 1319.
[41] Lehr, David & Ohm, Paul, (2017)"*Playing with the Data: What Legal Scholars Should Learn About Machine Learning*", UC Davis Law Review, Vol. 51, p. 658.
[42] Ibid, p. 660.
[43] Ibid,p. 660.

- 2) specifying a concrete outcome to be predicted in the data set,
- 3) deciding which possible influence variables are formed and provided to the training algorithm to be considered in the final model,
- 4) constructing a procedure to find the best influence variable that uses all other variables to predict the desired outcome, which can be used to make predictions about the outcome. For an example, the rating of a person, and finally,
- 5) the validation of the procedure with a retained part of the data set that was not used for training. [44]

In the training phase, a mathematical model with learning algorithms is optimised in an iterative process by gradually adjusting the parameters with the aid of feedback until the model is best adapted to the data set. Typically, one or more initial models are used, on which different learning algorithms are tried out until one of the learning algorithms produces the best performance of the model in terms of the most accurate prediction or estimation of the outcome.[45]

In the testing phase, the model is applied to the test data set and checked for "overfitting" or "underfitting". Usually, the problems of these two arise when the generated model "fits" too much with the training set only (does not generalise well) and does not "fit" well to the original data set from which the training data was taken.[46]After the two phases, the generated model can be used in the so-called productive phase to actually make predictions or classifications based on new data.

In comparison to conventional programming, machine-learning methods can save both time and money or enable the processing of complex data processing tasks in the first place. In the case of traditional programming, for an example, to detect and filter out spam emails, there is need to programme rules for individual terms, patterns or typical email components that are known to be common in spam email. This would require a complex list of all the features that define a spam email, which would also have to be programmed at great expense if spammers

[44] Kleinberg, J and others, "Discrimination in the Age of Algorithms", (2018), p.132.
[45] Géron, Aurélien, "*Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and techniques to Build Intelligent Systems*, (2019), p. 30.
[46] Ibid, pp. 26-29.

changed the components. When using machine learning for spam filters, emails that users have marked as spam and lets the learning algorithm detect the relevant components.[47]

Typical characteristics of machine-learning methods are able to process more dimensions of variables than conventional statistical methods, they are able to sometimes achieve higher accuracy in prediction or categorisation.

2.3 Stages of decision-making and uses

In the academic discussion it has been shown, generally, that three stages can be distinguished in processes of algorithmic decision-making, regardless the type of algorithm that is being used.[48] These distinctions are useful for the purposes of this paper, since different risks of discrimination can be seen to be involved on each of the three stages and different actors can be held responsible for such instances of discrimination.[49] Put differently, discrimination can infect algorithms from their inception to their end use, leading to consequences in the context of non-discrimination law. It is therefore important and necessary to provide clarity as to the various phases and intervening actors. Although it is certainly possible to further differentiate between different elements and phases of decision-making in the context of algorithmic discrimination, this threefold distinction is a convenient starting point for discussing algorithmic discrimination.

2.3.1 Planning stage

The first stage distinguished is that of problem analysis and planning.[50] Firstly, this encompasses defining a particular objective for the use of an algorithm by a company or public body. For an example, to deploy personnel more efficiently or to set prices for certain services or products etc. Once the objectives have been established, they can be used to decide which type of algorithm is best suited in serving that objective. One of the factors that need to be considered in this decision-making process is how the output of the algorithm will

[47] Ibid, pp. 4-6.
[48] Kulk S & others (2020), "*Legal aspects of algorithms that make decisions. An exploratory study*", p. 2.
[49] Ibid, p. 2.
[50] Ibid, p. 2.

be used. As mentioned above, efficiency increasing automated decision-making can sometimes require the use of rule-based algorithms, while self-learning algorithms can be more useful if the objective sets a need to predictions of human behaviour or profiling.[51] In making a choice it is also important to consider the specific characteristics of an algorithm.[52] Rule-based algorithms are typically highly predictable, in that all relevant parameters, variables and choices can be pre-determined as part of the development process. One the algorithmic model is ready, there will be no surprises. This also entails that that this sort of algorithm is relatively rigid – it cannot independently take account of any changing contextual circumstances, such as new ideas on what would be an acceptable price or a reasonable fine.[53] If such a rule-based algorithm stops generating expected decisions, it will have to be reprogrammed.

Supervised-learning algorithms are trained by using labelled and known sets of data, which usually reflect the situation at a particular moment. Therefore, they risk becoming outdated relatively quickly if contextual factors change. Such algorithms must be updated and revalidated rather frequently and may be less suitable to operating in highly dynamic contexts.[54]

In relation to unsupervised-learning and deep-learning algorithms, these are much more adaptable and flexible.[55] For example, if the data changes because of societal developments, these algorithms can train themselves to discover new patterns in the new data.[56] At the same time, the disadvantage of such flexible and adaptable algorithms is that it can be difficult to explain how they work and how they adapt, and users and even experts have little control over such adaptions.[57] In the planning stage, users will need to make other choices. They must decide whether they want to develop their own algorithm or purchase one that has already been produced by an external provider.[58]

---

[51] Surden, Harry, ”*Artificial Intelligence and Law: An Overview*” (2019), p. 1319.
[52] Larus, James & others, “*When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making*”, 2020, (Technical Report Informatics Europe & EUACM), p. 10.
[53] Géron, Aurélien, “*Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and techniques to Build Intelligent Systems”*, (2019), p. 18.
[54] Ibid., p. 19.
[55] Lehr, David & Ohm, Paul, (2017), “*Playing with the Data: What Legal Scholars Should Learn About Machine Learning”,* p. 698.
[56] Ibid., p. 698.
[57] Lehr, David & Ohm, Paul, (2017), “*Playing with the Data: What Legal Scholars Should Learn About Machine Learning”,* p. 698.
[58] Kulk S & others (2020), “*Legal aspects of algorithms that make decisions. An exploratory study*”, p. 2.

2.3.2 Development stage

In the development stage, data scientists and other experts write the computer code that are necessary to build the algorithm and, in the case of self-learning algorithms, allow them to be trained or engage in deep learning.[59]

If rule-based algorithms are to be built, this stage comprises the rules and decision-making processes into different steps and rebuilding them in computer code, making choices as to the appropriate variables and the type of decisions the algorithm eventually should make.[60] Technical experts often do this in cooperation with experts in the field, such as policy and legal experts, for systems that will assist in or take over administrative decision-making processes.[61]

If self-learning algorithms are to be used, an important part of the development stage is to decide exactly which type of learning must be applied in order to achieve the objectives set in the first stage. In addition, it must be decided which technologies for data analysis are best suited to achieving those objectives.[62] If self-learning algorithms are being developed, the development process further comprises the preparation of data, such as labelling in supervised-learning processes, to make them suited for the process of training and learning. The development stage also encompasses the actual training and feedback processes, and eventually the testing and validation of the algorithm. [63]

2.3.3 Decision-making stage and application

When the algorithm has been developed, tested and validated, it is ready to be used. In other words, it can be fed with new input data and can start generating output that can be used to achieve the objectives set in the first stage. How the algorithmic output is used differs for different types of algorithms and will depend on the objectives.

---

[59] Barocas, Solon and Selbst, Andrew (2016), "*Big Data's Disparate Impact*", p. 768.
[60] Ibid., p. 768.
[61] Kroll, Joshua & others, (2017), "*Accountable Algorithms*", 165 University of Pennsylvania Law Review 633, p. 702.
[62] Barocas, Solon and Selbst, Andrew (2016), "*Big Data's Disparate Impact*", p. 768.
[63] Kroll, Joshua & others, (2017), "*Accountable Algorithms*", p. 703.

One type of use is that the algorithmic directly generates a decision without further human intervention - automated decision making (ADM).[64] This type may be relevant for routine-decision-making processes, such as the imposition of fines in simple traffic offence cases or making bulk decisions in social security. In many such cases the decisions can be made using rule-based algorithms.[65] Self-learning algorithms may produce output that can also directly generate a decision and can be very powerful in doing so.[66] Many web shops nowadays use self-learning algorithms for online price determination, making personalised offers to users, ensuring targeted and individualised newsfeeds. [67] Furthermore, automated decisions can be made by means of a combination of different types of algorithms. For an example, self-driving cars, where many sensorial systems and different types of algorithms work together to allow the car to make the decision to brake or divert if it meets an obstacle.[68]

The output of the algorithm may also be used to support a human decision, which implies that there is a human in the loop where the actual decision making is concerned.[69] Put differently, the human and the algorithm produces a team effort. That is the case, for example, when a medical specialist uses an algorithmic analysis of medical imaging to check a patient's diagnosis or when a recruitment officer make decisions on the suitability of job applicants supported by an algorithmic analysis of success factors. In both cases, the human is the decision-maker **of** the team and can ultimately ignore or overrule the decision suggested by the algorithm. [70]

Finally, part of the final stage is monitoring to see whether the algorithm continues to generate reliable and acceptable outcomes. [71] Rule-based and supervised-learning algorithms may easily become outdated, while machine-learning or deep-learning algorithms may

---

[64] Larus, James & others, (2020) "*When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making*", p 5.
[65] Surden, Harry, "*Artificial Intelligence and Law: An Overview*" (2019), p. 1317.
[66] Larus, J (2018) and others, "*When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making*" , p. 6.
[67] Borgesius, Fredrick and Poort, Joost, (2017), "*Online Price Discrimination and EU Data Privacy Law*" Journal of Consumer Policy 40 347, p. 357.
[68] SOU 2018:16, "*Vägen till självkörande fordon – introduktion*", p. 563.
[69] Fagan, Frank and Levmore, Saul (2019), "*The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion*", Southern California Law Review 93 2, p. 13.
[70] Fagan, Frank and Levmore, Saul (2019), "*The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion*", p. 14.
[71] Castelluccia, Claude and Le Métayer, Daniel, (2019), "*Understanding algorithmic decision-making: Opportunities and challenges*", Panel for the Future of Science and Technology (STOA) of the European Parliament, p. 54.

unexpectedly generate unwarranted outcomes – discrimination – because they have learnt themselves to identify correlations that do not reflect causal relationships, or because they are not able to deal with certain types of new data.

2.4 Ways of training

Despite of what has been presented in this paper so far, algorithms cannot make any of the analyses or have any of the functions discussed above from their own motion. They require special development and training to analyse data in a particular way, find patterns, make calculations, predict future behaviour etc.[72] There are several different available ways of doing so and this section will briefly present these ways.

2.4.1 Supervised learning

The first method is supervised learning, which is often used in relation to classification of data. In conclusion, the algorithm is fed carefully selected and pre-categorised data – labelled data. For example, messages that clearly contain different forms of spam and is instructed that these data disclose a certain category.[73] New data are then fed into the algorithm and it is asked to recognise the same or similar patterns in that data. If the algorithm recognises the relevant patterns in the labelled data correctly, it is given positive feedback, whereas it is given negative feedback if it fails to categorise the data correctly.[74] Using that feedback as a springboard, the algorithm will gradually correct and improve itself until it has reached a satisfactory level. Once that point is reached, the algorithm can be validated. This means that it can operate safely outside the controlled context of a data lab and without the pre-labelled data.[75] The process is also called supervised machine learning, since data scientist are closely involved in giving feedback to the algorithm. [76]

[72] Denham, Elizabeth, (2017), ”*Big Data, artificial intelligence, machine learning and data protection*”, Information Commissioner’s Office, p. 9.
[73] Ibid., p. 9.
[74] Géron, Aurélien, “*Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and techniques to Build Intelligent Systems*”, (2019), p. 11.
[75] Ibid., p. 12.
[76] Ibid., p. 12.

2.4.2 Unsupervised learning

Other types of algorithms are based on unsupervised learning.[77] Unlike the supervised learning algorithms, these algorithms are provided with a set of instructions and a large amount of data in which they are instructed to discover correlations and patterns autonomously.[78] It is therefore not possible to constantly check whether they discover the correlations and patterns in the way that human beings would.[79] The only checks that can be made are in relation to the algorithms output. This is mainly the reason why these algorithms are called "black box" algorithms. On the basis of the output, feedback can then be given to the algorithm, which in turn can use that feedback to correct itself if needed.[80]

2.4.3 Reinforcement learning

The last form of machine learning is reinforcement learning. This means that an algorithm is confined to a certain environment and instructed to achieve a certain objective.[81] For an example the environment could be an employment context and the objective is to find the best-suited candidate for a certain job. Provided that the objective is achieved, the algorithm is given positive feedback.[82] If it is not, the feedback will be negative. In the end, using regression analysis, calculations of probabilities and recognition of patterns and correlations, the algorithm can learn which of a wide range of scenarios is related most closely to the aim that it is asked to realise, and which actions or decisions would best contribute to achieving that very aim.[83]

2.5 Summary

The aim of this section has been to clarify and define a number of basic notions related to algorithms. The main distinction is one between rule-based algorithms and machine-learning algorithms. Rule-based algorithms are based on a fixed set of instructions, variables and rules

[77] Ibid., p. 13.
[78] Ibid., p. 13.
[79] Denham, Elizabeth, (2017), "*Big Data, artificial intelligence, machine learning and data protection*", p. 10.
[80] Denham, Elizabeth, (2017), "*Big Data, artificial intelligence, machine learning and data protection*", p. 10.
[81] Hamon, Ronan and others, (2020), "*Robustness and Explainability of Artificial Intelligence*", Publications Office of the European Union, p. 10.
[82] Ibid., p. 11.
[83] Ibid., p. 11.

that are programmed into a computer, and that result in highly predictable output. Therefore, these algorithms can easily be used for automated decision-making (ADM).

Machine-learning or deep learning algorithms can be developed and trained to analyse data in a particular way, find relevant patterns, make calculations, trace correlations and predict future behaviour and much more. Since machine-learning algorithms are capable to express a bias that was neither intended nor expected by its user, these will be used in the analysis of this paper.

In relation to the application of algorithms, three different stages have been identified – the planning stage, the development stage and the decision-making stage. At each of these stages, different individuals and organisations may play a role, and the challenges, risks and problems related to non-discriminatory law may manifest themselves differently.

3. EU non-discrimination law

What actions are considered discriminatory is viewed differently in different societies, eras and local regions. The demarcation is the outcome of societal conflicts, negotiations and agreements. This is reflected, above all, in human and fundamental rights as well as in the laws and institutions that substantiate and enforce these rights. [84]

This paper follows the common understanding of discrimination in the EU and understands discrimination as disadvantageous, unjustified unequal treatment of persons in connection with a protected characteristic.[85] The unequal treatment is based on categories and attribution of characteristics to persons. The categorisation and formation of such characteristics may, for example, be based on stereotyping, rational calculations, prejudices or be unintentional. Several legal catalogues define these categories and characteristics as legally protected characteristics – called discrimination grounds – according to which persons must not be disadvantaged in unjustified ways. Unjustified primarily means that there is no objective reason or objective justification for the unequal treatment.[86] Put differently, unequal treatment may in itself also be acceptable from a societal point of view if a recognised objective reason exists for this.

3.1 The legal framework

Non-discrimination law is a general principle enshrined in Articles 2 and 3(3) TEU. Article 10 TFEU embodies the fight against discrimination in EU policies and activities. In addition to sex, racial or ethnic origin, religion or belief, disability, age and sexual orientation are all grounds as set out in Article 19 TFEU. Furthermore, Article 21 of the EU Charter of the Fundamental Rights sets out a prohibition against discrimination based on grounds of racial or ethnic origin, religion or belief, disability, age and sexual orientation, along with other grounds such as colour, social origin, genetic features, language, political or any other opinion, membership of a national minority, property and birth.

[84] Barnard, Catherine, "*EU Employment Law*", 4th Edition, Oxford University Press 2012, p. 254.
[85] Ibid., p.277.
[86] Ibid., p.287.

In terms of secondary law, discrimination in relation to racial or ethnic origin is prohibited by Directive 2004/43/EC (Racial Equality Directive) in employment matters, social protection, including social security and healthcare, social advantages, education and the access to and supply of goods and services.[87]

The grounds of religion or belief, disability, age and sexual orientation are protected under another framework, Directive 2000/78/EC (Equal Treatment Directive), which, unlike the Racial Equality Directive, specifically applies in employment matters.[88] As a consequence, discrimination on grounds of religion or belief, disability, age and sexual orientation is not prohibited in relation to education, social security, and access to goods and services including healthcare, housing, advertising and the media. This is a problem well known among lawyers that work with discrimination law and has been described as an undue hierarchy of grounds in EU equality law.[89] A proposal to bridge this gap has been proposed by the European Commission but the Council has not yet reached an agreement.[90]

3.2 Protected grounds and algorithmic discrimination

Beyond the gaps in the material scope of EU non-discrimination law and gender equality law, obscurities in the personal scope create further weakness in the light of the problem of algorithmic discrimination. As discussed above, the legal frameworks are characterised by their closed lists of grounds. Discrimination is prohibited only if it can be shown that be based on sex, race or ethnic origin, religion or belief, disability, sexual orientation and age or to disproportionately disadvantage a person or group identified by one of the listed characteristics. The forms of discrimination stemming from the use of algorithms set some challenges in relation to the protected grounds that define the personal scope of EU law.

[87] Council Directive 2000/43/EC of June 29 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin (Racial Equality DIrective) (200) OJ L 180/22.
[88] Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation OJ L 303.
[89] Howard, E (2006), "*The case for a considered hierarchy of grounds in EU law"*, 13 Maastricht Journal of European and Comparative Law 445.
[90] European Commission (2008) Proposal for a Council Directive on implementing the principle of equal treatment between persons irrespective of religion or belief, disability, age or sexual orientation COM (2008) 426 final, OJ C303/8 (European Union, 2008).

3.3 Charter of Fundamental human rights

In the light of what has been presented so far, algorithmic discrimination asks sharp questions to EU-non-discrimination law. With this in mind the Charter of Fundamental Rights of the European Union (the Charter) deserves closer attention.

The text of Article 21(1) states that "any discrimination based on any grounds such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation shall be prohibited".[91] Furthermore, Article 21(2) adds that within the scope of application of the Treaties and without prejudice to any of the specific provisions, any discrimination on grounds of nationality shall be prohibited.[92] Consequently, Article 21 establishes a non-exhaustive and open-ended list of discrimination grounds by prohibiting discrimination based on *any* ground as the characteristics listed.

Since 2009, the Charter has had the same value as the Treaties. However, the CJEU clarified through the judgements of *Chacon Navas*, *Coleman* and *Kaltoft* that only those grounds that find expression in secondary law can be held to be protected by the Directives.[93] In Coleman and Kaltoft, the Court stated that the scope of the Equal Treatment Directive should not be extended by analogy beyond the discrimination based on the grounds listed exhaustively in the directive.[94] Consequently, the CJEU held in *Kaltoft* that obesity could not as such be regarded as a ground in addition to those in relation to which the Equal Treatment Directive prohibits discrimination on and could only be protected to the extent it could relate to a ground already protected under EU law.[95] This has been understood as a dismissal of the Court to the potential of Article 21 as a way to introduce much needed flexibility in the personal scope of EU non-discrimination law.[96]

---

[91] Article 21(1) of the Charter of the Fundamental Rights of the European Union (CFR) OJ C 202/389.
[92] Article 21(2) CFR.
[93] Judgment of 11 July 2006, *Sonia Chacón Navas v Eurest Colectividades* SA C-13/05 EU:C:2006:456; Judgment of 17 July 2008, *S. Coleman v Attridge Law and Steve Law* C–303/06 EU:C:2008:415 and Judgment of 18 December 2014, *Fag og Arbejde (FOA) v Kommunernes Landsforening* (KL) C-354/13 EU:C:2014:2463
[94] Case C-303/06 *Coleman*, para 46 and Case C-354/13 *Kaltoft*, para 36.
[95] Case C-354/13 *Kaltoft, para 37.*
[96] Gerards, Janneke & Xenidis, Raphaele, *"Algorithmic Discrimination Europe: Challenges and opportunities for gender equality and non-discrimination law"*, p. 65.

The combination of the exhaustive list of protected grounds and the limits put by the CJEU risks giving a blunt answer to the questions of algorithmic discrimination, especially in relation to proxy discrimination. That being said, a broad and flexible interpretation of Article 21 could help better capture the specific types of discrimination that stems from the use of algorithms.

3.4 Discrimination defined in EU law

3.4.1 Direct discrimination

EU law defines direct discrimination as a situation in which "one person is treated less favourably than another is, has been or would be treated in a comparable situation" on the basis of one of the protected grounds defined in the already mentioned directives.[97] Direct discrimination aims at unfavourable treatment or differential treatment and captures situations in which a decision is made taking into consideration a protected ground, to the disadvantage of the person or group of persons related to that protected ground.[98] The concept adequately captures a wide array of situations of discrimination when performed by humans and constitutes a solid tool. Nevertheless, it has several weaknesses in the context of algorithmic discrimination.

One of the strengths of EU non-discrimination law is the irrelevance of intent. In relation to US discrimination law, where the notions of "motive" and "intent" are central to a finding of "disparate treatment".[99] Whether a protected ground was treated differently as a result of intention or not does not matter in EU law, which theoretically allows the concept of direct discrimination to capture a broad range of situations where protected grounds would be used as relevant variables by an algorithmic model even though discrimination was not the intention.[100]

[97] Barnard, Catherine, EU Employment Law, p. 277.
[98] Ibid., p. 277.
[99] McDonnel Douglas Corp. V. Green, 411 U.S. 792 (1973).
[100] Barnard, Catherine, *EU Employment Law*, p. 278.

Another strength of the concept is that it extends to situations where a person or a group of persons is treated unfavourably because she or he is associated with a protected group, without sharing the protected characteristic herself or himself. The CJEU has described this as discrimination by association in *Coleman*, where an employee was treated less favourably by her employer because she had to care for her sick child.[101] Despite the fact that she did not suffer from disabilities herself, the Court recognised that she had been harassed and directly discriminated against because of her relationship or association with her disabled child.[102] The CJEU reasoned that the principle of equal treatment enshrined in Directive 2000/78 in that are applies not to a particular category of person by reference to the grounds mentioned in Article 1.[103]

Particularly in the context algorithmic discrimination, this finding is important because it means that direct discrimination can potentially extend to some cases of algorithmic proxy discrimination and miscategorisation. As an example, in the present case, the concept of direct discrimination could cover cases of algorithmic profiling where users are classified within a protected category despite the fact that they do not share that characteristic themselves, but because of the proximity with the protected group. This sort of behavioural discrimination by association could happen where search and click data has been collected about an online user. For an example, the data reveals that the user is interested in whether a given restaurants are accessible to wheelchair users. Algorithmic profiling would result in the user being classified as disabled herself or himself, despite the fact the user is not but shares her or his life with someone who is. Such classification errors linked to the use of behavioural data as proxy in algorithmic profiling could result in scenarios where the user is denied given opportunities by because of misclassification. Therefore, the CJEU's approach in *Coleman* offers a valuable extension of the concept of direct discrimination in the context of algorithmic discrimination.

The concept of discrimination should arguably also extend to situations where individuals or groups are discriminated against because of a characteristic they are perceived to have, but that is not the case. Such a scenario differs from the discrimination by association since in this case it does not involve another individual who shares the protected characteristics, but

[101] Case C-303/06 *Coleman*, para 55. .
[102] Ibid, para 50.
[103] Ibid, para 38.

rather involves discriminatory ascriptions that are not founded in actual facts.[104] Put differently, a person should not need to share a protected characteristic to be recognised as a victim of direct discrimination based on that ground. Particularly in relation to algorithmic discrimination, situations of discrimination by perception, by ascription or by assumption are relevant since algorithmic profiling will result in ascribing given traits to people, while these inferences might not necessarily be correct.[105]

The European Commission has promoted such an interpretation by highlighting that the Equal Treatment directive and the Racial Equality Directive also prohibit a situation where a person is directly discriminated against on the basis of a wrongful perception or assumption of protected characteristics.[106] Nevertheless, this has not always been a promotion backed by the Court of Justice in cases of direct discrimination. In *Kaltoft*, a childminder was perceived as disabled by his employer because of his obesity, and, during a redundancy round, he was nominated for dismissal partly on that ground. [107] Despite the fact that the employee did not consider himself disabled, the employer's perception that the employee, impaired by his obesity, led to his dismissal. The CJEU, rather than recognising that it was the perception of obesity as a disability on the side of the employer that resulted in a differential treatment and discrimination, mandated the national court to find whether such an impairment existed in reality. The Court further stated that it would be the condition for obesity to count as a disability and ultimately for a finding of discrimination based on disability.[108]

This approach infers a restrictive application of the concept of direct discrimination in cases where protected grounds are perceived, ascribed or assumed. In particularly in relation to algorithmic profiling techniques it is problematic. If direct discrimination does not extend to such situations, it might very well be difficult to capture situations where algorithmically

[104] Barnard, Catherine, "EU Employment Law", p. 346.
[105] Hacker, Philipp, (2018), "*Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law*", p. 1151.
[106] European Commission, (2021), "*Report from the Commission to the European Parliament and the Council on the application of Council Directive 2000/43/EC implementing the principle of equal treatment between persons irrespective of racial or ethnic origin (the Racial Equality Directive") and of Council Directive 2000/78/EC establishing a general framework for equal treatment in employment and occupation (the Employment Equality Directive"*)", SWD(2021) 63 final (Brussels), p. 5.
[107] Case C-354/13 *Kaltoft*, para 24.
[108] Ibid, paras 62-64.

ascribed identities lead to the differential and disadvantageous treatment of individuals and groups.[109]

In addition to the uncertainty regarding the scope of direct discrimination, there are even bigger questions regarding its relevance in the context of algorithmic discrimination. As discussed in chapter 2, machine-learning algorithms are used to discover patterns in big datasets that combine a wide array of variables. These variables might be completely unrelated to protected grounds or may be proxies for protected categories. It has been argued that protected grounds themselves will not usually be used as inputs for such algorithms.[110] As a consequence, direct discrimination might be less likely to occur in algorithmic decision making when compared to "traditional" human decision making.[111] Another explanation for the unsuitability of the concept is that the treatment of data and its categorisation by algorithms might not be cognisable by human brains.[112] The categories and variables than an algorithm relies on might not mean anything to humans at all.[113] It would therefore be difficult to know whether they can be considered to stand for protected grounds. Moreover, the use of categories and variables of data in machine-learning algorithms in in constant evolution as the algorithm "learns". Since these categories are not static, but rather in constant motion, it would be difficult to know whether they relate to protected categories at given points in time. [114] That would require a thorough review of the algorithmic model and the way the statistical model used, treats the available data over time in order to find out whether unfavourable treatment arises.

Categorising algorithmic discrimination as direct discrimination is therefore likely to be a challenge given the opacity of algorithms, in particular due to the need to establish a comparator under current EU law. If the lack of transparency of the functioning of an algorithm prevents the gathering of evidence on how the algorithm as treated or would have

[109] Gerards, Janneke & Xenidis, Raphaele, *"Algorithmic Discrimination Europe: Challenges and opportunities for gender equality and non-discrimination law"*, p. 69.
[110] Hacker, Philipp, (2018), "*Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law*", p. 1152.
[111] Ibid., 1152.
[112] Leese, Matthias, (2014), "*The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union*", 45 Security Dialgoue, 494, p. 501.
[113] Ibid., p. 502.
[114] Ibid., p. 502.

treated a group that does not share the protected characteristic at stake, then a finding of direct discrimination might be precluded altogether.[115]

Moreover, the probability of direct discrimination occurring as unfavourable treatment might diminish in the context of algorithmic discrimination. Parts of the academic debate have suggested that, as awareness about relevant legal obligations increases, direct discrimination is likely to decrease in the developing phase of algorithms. [116] Furthermore, direct discrimination might diminish in the context of algorithms since the direct input of protected categories for decision making might yield lesser predictive accuracy.[117] Therefore, developers who are aware of these risks might remove protected categories from the available variables for algorithmic decision making in order to avoid direct discrimination completely. [118]

In conclusion, the concept of direct discrimination offers conceptual strengths given the lack of relevance of intent and its broad reach to situations of discrimination by association. Nevertheless, existing uncertainties regarding its applicability to cases of discrimination by proxy discrimination and ascription as well as doubts regarding the overlap between the law's static categorical approach to disadvantage and algorithms dynamic and mathematical approach to the categorisation of data might lead to a diminishing relevance and conceptual grasp of this central notion of non-discrimination law in the context of algorithmic discrimination.

3.4.2 Indirect discrimination

Another central notion of non-discrimination law is that of indirect discrimination. It is defined as situations "where an apparently neutral provision, criterion or practice would put members of a protected category 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".[119] Unlike direct discrimination,

[115] Kulk S & others (2020), "*Legal aspects of algorithms that make decisions. An exploratory study*", p. 5.
[116] Hacker, Philipp, (2018), "*Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law*", p. 1180.
[117] Ibid., p. 1181.
[118] Hacker, Philipp, (2018), "*Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law",* p. 1181.
[119] Eg. Article 2(2)(b) Directive 2000/78/EC.

which focus on the unfavourable treatment of given groups and individuals because of a given protected ground, the concept of indirect discrimination changes the focus on the disadvantageous effects of any given practice or measure.[120]

As for direct discrimination, the absence or presence of any intent to discriminate is irrelevant. Regardless of whether the developers of an algorithm, the company using an algorithm for commercial purposes, or the administration relying on an algorithm for decision-making purposes intend to discriminate, if they result in a disproportionately disadvantage of a protected group, it can be captured by the concept of indirect discrimination. This is the case even in situations of "masked" direct discrimination, or in other words, the concealing of direct discrimination through the use of a proxy for a given protected category in an algorithmic model.

In relation to the logic of algorithmic discrimination, the concept of indirect discrimination is better suited because instead of focusing on the treatment of various individuals based on their group membership, it shifts focus on the effect of any decision, policy or measure in terms of disadvantage experienced by protected groups. [121] Since algorithmic discrimination arises from the mining of large datasets, it concerns population groups who share common characteristics.[122] In theory, indirect discrimination offers a few basic features which might prove more effective than direct information in relation to algorithmic discrimination. Furthermore, since indirect discrimination aims at the discriminatory effects of algorithms rather than on their operations, it would offer a way a safe passage through the deep-sea areas that are the content of algorithms, the understanding of their operations, the chain of algorithmic decision making etc.

In addition, the concept of indirect discrimination as interpreted by the CJEU is capable of adequately addressing scenarios of proxy discrimination where decisions are made on the basis of characteristics related to, but different from, protected grounds. For example, in *CHEZ* the Court recognised that residency could be a proxy for ethnicity. The case concerned an area mostly populated by Roma people, which was excluded from accessing certain electricity services based on the racist stereotypes the company held against the Roma

[120] Barnard, Catherine, "*EU Employment Law*", p. 364.
[121] Xenidis, Raphael, (2020), "*Tuning EU equality law to algorithmic discrimination: Three pathways to resilience*", 27(6) Maastricht Journal of European and Comparative Law 736, p. 757.
[122] Barocas, Solon and Selbst, Andrew (2016), "*Big Data's Disparate Impact*", p. 768.

population of the area. [123] The company, explaining that it feared electricity theft in the area, installed electricity meters out of reach for the clients so they could not reach them to control their electricity consumption, a policy which it did not pursue in other areas of service provision.[124] The applicant in the case was not of Roma origin herself but the CJEU concluded that she was the victim of the racist stereotypes and ascriptions that targeted her neighbourhood.[125] The judgement shows that the concept of indirect discrimination is capable to address proxy discrimination, which is vital in the context of algorithmic discrimination.

Capturing situations such as those described in section XXX seems to provide some useful tools to the current EU law regime. Although the concept provides a safe harbour for tackling proxy discrimination when there is doubt as to whether the link between a proxy and protected ground is direct enough for the concept of direct discrimination to be applicable.[126] This is shown in *Jyske Finans*, where the CJEU considered whether a credit institution directly discriminated against a Danish citizen born in another country by asking him to provide additional proof to support his identity compared to native-born Danish citizens. With reference to the decision in *CHEZ*, the CJEU recognised that "the concept of ethnicity origins from the idea of societal groups marked in particular by common nationality, language, cultural and traditional origins and religious faith". [127] With emphasis on the word "in particular", the Court acknowledged that the list is not exhaustive and therefore it cannot be ruled out that a person's country of birth might be included among those criteria.[128] The Court established a link between the applicant's country of birth and his ethnicity. Nevertheless, it refused to consider the first ground as a proxy for the second, underlining that a person's country of birth is only of the specific factors which may justify the conclusion that a person is a member of an ethnic group and is not decisive in that regard.[129] This led to the Court to finding no differential treatment based on a proxy for ethnic origin.[130] By holding that a person's country of birth cannot, in itself, justify a general presumption that a person is a member of a given ethnic group such as to establish the existence of a direct link

[123] Judgment of 16 July 2015, '*CHEZ Razpredelenie Bulgaria' AD v Komisia za zashtita ot diskriminatsia* C-83/14 EU:C:2015:480.
[124] C-83/14 *CHEZ*, para 22.
[125] Ibid., para 82.
[126] Judgement of 6 April 2017, *Jyske Finans A/S v Ligebehandlingsnaevnet, acting on behalf of Ismar Huskic* C-668/15 EU:C:2017:278.
[127] Ibid., para 17.
[128] Ibid., para 18.
[129] C-668/15 *Jyske Finans*, para 25.
[130] Ibid., para 25.

between those two concept, the CJEU indicated a limited capacity of the notion of direct discrimination to tackle proxy discrimination.[131]

This is alarming, given the demonstrated examples of the use of residency data by algorithms as a way of inferring people's ethnicity. Nonetheless, the concept of indirect discrimination might apply in order to capture the group advantage at stake, although the CJEU considered that it was not applicable in *Jyske Finans*.[132] This was later confirmed by the Court in another case, *Maniero*. The case concerned the award of a scholarship for law students in Germany.[133] The scholarship was conditional on holding the German "First State Examination" in law and the applicant argued that this condition amounted to indirect discrimination on grounds of ethnic or racial origin since it had the effect of placing people of foreign ethnic origin with an equivalent diploma acquired abroad at a disadvantage.[134] Although the Court came to another conclusion, it confirmed that the concept of indirect discrimination is more well-suited to capture group disadvantages than the concept of direct discrimination.[135]

The notion of indirect discrimination is suitable to redress proxy discrimination even in situations where the wronged individual or group does not possess a protected characteristic. In *CHEZ*, the Court replicated the Coleman decision prohibiting discrimination by association in cases of indirect discrimination. Although the applicant in *CHEZ* was not of Roma origin herself, the CJEU underlined that the principle of non-discrimination is intended to also benefit persons who, although not themselves a member of the protected group concerned, suffer less favourable treatment or a particular disadvantage on one of those grounds.[136] Such an interpretation, which extends the scope of the concept of indirect discrimination to situations of discrimination by association, offers a potent tool in relation to redressing proxy discrimination arising from algorithmic discrimination.

---

[131] Ibid., para 33.
[132] Ibid, para 35.
[133] Judgment of 15 November 2018, *Heiko Jonny Maniero v Studienstiftung des deutschen Volkes eV* C-457/17 EU:C:2018:912.
[134] Ibid, para 21.
[135] Ibid, paras 45-50.
[136] C-83/14 *CHEZ*, paras, 49 and 56.

In conclusion, the concept of indirect discrimination might become more relevant in the context of algorithmic discrimination.[137] While it is relatively easy to filter out protected grounds to avoid direct discrimination, it might not be feasible to do so with their proxies so that discrimination might take place despite these precautions. Certainly, societal and structural discrimination may affect the operations of algorithms if the data is used to program or train them reflects biases and stereotypes that have developed into patterns of inequality over time.[138] If strategies to debias are not put in place and the data is not cleansed, it will reflect structural forms of inequality that originate from the institutionalism of past discrimination over the course of human history. If "infected" data is used by algorithms as training material, the patterns and structure of inequality engrained in that data will not only further be reproduced, but there is also a substantial risk that it will be amplified resulting in even deeper inequality.[139] The operation of algorithms, because of their reliance on biases social data increases the likeliness of occurrences indirect discrimination.

The solid doctrine developed by the CJEU around indirect discrimination could help address the issues and, from a substantive perspective, the notion has several advantages compared to that of direct discrimination. However, instrumentally it has been discussed that indirect discrimination could be understood as an instrumental devise that assists in the enforcement and expands the scope of direct discrimination.[140] For example, by helping to overcome problems of proving direct discrimination or to enable the selective protection of other groups than those explicitly captured by protected grounds of discrimination law.

However, the concept of indirect discrimination and its doctrinal application also entail practical difficulties. While a finding of direct discrimination excludes justification apart from a closed and restricted list of exceptions, establishing a case of indirect discrimination opens a wide array of possible justifications. [141] From the directives it follows that no indirect discrimination is to be found where the implicated provision, criterion or practice is objectively justified by a legitimate aim, and the means of achieving that aim are appropriate

---

[137] Hacker, Philipp, (2018), "*Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law",* p. 1175.
[138] Kim, Pauline, (2017)"*Data-Driven Discrimination at Work*", 58 William and Mary Law Review 857, p. 891.
[139] Cofone, Ignacio, (2019), "*Algorithmic Discrimination Is and Information problem*", 70(6) Hastings Law Journal 1389, p. 1398.
[140] Maliszewska-Nienartowicz, Justyna, "*Direct and Indirect Discrimination in European Union Law – How to Draw a Dividing Line?",* 111(1) International Journal of Social Sciences 40, p. 54.
[141] Barnard, Catherine, "*EU Employment Law*", p. 364.

and necessary.[142] Once a case of indirect discrimination has been established, the burden of proofs shifts to the defendant, which strengthens the position of the applicant.[143] Furthermore, the provision then offers the possibility for a defendant to invoke any justification. Thereafter, a court applies a probability test, the aim of which is to find out whether the existence of a disproportionate disadvantage can be justified by a measure serving a legitimate interest. Further conditions for that measure to be accepted as justification for an existing disadvantage are its objective, effective and proportionate contents, and its necessity, that the absence of any other measure that could fulfil the same aim and that would be less detrimental to the wronged individual or group. Hence, the grasp of the concept of indirect discrimination is not as stringent as that of the concept of direct discrimination.

In relation to algorithmic discrimination, the openness of the indirect discrimination test poses several problems. Firstly, it opens to a great number of justifications. Secondly, the legal certainty for potential victims decreases. A victim must show that a seemingly neutral rule, practice or decision disproportionately affects a protected characteristic and therefore is discriminatory.[144] However, indirect discrimination can remain hidden to both parties. For an example, suppose that somebody applies for a loan on a website of a bank. The bank uses an algorithmic system to decide on such requests. If the bank automatically denies a loan to a customer on its website, the customer does not why the loan was denied. Furthermore, the customer cannot see whether the bank's system denies loans to a disproportionate percentage of, for example, women. Even if customers knew that an algorithm rather than a bank employee decided, it would be difficult for them to discover whether the algorithm is discriminatory. [145]

Thirdly, developers and users of algorithmic systems will decrease as the appreciation of the validity of potential justifications would exclusively be bestowed upon courts. In particular, the application of the necessity part of the proportionality test by courts poses questions

[142] Article 2(2)(b) Directive 2000/43/EC; Article 2(2)(b) Directive 2000/78/EC; Article 2(b) Directive 2004/113/EC; Article 2(1)(b) Directive 2006/54/EC.
[143] Article 8(1) of Directive 2000/43/EC, Article 10(1) of Directive 20000/78/EC, Article 9(1) of Directive 2004/113/EC and Article 19(1) of Directive 2006/54/EC.
[144] ECtHR, D.H. and Others v. Czech Republic (Grand Chamber), No. 57325/00, 13 November 2007, paras. 187–8.
[145] Larson, Jeff and others, 'These Are the Job Ads You Can't See on Facebook If You're Older', (The New York Times, December 20, 2017), < https://www.nytimes.com/2017/12/20/business/facebook-job-ads.html > accessed 25 May 2021.

considering the trade-off between accuracy and performance on the one hand and non-discrimination on the other that might arise in cases of algorithmic discrimination.

Finally, the distinction between direct and indirect discrimination might become increasingly blurred in cases of algorithmic discrimination. [146] It has been debated that, on the one hand, doubts might arise regarding the possibility of directly relating proxy data to any protected grounds therefore putting in jeopardy the application of the notion of differential treatment based on a protected ground. On the other hand, the doctrinal distinction will be undermined if cases of algorithmic discrimination fall within the indirect discrimination category for a lack of transparency on the input data used by the algorithmic model. [147]

3.5 Algorithmic discrimination and GDPR

Discussions between academics have drawn attention to potential role of data protection law and legal obligations related to privacy in the prevention of algorithmic discrimination.[148] In summary, the rationale is that certain categories of data, such as gender, race, religion etc., are particularly sensitive because they can easily lead to unlawful discrimination if processed without any precautions. This is reflected in the GDPR, which "identifies special categories of personal data or sensitive data.[149] The GDPR recognises that "personal data which are, by their nature, particularly sensitive in relation to fundamental rights and freedoms merit specific protection as the context of their processing could create significant risks to the fundamental rights and freedoms".[150] Especially recital 71 of the GDPR indicates that "in order to ensure fair and transparent processing in respect of the data subject the data controller should prevent, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or processing that results in measures having such an

[146] Senden, Linda & Xenidis, Raphael, "*EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination*", in Ulf Bernitz et al (eds), *General Principles of EU law and the EU Digital Order*, Kluwer Law International, 2020, p. 19.
[147] Xenidis, Raphael, (2020), "*Tuning EU equality law to algorithmic discrimination: Three pathways to resilience*", p. 751.
[148] Henderson, Tristan, "*Does the GDPR Help or Hinder Fair Algorithmic Decision-Making?"* (2017), LLM dissertation, Innovation, Technology & The Law, University of Edinburgh, 2017, p. 17.
[149] Recital 10 of Regulation 2016/679 of 27 April 2916 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 Protecton Regulation) (2016) OJ L 119/1.
[150] Recital 51 of the General Data Protection Regulation (GDPR).

effect". This embodies the interaction between data protection law and non-discrimination law. However, the interaction poses a number of questions as the provisions do not neatly overlap.

The list of categories of data the processing of which could give risk to discrimination does not neatly fit with the lsit of protected grounds under EU non-discrimination law and EU gender equality law. Most importantly, the issue of gender equality or sex discrimination is completely absent from the GDPR and neither gender nor sex are mentioned as sensitive categories of personal data. Racial or ethnic origin, religion or belief and sexual orientation are explicitly mentioned both in relation to discrimination (recital 71) and in relation to the prohibition of processing such data, but the recital does not refer to sex or grounds such as age and disability. Similarly, Article 9(1) GDPR prohibits the "processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation".[151] Although the list is much broader compared to the list in Article 19 TFEU on the prohibition of discrimination, it does not explicitly mention sex, disability and age. I t might be inferred that disability is understood to be included in the terms of health status or data concerning health but age and sex, in the form of protected grounds, are more difficult to read. The absence of an outright prohibition on processing such data categories could be justified by several reasons pertaining to the possibilities for legitimate and non-discriminatory use of the data.

The approach taken by EU data protection law, and the GDPR in particular, towards preventing discrimination differs from that of non-discrimination law and pertains to the degree of automation of the processing of sensitive data. The GDPR considers the presence of a human in the loop as a form of preventive safeguard.[152]

In spite of the recital, there is no clarity concerning the form of human supervision involved. This is problematic in view of the existence of human bias. Prohibiting full automation does not ensure the absence of discrimination.[153] Furthermore, Article 5(1) GDPR provides some

[151] Recital 51, GDPR.
[152] Recital 71, GDPR.
[153] Henderson, Tristan, "*Does the GDPR Help or Hinder Fair Algorithmic Decision-Making*? (2017), p. 28.

clarity by stating that the principles underpinning the processing of personal data are "lawfulness, fairness and transparency". In addition, Article 5(2) mentions the principle of accountability. By comparison, the notion of "discrimination" is only mentioned three times to describe the risks posed by the processing of sensitive data.[154] By contrast, "equality" is only mentioned twice in relation to processing data in the context of employment.[155]

As a result, the concepts on which the GDOR relies in relation to the issue of algorithmic discrimination are different from those to non-discrimination law and gender equality and the link between the two areas is not made explicit by the framework. Nevertheless, the approach taken by the GDPR to sensitive data offers some guarantees regarding some of the protected grounds covered by non-discrimination law. But there are also evidence gaps, especially in relation to the protection of gender equality. Despite different conceptual approaches to the question of algorithmic discrimination, EU data protection law with the GDPR as a spearhead can provide important tools to non-discrimination law and gender equality within the EU.

3.6 Summary

Non-discrimination has become a core tenet of EU market law. It is not only a fundamental right in EU, but also seen as a principle governing market interaction. There are several directives that together forms a framework, which establishes legally protected grounds of discrimination. In all of the directives, the core is the distinction between direct and indirect discrimination and the different criteria that must be met to justify them, respectively.
In short, EU non-discrimination and equality law show a limited capacity to deal with algorithmic discrimination. Algorithmic discrimination raises questions in relation to the exhaustive list of protected grounds that form the substance of the personal scope of the law. Because of the proxies and correlation challenge, it is questionable whether this is sufficient for protection against algorithmic discrimination. Since algorithms change the very nature of discrimination, they increase the risk of miscategorization based on user's data as well as the risk of intersectional discrimination. It will be rare for an algorithm to discriminate only or decisively on the basis of a protected ground, since it will usually base its data input on an array of different factors and variables.

[154] Recitals, 71, 75 and 85, GDPR.
[155] Recital 155 and Article 88, GDPR.

Above all, this raises the issue of whether a protected characteristic such as gender can sufficiently be defined as to cover a range of proxy grounds. Moreover, in many cases the output of an algorithm will be based on a combination of characteristics and behaviour that is unique to a particular group of persons. Consequently, the limited focus on a few protected grounds and the lack of recognition of notions of multiple and intersectional discrimination in the current legislation means that such cases of highly differentiated discrimination cannot be effectively redressed.

The issue of algorithmic discrimination does not fit the traditional concepts of EU non-discrimination law. In the literature on algorithmic discrimination, it has been pointed out that direct discrimination is unlikely to arise because it is improbable that designers of algorithmic systems directly input protected categories as negative variables in supervised algorithmic models.[156] Scholars have also claimed that biases in the processing of data would lead to indirect rather than direct discrimination.[157] Therefore, the concept of indirect discrimination would be a better fit for algorithmic discrimination. The notion of direct discrimination might prove to be potent against discriminatory wrongs of algorithms. However, that might increase legal uncertainty given various objective justifications.

Some scholars have shown that algorithmic decision-making has revealed patterns of structural discrimination that are based on existing ineqauli9ties related to characteristics such as health status, education, income and socioeconomic status. This is particularly problematic since these characteristics are not protected under non-discrimination law.[158] Furthermore, it has been shown that algorithms may discriminate on the basis of seemingly irrelevant characteristics such as being a cat-owner, simply because there is a correlation to be made between owning a cat and being open to certain forms of advertising (this will be explained further in chapter 4).[159] Consequently, it raises the question of whether the increasing use of algorithms may create new forms and types of discrimination.[160]

---

[156] Hacker, Philipp, (2018), "*Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law",* p. 1152.
[157] Senden, Linda & Xenidis, Raphael, "*EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination*", p. 22.
[158] Wachter, Sandra, (2020), "*Affinity profiling and discrimination by association in online behavioural advertising*" Berkely Technology Law 35(2), p. 54..
[159] Ibid., p. 54.
[160] Ibid., p. 60.

4 Principles on evidence to algorithmic discrimination

Traditional legal principles rely on two cornerstones – effective access to facts and personal accountability.[161] Historically, accountability requires agency, but the growing number of discriminatory acts caused by algorithmic systems has triggered the emergence of a discussion about AI agency.[162] The topic of AI agency is complex and triggers a vast array of interesting topics. On the one hand, it is wrong to reduce AI systems as mere commodities with no regard to their human-like abilities. But on the other hand, algorithmic systems are not capable of being personally accountable.

4.1 The presumption of human agency

In the context of new technologies, agency is defined in reference to the legal status of new technologies. Algorithmic systems and ML systems are defined as objects.[163] As presented above, they are viewed as something different than computers and washing machines since they are capable to learn, adapt, react, and adjust to, external stimuli but a lot less than humans. For example, in criminal liability, damage caused by autonomous systems is by default attributed to the user or the programmer of the system since proof of an intent to harm is necessary in order to establish liability. Although autonomous algorithmic decisions cannot be considered as intentional *per se*, the adduced evidence will result in designating the human in command as accountable.[164] These tests are easily bypassed when black-box algorithmic systems are the actor or decision-maker.[165]

Provided that damage occurs as the result of an algorithmic decision, the so-called respondent superior principle is applied. It makes it possible for a principal to be held accountable for the conduct of an object or an agent he or she employs.[166] The starting point of the doctrine is that a principal's use of an agent entails a general command under which the agent or the

[161] Craig, Paul & De Búrca, Gráinne, *"EU Law: Text, Cases, and Materials*", p. 239.
[162] Bathaee, Yavar, "*The Artificial Intelligence black box and the failure of intent and causation*", p. 893.
[163] Ibid., p. 897.
[164] SOU 2018:16, p. 554.
[165] Bathaee, Yavar, "*The Artificial Intelligence black box and the failure of intent and causation*", p. 919.
[166] Craig, Paul & De Búrca, Gráinne, *"EU Law: Text, Cases, and Materials*", p. 938.

object accomplish the principal's preassigned goals.[167] In cases where the general command is not manifested, it is assumed.

Contrary to the evidentiary requirements in criminal liability, the system of evidence in EU non-discrimination law, does not include proof of an intent to discriminate. Regardless of a bias was intended or not, is in fact, of no importance.[168] Proof of discrimination is based on the effects of an action. The issue of agency being is irrelevant and the evidence as such, inadmissible. However, such a system of evidence implicitly relies on the human agent postulate because discrimination has never been viewed as being caused by non-human systems.[169] The issue of who the accountable agent is, or should be, has not until recent years with the evolution of autonomous vehicles, created a specific evidentiary debate to take place. For an example, in the field of labour law, when actions are brought against an employer, victims are usually able to identify the agent having introduced a difference in treatment in a recruitment practice or wage system. Therefore, if a job applicant brought a claim of discriminatory online recruitment, they would designate the legal or natural person acting as user of an HR algorithm as respondent. Consequently, disregarding the possibility that the unfair bias may have occurred as a result of the algorithm's autonomous decision.

From a perspective of classic agency principles, this leads to that AI users might be allowed to avoid being held accountable in cases of unsupervised algorithmic discrimination. To recognize that autonomous algorithmic systems are agents, is to extend the concept of agency to artificial systems and to abandon the concept of human agency. Nevertheless, from the perspective of fairness in legal evidence theory, the answer might be different. Personal accountability should not be presumed based on a mere use of AI. Such users should at least be given the possibility to argue that such a discriminatory act was the result of a non-transparent algorithmic decision-making process. In the view of exercising their right to an effective defence, respondents should be given the possibility to rebut the presumption of human agency, by proving that they did not, personally, cause a given damageable act.

[167] Ibid., p. 934.
[168] Craig, Paul & De Búrca, Gráinne, "*EU Law: Text, Cases, and Materials*", p. 937.
[169] Gerards, Janneke & Xenidis, Raphaele, *"Algorithmic Discrimination Europe: Challenges and opportunities for gender equality and non-discrimination law"*, p. 70.

4.2 The feasability of evidence

The principle of feasibility expresses an overall standard of reasonableness in legal evidence theory by providing that the procedural requirements for the parties should not be irrational or unrealistic. Parties in a trial should not be held to adduce evidence that in practice is difficult or impossible to access and adduce. Furthermore, during pleadings, the parties should not be expected to meet unrealistic evidentiary requirements. As mentioned above, algorithms are evidentially challenging due to the fact that their decisions may be unknown to the humans in command. In a situation where both victims and respondents run the risk of being left without the right to effectively bring their claims before a competent authority.

4.2.1 The system of evidendce in EU non-discrimination law

The non-discrimination directives regulate evidence of discrimination by placing the initial burden of proof on the victims.[170] The claimants are required to establish the basic facts from which discrimination may be presumed. Once such a presumption is established, the burden of proof shifts to the defendant.[171]

In general, it could be said that the system of evidence in EU non-discrimination law reflects two ground pillars - effectiveness and fairness.[172] Fairness underlies the allocation of the burden of proof between claimantsaand defendants.[173] Through the lens of the non-discrimination directives, the burden does not shift to the defendant based on a mere assertion of discrimination by the claimant. It is rather the latter who is required to support their claim with evidence. Although, neither direct nor indirect discrimination are presumed lightly, but must be plausible proven.[174]

Effectiveness is expressed in the right to access evidence, which plays a crucial role, especially in indirect discrimination cases. Unlike in the case of direct discrimination, which many often be self-evident, indirect discrimination is more challenging to prove.[175] In the

[170] Article 10 of Directive 2000/78, Article 8 of Directive 2000/43 and Article 19 of Directive 2006/54.
[171] Craig, Paul & De Búrca, Gráinne, "*EU Law: Text, Cases, and Materials*", p. 959.
[172] Ibid, p. 955.
[173] Ibid, p. 959.
[174] Ibid, p. 960.
[175] Opinion of AG Lenz, Case C-109/88, *Danfoss*, EU:C:1989:228, para 30.

case of *Danfoss*, the AG pointed out that in indirect discrimination cases, the plaintiff's burden is "made considerably more difficult because she has to show that a neutral criterion which is applied in like manner to men and women is in practice in the majority of cases satisfied by women and they as a sex are therefore disadvantaged."[176]

The right to access evidence does seem to be the norm under the EU non-discrimination law, but that is not the case regarding job applicants. Recruiters do not seem to have an obligation to facilitate the access to facts for applicants who suspect they have undergone discriminatory recruitment processes.[177] In general, victims of indirect discrimination should benefit from a right to access evidence and may rely on EU non-discrimination law in order to lift any restrictions on this right, stemming from national regulations or practices.[178] However, it should be noted that under the non-discrimination directives the procedural aspects pertaining to the ways in which evidence is presented and assessed remain unclear. In the case of *Belov*, the referring court asked the ECJ if the facts established by the claimant must allow a conclusion that there has been discrimination or whether the mere presumption in that there has been discrimination is sufficient.[179] The provisions in those Directives do not define standard of proof or the principles of evidence assessment. This is due to fact that the main procedural competence under those Directives lies with the Member States.[180]

Useful guidance can be drawn from the case law of the ECJ. On the point of admissibility, the court seem to allow a broad definition of the category of evidence able to support and rebut a presumption of discrimination.[181] Regarding the standard of proof, the norm in EU non-discrimination law seems to be more similar to civil law, as opposed to criminal law's rule of beyond reasonable doubt. This is understandable because claimants are required to prove discrimination plausibly – not conclusively.[182] Put in other words, the more a fact is

---

[176] Ibid, para 30.
[177] Köchling, Alina & Wehner, Marius, "*Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development*", Bus Res 13, 795–848 (2020), p. 801.
[178] See Directive 2000/78, Art. 10(1); Directive 2000/43, Art. 8(1); Directive 2006/54, Art. 19(1).
[179] Judgement of 31th of January 2013 *Valeri Hariev Belov v CHEZ Elektro Balgaria AD* Case C-394/11, EU:C:2012:585.
[180] See Directive 2000/78, Art. 10(1); Directive 2000/43, Art. 8(1); Directive 2006/54, Art. 19(1)´
[181] Judgement of the Court of 23 October 2003 *Hilde Schönheit v Stadt Frankfurt am Main (C-4/02) and Silvia Becker v Land Hessen (C-5/02)* Joined Cases C-4 & 5/02, EU:C:2003:583, para 72. Köchling, Alina & Wehner, Marius, "*Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development*", p. 836.
[182] Köchling, Alina & Wehner, Marius, "*Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development*", p. 836.

undisputed, the less additional evidence will be required. Therefore, the best evidence of a party are so-called notorious facts.[183] These types of facts are more frequent in cases of direct discrimination. For example, in *Feryn*, declarations by which an employer openly expressed bias against foreigners in his recruitment policy were considered to support a presumption of discrimination.[184]

In the absence of undisputed facts, which is more usually the case in cases of indirect discrimination, a claimant may include facts that reveal so called residual discriminatory biases.[185] This can be achieved through statistical evidence or situation testing. In the case of *Danfoss*, AG Lenz considered that if an employee has no access to the facts required to prove indirect discrimination "a rule of evidence applies to the effect that on proof of a lower average wage for women with a representative group of employees there is a presumption of discrimination".[186]

Concerning the principles on the assessment of evidence in the non-discrimination directives, it may be drawn from the ECJ case law that national authorities should never apply procedural rules that discharge a national authority from evaluating the evidence presented in support of an allegation. In the *Otero Ramos* case, a nurse working in the emergency unit of a Spanish hospital requested a medical certificate stating that her activities presented a risk to her child whom she was breastfeeding.[187] A certificate was requested for the purpose of benefitting from financial allowance in respect of risks during breastfeeding. The competent authority took into account a statement issued by the hospital's HR director, certifying that the nurse's job was risk-free as well as a report issued by a doctor stating that the nurse was fit to carry out her work. Based on these reports it was concluded that the nurse's work did not pose a risk for the breastfeeding of her child and her request for financial allowance was rejected. The decision was contested and eventually brought before the ECJ.[188]

---

[183] Ibid., p. 838.
184 Judgement of 10th July 2008, *Centrum voor gelijkheid van kansen en voor racismebestrijding v Firma Feryn NV*, Case C-54/07 EU:C:2008:397.
[185] Judgment of 17th October 1989, *Handels- og Kontorfunktionærernes Forbund I Danmark v Dansk Arbejdsgiverforening, acting on behalf of Danfoss,* C-109/88 EU:C:1989:383, para 9.
[186] Opinion of AG Lenz, Case 109/88, Danfoss, para 52.
[187] Judgement of 19th October 2017, *Elda Otero Ramos v Servicio Galego de Saude and Institutio Nacional de la Seguridad Social*, Case C-531-15 EU:C:2017:789.
[188] Ibid., para 28.

From the perspective of evidence, the main issue in *Otero Ramos* was whether a failure from the authorities to assess the occupational risks for a worker, who is breastfeeding, in accordance with Article 4(1) of Directive 92/85[189], could constitute discrimination on the ground of sex, under Directive 2006/54.[190] Regarding this point, the court held that by virtue of Directive 92/85, the risk assessment of the work carried out by a breastfeeding worker should not rely entirely on reports provided by the defendant, but must also include a specific assessment taking into account the worker's individual situation.[191] In conclusion, the Court stated that there was a less favourable treatment, qualified as direct discrimination on the grounds of sex, of a female workeradue to her being a breastfeeding woman, insofar as the authorities considered themselves discharged from having to assess the claimant's allegations, by classifying, the report on the absence of risk as conclusive evidence. However, it was left for the referring court to ascertain whether the defendant had failed to conduct the risk assessment in accordance with the requirements laid out in Directives 92/85 and 2006/54.[192]

The judgement confirms the ECJ's general tendency to define the criteria for the assessment of evidence under Union law, while leaving it to the referring court to apply them in a specific case. From the perspective of the effectiveness, the case arguably sets out a general rule of evidence assessment. When an allegationais brought before a competent authority on the grounds of EU law, the evidence in support of this allegation should always be examined and assessed, and any national rules that create a discharge from such assessment should be disapplied.

The mentioned cases shows that the effective application of the non-discrimination Directives depends on the victim's ability to cast a wide net on the indicia they may present in support of their claims, having unrestricted access to thoseaindicia and realistic standards of proof. It appears, however, that despite the protective attitude of the ECJ towards victims, discrimination in workers recruitment involves a notable exception. While employers are held to ensure access to evidence if discrimination in favour of their employees, recruiters do not

[189] Council Directive 92/85 on the introduction of measures to encourage improvements in the safety and health at work of pregnant workers and workers who have recently given birth or are breastfeeding, O.J. 1992, L 348/1.
[190] Ibid,., para 2.
[191] Ibid., para 77.
[192] Ibid., para. 76.

seem to be bound by the same requirements in relation to job applicants. This is illustrated through the cases of *Danfoss* and *Meister*.[193]

4.2.2 Limits of feasability principle in cases of recruitment

In the case of Danfoss, the criteria used by an employer to supplement workers basic pay were opaque. This made it difficult for female workers to identify the reasons underlying the difference in salary between them and their male colleagues performing the same tasks.[194] The ECJ held that it was for the employer to prove that their practice regarding wages was not discriminatory, since the female workers were unable to provide evidence and build a prima facie case of discrimination.[195] Therefore, the defendant was forced to make his system of pay transparent. [196] The Danfoss case reinforces the case law of *Bilka.-Kaufhaus* and *Enderby*, in which the ECJ stated that the there may be a shift in the onus when called upon in order to avoid depriving workers who appear to be the victims of discrimination of any effective means of enforcing the principle of equal pay.[197]

The *Meister* case provides the scenario of recruiters. In the case a third-country national applied for three job positions in the Union but did not get called for interview. She requested information on the criteria used to shortlist the applicants, which the employer refused to share. The claimant then argued discrimination on the grounds of racial and ethnic background before the court. However, the German courts considered that she did not adduce enough evidence to support the presumption of discrimination, although they stated that she had clearly suffered a less favourable treatment than other persons in a comparable situation.[198]

Similarly, to *Danfoss*, the opacity in *Meister* constituted an obstacle for the claimant since it prevented her from adducing evidence supporting a presumption of discrimination. Nevertheless, in the latter case, the ECJ considered that the non-discrimination directives did

---

[193] Case C-109/88, *Danfoss*, and judgement of 19th April 2012, *Galina Meister v Speechg Design Carrier Systems GmbH*, Case C415/10, EU:C:2012:217.
[194] Case C-109/88, *Danfoss*, para 13.
[195] Ibid., para 15.
[196] Judgement of 13th May 1986, *Bilka – Kaufhau8s GmbH v Karin Weber von Hartz*, Case C-278/84 EU:C:1986:204; judgement of 27th October, *Dr. Pamela Mary Enderby v Frenchay Health Authority and Secretary of State for Health*, Case C-127/92 EU:C:1993:859.
[197] Case C-127/92 *Enderby*, para 14.
[198] Case C-415/10 *Meister*, para 13.

not entitle a worker who claims plausibly that he meets the requirements listed in a job advertisement to have access to information indicating whether the employer had engaged another applicant at the end of the recruitment process.[199] The court also stated that it cannot be ruled out that a defendant's refusal to grant access to information may be on of the factors to take into account in the context of establishing facts from which it may be presumed that there has been direct or indirect discrimination.[200] In conclusion, a job applicant is not entitled under EU law to require from a recruiter access to facts able to support their claim.

From a practical perspective, the ECJ's ruling in *Meister* is reasonable. Provided that recruitment, algorithmic or not, is a biased process, proving that a recruiter used a skill-bias as a cover-up for a discriminatory selection of job applicant is difficult. If the approach of *Danfoss* is applied to workers recruitment, frustrated and rejected job applicants risk posing a huge burden to recruiters through an ocean of claims demanding transparency, even though they were perfectly skill-based.

However, from the perspective of fairness and effectiveness, *Meister* shows that in the absence of an effective right for victims to access facts able to reveal that a recruitment process was biased, indirect discrimination in such processes is difficult to prove. Rejected candidates seem to be left with the recruiter's conduct as a sole indication of possible discrimination.

Both *Danfoss* and *Meister* concern cases of human conduct. However, if those procedures were automated the evidentiary difficulties in proving discrimination would be even greater.[201] When a recruitment algorithm is charged with hiring job applicants there is no human conduct per se that could hint at the presence of unfair bias, and recruiters who use the algorithm may, themselves, not be able to plausibly explain how the bias occurred. [202] In such a case, the obstruction to evidence, for the claimant as well as for the defendant, seems to be absolute. The implication being that algorithmic discrimination falls altogether outside the scope of application of the non-discrimination directives.

---

[199] Ibid., para 46.
[200] Ibid., para 47.
[201] Gerards, Janneke & Xenidis, Raphaele, *"Algorithmic Discrimination Europe: Challenges and opportunities for gender equality and non-discrimination law"*, p. 74.
[202] Ibid., para 47.

5. Effectiveness and fairness: how to rebalance the procedural rights of the parties

5.1 Claimants – a right to transparency

A claimant's access to evidence in e-recruitment cases can be made more effective in two different ways. On the one hand, redefining the allocation of the burden of proof or, on the other hand, redefining the modalities of adducing evidence.

In the White Paper on AI, it was presented that it is possible to draw inspiration from principles on strict liability and argue that algorithmic recruitment always creates an opportunity for discrimination.[203] However, it has also been discussed, assuming that AI presents a risk of discrimination would lead to a lowering of the standard of proof for claimants in the sense that in cases dealing with algorithms, a general presumption of algorithmic discrimination would apply.[204] Such a presumption would not comply with the current system of evidence in EU non-discrimination law, under which the onus on the victims is lessened but not removed.[205] Furthermore, a risk-based approach would make it easier for job applicants to claim that they have been discriminated against, even if they were not.

A second solution would be to extend the case law in *Danfoss* to job applicants who underwent algorithmic recruitment. Unlike, a presumption of discrimination, it would not lead to a removal of the onus on the victims. Instead, it implies a revision of the modalities according to which evidence of algorithmic discrimination is adduced, with regard to access to indicia.

A right to access evidence would be the result of a joint reading of EU non-discrimination law and the GDPR. Firstly, the non-discrimination directives complement and the GDPR complement each other. The safeguards of the GDPR on automated decisions provide the basis for the recognition of a right to transparency, which in the light of the non-

[203] White Paper, "*On Artificial Intelligence – A European approach to excellence and trust*", COM (2020)65 final, p. 18.
[204] Bathaee, Yavar, "*The Artificial Intelligence black box and the failure of intent and causation*", p. 928.
[205] Case C-415/10 *Meister*, para 22.

discrimination directives, would result in a right to access evidence in favour of victims of algorithmic discrimination.[206] Provided that such a right would be recognized, the practical procedural aspects related to a victim's provision of evidence would be defined in reference to the requirements set out in the system of evidence in EU non-discrimination law, as well as in the relevant ECJ case law.

Secondly, the GDPR and the non-discrimination directives share similar objectives. Under the GDPR, data controllers are required to handle personal data that takes account of the potential risks involved for the interests and rights of the data subject and that prevents discriminatory effects on natural persons on the basis of the grounds similar to those of the non-discrimination directives. [207] Furthermore, Article 9(1) GDPR sets out a general prohibition on processing personal data revealing, for example, racial or ethnic origin or political opinions. The prohibition converges with the prohibitions set out in Article 1 of Directive 2000/78 and Article 21(1) EUCFR.[208]

Under the GDPR, e-recruitment may be classified as an individual automated decision. [209] For such decisions, Article 22(1) GDPR establishes a right to object for the data subject, for example a job applicant. However, there are a limited number of cases in which the provisions does not apply. These include automatedadecisions that are "necessary for entering into", or "performance of", a contract between the data subject and a data controller.[210] The application of this article is subject to debate and it is not clear how "contractual necessity" should be interpreted".[211] Nevertheless, assuming that the selection of job applicants is a necessary pre-contractual phase to entering into an employment, algorithmic recruitment would fall outside the scope of application of Article 22(1) GDPR. The data controller – recruiter, is then held to implement suitable measures to safeguard the data subject's rights, including the right to obtain human intervention on the part of the

[206] Article 15 GDPR.
[207] Recital 71, GDPR.
[208] Hacker, Philipp, (2018), "*Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law*", p. 1182.
[209] Edwards, Lilian & Veale, Michael, "*Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling*", 34 Computer Law & Security Review (2018), 398–404, p. 400.
[210] Article 22(2)(a) GDPR.
[211] Mendoza, Isak and Bygrave, Lee A., "*The Right Not to Be Subject to Automated Decisions Based on Profiling*" (May 8, 2017), EU Internet Law: Regulation and Enforcement (Springer, 2017, Forthcoming), University of Oslo Faculty of Law Research Paper No. 2017-20, p. 15.

controller, to express her or his view and to contest the decision.[212] In the contest of this paper, after having gone through a recruitment process, a job applicant could require intervention from the recruiter, demand their opinion and possibly contest the decision.

Article 22(4) GDPR establishes another important requirement regarding the lawfulness of individual automated decisions not covered by the first paragraph. The decisions should not be based on the categories of personal data outlined in Article 9(1) GDPR. Instead, the exceptions to this prohibition include several cases where explicit consent had been given by the data subject.[213] Firstly, cases where the data subject had given explicit consent. Secondly, if the processing is necessary for carrying out obligations and exercising specific rights of the controller or of the data subject in the field of employment and social security and social protection law in so far as it is authorised by Union or Member State law or a collective agreement pursuant to Member State law providing for appropriate safeguards for the fundamental rights and the interests of the data subject.[214] It can be drawn from these provision that, in the context of e-recruitment, an employer would not be allowed without the consent of the job applicants to include biases based on protected characteristics, since this would not be authorized by Union law. In particular, provisions which prohibit discrimination in the context of employment[215], access to employment[216] and of data processing.[217]

In relation to recruiters using algorithms to screen job applicants, Article 22 GDPR creates three obligations. First, an obligation to inform the job candidates of the nature of the recruitment and provide an explanation of the functionality of the HR algorithm.[218] Second, an obligation of explanation and human intervention, aimed at revealing the rationale behind a specific decision.[219] Third, an obligation to process personal data in compliance with EU law, which includes non-discrimination principles.

---

[212] Article 22(3) GDPR.
[213] Article 9(2)(a) GDPR.
[214] Article. 9(2) GDPR.
[215] Article. 1, Directive 2000/78.
[216] Article. 1(a), Directive 2006/54.
[217] Article. 9(1), GDPR.
Wachter, Sandra, Mittelstadt, Brent, & Floridi, Luciano, "*Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation*". (2017) International Data Privacy Law, 7(2), 76–99, p. 79.
[219] Ibid., p. 79.

From the perspective of evidence, these requirements are beneficial for claimants. Provided that a job applicant suspected discriminatory algorithmic recruitment, they could require an explanation from the employer under Article 22(3) GDPR, and based on the explanation, claim non-compliance of the recruiter – or the algorithm – with the prohibition set out in Article 22(4) GDPR. In conclusion, through the application of the safeguards on automated decisions of the GDPR, a job applicant may force a recruiter to make the recruitment process equally transparent as for employers under the case of *Danfoss*.

Nevertheless, it is yet to be clarified how this right should be applied in practice.[220] Despite the lack of clarity in the context of algorithmic discrimination, some conclusions can still be made. It seems probable that a claimant can rely on the safeguards of the GDPR applied to automated decisions to gain access to indicia of discrimination and demand that the veil of opacity shall be lifted. In this sense, the GDPR may serve as a basis for a right to access evidence whenever alleged discrimination is the result of automated data processing. However, the scope and exercise of the right to transparency will still largely depend on the availability of facts.[221] It is reasonable to assume that a recruiter's duty of human explanation would consist of facts within reach, such as shortlisted candidates and the selection criteria fed into the algorithm. Ideally, an unfair bias could be detected on scrutiny of the features of the shortlisted candidates. This was how the preference for male workers in Amazon's recruitment system was revealed.[222]

Regarding the human explanation, it has been discussed that the human explanation in algorithmic recruitment may include information on the functionalities of the algorithm.[223] Especially in cases when biases cannot be inferred from the mere disclosure of a shortlist of candidates. In the case of Amazon, in order to find out why the algorithm discriminated against women (contrary to the neutral selection criteria initially fed in) it was a closer

[220] Edwards, Lilian & Veale, Michael, "*Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling*", pp. 402-403.
[221] Wachter, Sandra, Mittelstadt, Brent, & Floridi, Luciano, "*Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation*", p. 88.
[222] Jeffrey Dastin, "*Amazon scraps secret Ai recruiting tool that showed bias against women*", (Reuters, 11 October 2018), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G <accessed 24 May 2021>.
[223] Wachter, Sandra, Mittelstadt, Brent, & Floridi, Luciano, "*Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation*", p. 79.

inspection of algorithms functionalities that revealed the latter's systematic disregard of applicants mentioning the word "women". [224]

In conclusion, while it is difficult to know for sure how the right to transparency will entail when applied in EU non-discrimination law, it is possible to speculate that the starting point of human explanation will be the disclosure of the shortlisted of job candidates, presented by the algorithm. In such cases where discrimination is less evident, further proof may be needed. The human explanation may then include disclosure of the employer's recruitment criteria, recruitment history and the functionalities of the algorithm.

If the ECJ would recognize a right to transparency through a joint reading of the non-discrimination directives and the GDPR, victims of algorithmic discrimination would have a procedural possibility to present a case of algorithmic discrimination. However, the disadvantage of such right would be the overburdening of recruiters who have used algorithms in their recruitment processes. As users of such algorithms, recruiters would carry a large share of the evidentiary onus since they would need to lift the veil of opacity. Not only so that claimants can make their case, but also because they themselves can prepare their defence. Consequently, respondents should also benefit from a procedural right that would allow them to participate in a debate on evidence of algorithmic discrimination. Such a right could, for an example, could include a rebuttal of the presumption of human agency.

5.2 Respondens – grounds of defence

Without the possibility to rebut the presumption of human agency, a respondent would by default be held accountable for an algorithmic discriminatory bias and would only be able to present arguments that are based on the assessment of legitimacy, necessity and proportionality. Drawing on the case law of the ECJ, the respondent would be required to have knowledge of the rationale behind a given difference in treatment. In the case of *Abercrombie & Fitch Italia*, the Italian government was able to justify the application of an

[224] Jeffrey Dastin, "*Amazon scraps secret Ai recruiting tool that showed bias against women*", <accessed 24 May 2021>.

age limit for certain types of employments contracts because the Italian legislation pursued a social objective encouraging entry of young workers in the labour market.[225]

In the scenario of a black-box algorithm, the recruiter would be entirely unfamiliar with the reason why an algorithm would draw the conclusion that gender correlates with a job applicant's level of productivity. If the age limit in *Abercrombie & Fitch Italia* had been set by an algorithm, the respondent would first need to understand why the algorithm had chosen age as a criterion for the selection of workers under short-term contracts. Only when it is established that for these certain contracts the algorithm had given preference to job applicants without prior work experience, can it plausible be argued that the algorithm tried to encourage first entry of young workers in the labour market. In this hypothetical case it becomes obvious that in a black-box scenario, arguments to justify algorithmic biases would require a comprehensive understanding of the logic behind the algorithm's decision. From that understanding it is then possible to draw up arguments pertaining to the legitimacy, necessity, and proportionality of the difference in treatment resulting from that decision.

In order to gain insight into the logic behind an algorithmic biased decision, a recruiter would have the choice between either induce the logic probably followed in the making of such a decision or call upon experts to reverse-engineer the process. In the first scenario, the recruiter might rely on a vast array of factor, for example, the nature of and requirements for given job, the instructions fed into the algorithm and the available data on requirement practices in the specific sector. A practical test, in order to construct a convincing narrative on the development of a given bias, might be that of a hypothetical human recruiter.[226] The objective of the test would be to determine whether a human would express the same bias as that expressed by an algorithm. For example, based on the scenario that an algorithm is given an instruction to select committed worker, it might disregard women with children, as its training data may contain residual labour market tendencies according to which such women tend to focus on their families and, as a consequence, make slower career progress.[227] Given that in many white-collar professions, the yet prevailing view seems to be that women with children are less committed to their work, it is not unlikely that a conventional human

[225] Judgement of 19th July 2017, *Abercombie & Fitch Italia Srl v Antonino Bordanoro*, Case C-143/16 EU:C:2017:566.
[226] STOA (Panel for the Future Science and Technology), "*A governance framework for algorithmic accountability and transparency*", 2019, European Parliament, p. 73
[227] Wajcman, Judy, "*New connections: Social studies of science and technology and studies of work*", 20 Work, Employment & Society (2006), 773–786, p. 781

recruiter might express the same bias.[228] Indeed, associations such as those between gender and productivity, race and salary, ethnic background and levels of education are well documented in long-standing data and research. Provided that an algorithm's decision can be presumed to be based on such a "familiar" bias, a recruiter may easily argue that the data used to train the algorithm was probably flawed.

However, as has been presented in earlier sections in this paper, a HR algorithm may express biases which, although present in the market, are not easily detected. For example, researchers have shown that criminal offenders' postcodes correlate with races, since areas in which the crime rate is high are more racially diverse than areas where the crime rate is low.[229] In a hypothetical scenario, an algorithm may give preference to candidates who reside in neighbourhoods classed as "decent" and shortlist only white candidates. Nevertheless, in such a scenario, the recruiter would face obstacles in figuring out that the algorithm used the candidate's postcodes as a criterion in its selection. When the answer to how of an algorithmic bias is expressed, when there is not an obvious association with a long-standard residual market bias, the recruiter might consider the second alternative – reverse engineering of an algorithm.

Provided that a recruiter did manage to gain more insight into the occurs of an algorithmic bias, it is still required to provide an argument based on legitimacy, necessity, and proportionality. Legitimacy is usually a matter of pursuing a public policy objective or when discrimination occurs in a recruitment context, a matter of essential professional requirements.[230] In its case law, the ECJ, has admitted that a difference in treatment may pursue a legitimate aim if a protected characteristic is job-related.[231] In the hypothetical variant on *Abercombie & Fithc Italia*, where an algorithm had chosen age as a selection criterion, the respondent could argue that the decision of the algorithm was in line with a policy7 objective of battling unemployment among young workers.

Nevertheless, it is obvious that arguing the legitimate aim in relation to discrimination stemming from black-box algorithms is a challenge. The main challenge is found in the fact

[228] Gerards, Janneke & Xenidis, Raphaele, *"Algorithmic Discrimination Europe: Challenges and opportunities for gender equality and non-discrimination law"*, p. 41.
[229] Bathaee, Yavar, "*The Artificial Intelligence black box and the failure of intent and causation*", p. 920.
[230] Barnard, Catherine, "EU Employment Law", p. 286.
[231] Barnard, Catherine, "EU Employment Law", p. 287.

that user would assign a legitimate aim to a bias that they themselves did not express. In extension it also affects the respondent's arguments on necessity and proportionality. Essentially, the proof on these points aims at establishing that a difference in treatment is necessary and is proportionate. In other words, the respondent discriminates, but in order to achieve a legitimate aim and the discrimination is appropriate for the achievement of that aim. A respondent could argue that a discriminatory practice or measure is necessary and proportionate if a protected characteristic is a relevant indicator of a worker's productivity.

For an example, in the case of *Wolf*, an age limit for the recruitment of voluntary firefighters was considered necessary and proportionate because of the occupation required a certain level of physical fitness and the traditional indicator of that is age.[232] Subsequently, if an algorithm was used in the same recruitment process, it would most likely express the same bias. However, if in addition to the age indicator, the algorithm also considered gender and sexual orientation, and shortlisted only white heterosexual men in their early thirties, an unbiased human recruiter would have some difficulty in explaining how gender and sexual orientation affect the level of fitness required for the occupation of firefighter.

In conclusion from the several hypothetical situations presented in this section, if the presumption of human agency continues to be irrefutable, the consequence will that a respondent's defence will be weak due to its inferential nature. Attempting to infer the underlying logic of an unfair bias in an algorithm is guesswork. It contradicts the basic legal principle, as expressed in the case law of the ECJ on the proof of risk, that when evidence is inferential it should never be based on pure hypothesis but on facts, which when added together, point out the most probable of many possible scenarios.[233] In the case of black-box algorithms, there are hardly any tangible facts that the respondent could refer to in order to detect, with a high degree of certainty, the logic followed in the making of a decision stemming from such an algorithm. Consequently, the implication is that whenever presumptions of algorithmic discrimination are established, they would in fact be irrebuttable as their justification would be practically unfeasible.

[232] Judgement of 12th January 2010, *Colin Wolf v Stadt Frankfurt am Main*, Case C-229/08 EU:C:2010:3.
[233] Judgement of 9th September, *Monsanto Agricoltura Italia SpA and Others v Presidenza del Consiglio dei Ministri and Others*, Case C-236/01 EU:C:2003:431, para 106.

Read together with the goals of effectiveness and fairness, as mentioned above, it may seem desirable to acknowledge that when human users in e-recruitment cases claim to be unbiased, they should be given the procedural possibility to prove it.

5.2.1 Reinforcing the grounds of defence

A rebuttal of the presumption of human agency would imply that the exact origin or cause of an unfair bias becomes a relevant fact in cases of algorithmic discrimination, as discussed in above. In addition to the traditional justification of legitimacy, necessity and proportionality, a respondent would be able to prove themselves unbiased and that a specific case of discrimination was entirely at an algorithm's doing. Provided that such evidence is declared admissible, the recruiter has two options at his or her disposal. Either seek to establish that the algorithmic bias was embedded or that it was machine-learnt.

As mentioned above, evidence to support embedded biases could be data used to train an algorithm. This could be obtained through the developers of the algorithm, who may be called to provide information on types of data used during the training phase of the algorithm. Such testimonial evidence has already been used in cases of algorithmic liability in USA.[234] In addition to witness evidence, a recruiter could present statistical evidence on residual biases. [235] For example, if an algorithm shortlisted onlyafemale candidates for a part-time position, it could be argued that it perpetuated a residual market bias of part-timers mostly being women.[236] This could be supported with available statistics and, insofar as the algorithm was trained on data pertaining to a particular sector in the labour market, reinforce the plausibility of the fact that a residual bias had affected the algorithms performance in practice.

In other scenarios, it could be argued that the bias was machine-learnt. For example, if an algorithm includes irrelevant data. Following the recruiter's instructions to select candidates who are high achievers, the algorithm may take membership of country clubs as a relevant criterion and consequently give preference to male candidates. In such a scenario, a recruiter would need to provide evidence revealing an algorithm's autonomous deviation from its

[234] US Court of Appeals for the 7th Circuit, US v. Coscia, 866 F.3d 782 (2017).
[235] Wajcman, Judy, "*New connections: Social studies of science and technology and studies of work*", p. 781.
[236] Ibid., p. 782.

original instructions. In EU law of evidence, such proof of deviations is admitted under principles governing the rebuttal of presumptions of fault. In a case concerning competition law, *Akzo*[237], a subsidiary's conduct isaimputed to the head company, unless it is established that the latter had distanced itself from the directives given by the head company.[238] Applied to a case with algorithms, a recruiter could seek to establish that the algorithm has distanced itself from its original directives.

Nevertheless, there are mainly two things that need clarification. Firstly, the procedural consequences of the rebuttal of the presumption of human agency and, secondly the evidentiary requirements that recruiters will be held.

Regarding aftermath of a rebuttal of the presumption of human agency, the already very much debated topic of accountability takes form. Provided that a respondent managed to successfully establish the algorithmic origins of a bias, it will remain unclear as to who should then be accountable. From the current legal debate, it can only be assumed that if AI personhood is not an option, the human user of an algorithmic system would continue to be held accountable, even when proven that an unfair bias was entirely originated on the algorithm.[239] In the surface, rebutting the presumption of human agency does not seem to add value to a respondent in a case of algorithmic discrimination. Even though the respondent manages to prove than an algorithm had developed a bias autonomously, it would still not be able to avoid accountability. Simply because algorithms cannot be held accountable.

Regarding the standard of proof for respondents in algorithmic discrimination cases is unclear. It is possible that in order not to overburden a respondent, the sufficiency threshold is kept low. For example, evidence that a business favours unbiased recruitment or than an algorithm was likely trained on biased data, may be enough to rebut the presumption of human agency. Nevertheless, if the standard of proof is higher, courts may develop a system of expert evidence since algorithmic screening is the only available alternative to reverse-engineer algorithmic decisions and highlight the underlying reason of a given bias.

---

[237] Judgement of 10th September 2009, "*AKZO Nobel NV and others v Commission of the European Communities"*, Case C-97/08 P EU:C:2009:536.
[238] Jones, Allisson & Sufrin, Brenda, "*EU Competition Law: Text, Cases and Materials*" (7th Edition, Oxford University Press (2019), p. 156.
[239] Schlaepfer, Daniel & Kruyne, Hugo, "*AI and robots should not be attributed legal personhood*", (Euractiv, 26th Mars 2018), AI and robots should not be attributed legal personhood – EURACTIV.com, <accessed 29th June 2021>.

A more subtle analysis of the notion of use of algorithms may lead to a more nuanced perspective. The doctrine on criminal liability may provide useful guidance. For the purpose of designating the accountable agent for AI decisions, there have been a proposal that the degree of human supervision and the degree of AI autonomy shall be used as reference points. [240] The proposal presents four stages of liability, corresponding to four levels of transparency, which are gradually increasing from transparent and supervised, to quasi or fully unsupervised algorithmic decisions. In situations where AI is autonomous but supervised, it is argued that the principal-supervision from agency law applies. The test being whether the creator or user of the system exercised reasonable care in the development and deploying phases. However, in those situations where AI is both unsupervised and autonomous, the main issue would be whether it is reasonable to have use such a system at all.

It can be argued that the degree of human supervision over the HR algorithm will be key in determining where accountability should lie. By virtue of the traditional supervision-rule, the recruiter is assumed to exercise a general direction over the algorithm, insofar as they define the selection criteria for the recruitment process and instruct those criteria to the algorithm. Nevertheless, with regard to the close monitoring of the performance of the algorithm, the recruiter could argue that, insofar as the bas was concerned, the algorithm was unsupervised due to the opacity of an algorithmic decision. If a court concluded that a bias was either embedded or machine-learnt, the recruiter could consequently rely on that decision and bring an action of product liability against the company that supplied the algorithm. Therefore, in a case of algorithmic discrimination, the respondent would continue to be held accountable, whatever the cause of the bias. But if a court ruled that the bias was developed outside the recruiter's supervision or sphere of influence, the decision would provide it with a tool to bring an action against the developer of the algorithm on the grounds of product liability.[241]

[240] Bathaee, Yavar, "*The Artificial Intelligence black box and the failure of intent and causation*", p. 936.
[241] *Borghetti, Jean-Sébastien, "Civil Liability for Artificial Intelligence: What Should its Basis Be?*", La Revue des Juristes de Sciences Po, June 2019, n°17, 94-102, p. 99.

6. Conclusion

The increased usage of algorithms in all areas of society poses challenges in terms of discriminations and raises legal challenges, has been know for some time. If the increasing use of algorithms in decision-making, in areas such as recruitment processes, risks systematically amplifying and magnifying existing structural inequalities, it is not enough to only focus on tools which aims to combat algorithmic discrimination. It is of equal importance to make sure that there are effective ways to procedurally challenge such discriminatory decisions.

This paper aimed to outline how new technologies may challenge traditional legal principles on agency and feasibility of evidence, in the context of EU non-discrimination law. The opacity of algorithmic decision-making processes is a complex and serious hurdle to the access to evidence in proceedings, as well to the plaintiff as to the defendant. Such a lack of access to evidence is a challenge to the effective application of the non-discrimination directives. Without sufficient evidence, cases of algorithmic discrimination would be inadmissible. Therefore, it is relevant to reflect on the regulatory choices that could or should be made in order for EU non-discrimination law to be effective.

The issue of ML systems reconfigures the traditional divide between direct and indirect discrimination. It has been pointed out that direct discrimination is unlikely to arise, since biases in the processing of data would lead to indirect rather than direct forms of discrimination. Furthermore, another challenge is that of the transparency and explainability of ML systems. Such systems are “black boxes” – their inner workings are not intelligible, not even to experts, which makes it difficult to trace and isolate the source a given discriminatory output. Moreover, it is also hard to establish whether the bias stems from biases data or from another source. Even though the source is pinpointed, ML systems are in constant evolution and consequently, the source of discrimination can change over time or even disappear altogether.

Unlike direct discrimination, which may often be self-evident, indirect discrimination is more challenging to prove. This is made even more challenging in the context of ML systems in a recruitment process. In indirect discrimination cases, the plaintiff’s burden is considerable

more difficult since he or she has to show that a neutral criterion is in practice discriminatory. The evidentiary difficulties in proving discrimination in conventional recruitment processes are great and would be even greater if such processes were automated by ML systems. In such a case, the obstruction to evidence for both parties seem to be absolute, leaving algorithmic discrimination outside the scope of application of the non-discrimination directives altogether since there are no facts to support either side.

My opinion is that there are arguably two possible directions in which the system of evidence in EU-non discrimination law may evolve, in order to remedy the inbalance. First, the increasing amount of algorithmic discrimination cases may push the ECJ to interpret the evidentiary rights and obligations of the parties under the non-discrimination directives in such a that their procedural rights would be reinforced. In such a case, victims would benefit from a right to transparency, forcing for example recruiters to provide access to information that could support presumptions of algorithmic discrimination. In addition, respondents would be given a chance to rebut the presumption of human agency by showing that a discriminatory bias were machine-learnt, or embedded. Such a change of the procedural rights and duties of the parties would require a reinterpretation of the system of evidence in two regards.

Firstly, the principles on admissibility and relevance would require a slight adjustment. Assuming that the origins of unfair biases become evidentially relevant in algorithmic cases of discrimination, the scope of evidence would need to be broader to include proof of that a bias was autonomously developed by an algorithm. Secondly, provided that algorithmi9c decisions constitute cases of automated data processing, a joint reading of the GDPR and the non-discrimination directives may provide a possibility to victims of algorithmic discrimination to gain access to vital evidence pertaining the instructions according to which recruitment was conducted. This would allow them to access information which they could include in a lawsuit. Such a rights-based approach to the evidence of algorithmic discrimination would achieve three important goals in trying to make a procedure more balanced. First, the balance of the allocation of the burden of proof under EU-non discrimination law would remain unmodified. Secondly, both the plaintiff and the defendant would gain more access to evidence in support of their different arguments. And, thirdly, which might be the most important, by reinforcing the right to access evidence, cases of algorithmic discrimination would change from hopeless to only very difficult to bring

forward. In summary, all parties in algorithmic discrimination cases would not be left entirely without protection from EU-law.

Another possibility for the evolution of EU non-discrimination law is for it to include a general presumption of algorithmic discrimination, adapting a more risk-based approach. However, such a presumption would be more advantageous to victims, who would benefit by being discharged of the initial burden of proof. Simultaneously, the respondents would be faced with the difficult task of having to rebut the presumption of human agency. The risk-based approach is a convenient alternative but not a fair one since it upsets the existing allocation of the burden of proof under current EU non-discrimination law.

In both scenarios, the key issue is whether the presumption of human agency should be irrefutable or not. However, the answer to that question depends on the conclusion of a far deeper issue – how should AI be classified in law? Given the rapid development of AI, the issue of AI agency will become more pressing. Depending on how this issue is resolved, more light will be shed on ways in evidence in algorithmic discrimination cases will be assessed and adduced in the future.

In the end, the application of machine learning to ever more important societal and economic decisions should not only be perceived as a risk, but also as a potential opportunity.

7. Bibliography

Legal Sources

European Union

Treaties & Agreements

Charter of Fundamental Rights of the European Union (2000/c 364/01) (CFR), OJ C 2012/C 326/02

Consolidated version of the Treaty on the European Union (TEU) , OJ C C115/13

Consolidated version of the Treaty on the Functioning of the European Union (TFEU), OJ C 326/01

Directives

Council Directive 92/85 on the introduction of measures to encourage improvements in the safety and health at work of pregnant workers and workers who have recently given birth or are breastfeeding, O.J. 1992, L 348/1

Council Directive 2000/43/EC of June 29 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin (Racial Equality DIrective) (200) OJ L 180/22.

Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation OJ L 303.

Directive 2006/54/EC of the European Parliament and of the Council of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (recast) (Gender Recast Directive) [2006] OJ L 204/23.

Regulation

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) OJ L 119/1

Official Documents

European Union

European Commission (2018), *Communication from the Commission to the European Parliament, The European Council, The Council, The European Economic and Social Committee and the Committee of the Regions – Artificial Intelligence for Europe*, COM(2018) 137 final.

STOA (Panel for the Future Science and Technology), "*A governance framework for algorithmic accountability and transparency*", 2019, European Parliament

Castelluccia, Claude and Le Métayer, Daniel, (2019), "*Understanding algorithmic decision-making: Opportunities and challenges*", Panel for the Future of Science and Technology (STOA) of the European Parliament

Gerards, Janneke & Xenidis, Raphaele, *"Algorithmic Discrimination Europe: Challenges and opportunities for gender equality and non-discrimination law*, (2021) Publications Office of the European Union.

White Paper, "*On Artificial Intelligence – A European approach to excellence and trust*", COM (2020)65 final,

Gerards, Janneke & Xenidis, Raphaele, *"Algorithmic Discrimination Europe: Challenges and opportunities for gender equality and non-discrimination law*, (2021) Publications Office of the European Union.

European Commission, (2021), "*Report from the Commission to the European Parliament and the Council on the application of Council Directive 2000/43/EC implementing the principle of equal treatment between persons irrespective of racial or ethnic origin (the Racial Equality Directive") and of Council Directive 2000/78/EC establishing a general framework for equal treatment in employment and occupation (the Employment Equality Directive"*)", SWD(2021) 63 final (Brussels)

Proposal for a *REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS,* COM/2021/206 final.

Sweden

SOU 2018:16, "*Vägen till självkörande fordon – introduktion*"

Articles

Barocas, Solon & Selbst, Andrew, "Big Data's Disparate Impact", (2016), 104 Californa Law Review

Bathaee, Yavar, “The Artificial Intelligence black box and the failure of intent and causation”, 31 Harvard Journal of Law & Technology (2018)

Borgesius, Fredrick and Poort, Joost, (2017), “Online Price Discrimination and EU Data Privacy Law” Journal of Consumer Policy 40 347

Borgesius, Fredrick, “Discrimination, Artificial Intelligence, and Algorithmic Decision-Making”, (Council of Europe, Directorate General of Democracy 2018)

Borghetti, Jean-Sébastien, “Civil Liability for Artificial Intelligence: What Should its Basis Be?”, La Revue des Juristes de Sciences Po, June 2019, n°17 ISSN 2111-4293, 94-102,
Cofone, Ignacio, (2019), “Algorithmic Discrimination Is and Information problem”, 70(6) Hastings Law Journal 1389,

Edwards, Lilian & Veale, Michael, “Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling”, 34 Computer Law & Security Review (2018), 398–404
Fagan, Frank and Levmore, Saul (2019), “The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion”, Southern California Law Review 93 2

Gurney, Kevin, “An introduction to neural networks”, first published 1997, UCL Press

Hacker, Philipp, (2018), “Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law”, 55 Common Market Law Review 1143

Henderson, Tristan, “Does the GDPR Help or Hinder Fair Algorithmic Decision-Making?” (2017), LLM dissertation, Innovation, Technology & The Law, University of Edinburgh, 2017

Howard, E (2006), “The case for a considered hierarchy of grounds in EU law”, 13 Maastricht Journal of European and Comparative Law 445.

Jareborg, Nils , ‘Rättsdogmatik som rättsvetenskap’ (2004) SvJT 1

Kim, Pauline, (2017)“Data-Driven Discrimination at Work”, 58 William and Mary Law Review 857

Kleinberg, J and others, “Discrimination in the Age of Algorithms”, (2018), Journal of Legal Analysis, Vol 10.

Kroll, Joshua & others, (2017), “Accountable Algorithms”, 165 University of Pennsylvania Law Review 633

Kulk S & others (2020), “Legal aspects of algorithms that make decisions. An exploratory study”, The Hague, WODC 2020

Köchling, Alina & Wehner, Marius, ”*Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment*

and HR development", Bus Res 13, 795–848 (2020)

Leese, Matthias, (2014), "The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union", 45 Security Dialgoue, 494

Lehr, David & Ohm, Paul, (2017)"Playing with the Data: What Legal Scholars Should Learn About Machine Learning", UC Davis Law Review, Vol. 51

Maliszewska-Nienartowicz, Justyna, "Direct and Indirect Discrimination in European Union Law – How to Draw a Dividing Line?", 111(1) International Journal of Social Sciences 40, McCrudden, Christopher, "Legal research and the social science", (2006), The Law Quarterly Review, Oxford Legal Studies Research Paper No. 33/2006.

Mendoza, Isak and Bygrave, Lee A., "The Right Not to Be Subject to Automated Decisions Based on Profiling" (May 8, 2017), EU Internet Law: Regulation and Enforcement (Springer, 2017, Forthcoming), University of Oslo Faculty of Law Research Paper No. 2017-20

Parasuraman, Raja & Riley, Victor, "Humans and Automation: Use, Misuse, Disuse, Abuse", 39 Human Factors, 1997:39(2)

Senden, Linda & Xenidis, Raphael, "EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination", in Ulf Bernitz et al (eds), General Principles of EU law and the EU Digital Order, Kluwer Law International, 2020.

Surden, Harry, "Artificial Intelligence and Law: An Overview" (2019), Georgia State University Law Review 1305

Wachter, Sandra, (2020), "Affinity profiling and discrimination by association in online behavioural advertising" Berkely Technology Law 35(2)
Wachter, Sandra, Mittelstadt, Brent, & Floridi, Luciano, "Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation". (2017) International Data Privacy Law, 7(2), 76–99

Wajcman, Judy, "New connections: Social studies of science and technology and studies of work", 20 Work, Employment & Society (2006), 773–786

Xenidis, Raphael, (2020), "Tuning EU equality law to algorithmic discrimination: Three pathways to resilience", 27(6) Maastricht Journal of European and Comparative Law 736,

Books

Craig, Paul & De Búrca, Gráinne," *EU Law: Text, Cases, and Materials*" (6th edn, Oxford University Press 2015), p. 266.

Géron, Aurélien, "*Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and techniques to Build Intelligent Systems"*, (2019), O'Reily Media, p.177.

Hettne, Jörgen and Eriksson, Ida, ”*EU-rättlig metod, Teori och genomslag i svensk rättstillämpning*”, (2011), 2nd edition, Nordstedts juridik

Jones, Allisson & Sufrin, Brenda, “*EU Competition Law: Text, Cases and Materials*” (7th Edition, Oxford University Press (2019), p. 156.

Internet sources

Dastin, Jeffrey, ”Amazon scraps secret Ai recruiting tool that showed bias against women”, (Reuters, 11 October 2018), <accessed 24 May 2021>.

Decille, Joe, “Leaky Data: How Wonga Makes Lending Decisions.” (Charisma: Consumer Market Studies, 20 May 2013), https://www.charisma-network.net/finance/leaky-data-how-wonga-makes-lending-decisions, <accessed 24 May 2021>

Hunt, Elle, ”Tay, Microsoft’s AI Chatbot, gets a crash course from Twitter”, (The Guardian, 24 Mars 2016), https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter, <accessed 24 May 2021>.
Larson, Jeff and others, ‘These Are the Job Ads You Can’t See on Facebook If You’re Older’, (The New York Times, December 20, 2017), < https://www.nytimes.com/2017/12/20/business/facebook-job-ads.html > accessed 25 May 2021.

Lentz, Aurore, “Garbage in, garbage out: is AI discriminatory or simply a mirror of IRL inequalities?” (Universal Rights Group, 18 January 2021), https://www.universal-rights.org/blog/garbage-in-garbage-out-is-ai-discriminatory-or-simply-a-mirror-of-irl-inequalities/ ,<accessed 3 April 2021>.

Lobosco, Katie, “Facebook Friends Could Change Your Credit Score”, (CNNMoney, 27 August 2013), https://money.cnn.com/2013/08/26/technology/social/facebook-credit-score/index.html, <accessed 24 May 2021>.

Schlaepfer, Daniel & Kruyne, Hugo, “*AI and robots should not be attributed legal personhood*”, (Euractiv, 26th Mars 2018), AI and robots should not be attributed legal personhood – EURACTIV.com, <accessed 29th June 2021>.

Reports

Gerards, Janneke & Xenidis, Raphaele, *“Algorithmic Discrimination Europe: Challenges and opportunities for gender equality and non-discrimination law*, (2021) Publications Office of the European Union.

Borgesius, Fredrick, "*Discrimination, Artificial Intelligence, and Algorithmic Decision-Making",* (Council of Europe, Directorate General of Democracy 2018), p. 10.

Table of Cases

EU

Judgement of 13th May 1986, *Bilka – Kaufhau8s GmbH v Karin Weber von Hartz*, Case C-278/84 EU:C:1986:204

Judgment of 17th October 1989, *Handels- og Kontorfunktionærernes Forbund I Danmark v Dansk Arbejdsgiverforening, acting on behalf of Danfoss*, Case 109/88, ECLI:EU:C:1989:383

Judgement of 27th October 1993, *Dr. Pamela Mary Enderby v Frenchay Health Authority and Secretary of State for Health,* Case C-127/92 EU:C:1993:859

Judgement of 9th September 2003, *Monsanto Agricoltura Italia SpA and Others v Presidenza del Consiglio dei Ministri and Others,* Case C-236/01 EU:C:2003:431

Judgment of 11 July 2006, *Sonia Chacón Navas v Eurest Colectividades* SA C-13/05 EU:C:2006:456;

Judgement of 10th July 2008, *Centrum voor gelijkheid van kansen en voor racismebestrijding v Firma Feryn NV*, Case C-54/07 EU:C:2008:397

Judgment of 17 July 2008, *S. Coleman v Attridge Law and Steve Law* C–303/06 EU:C:2008:415

Judgement of 10th September 2009, "*AKZO Nobel NV and others v Commission of the European Communities"*, Case C-97/08 P EU:C:2009:536.

Judgement of 12th January 2010, *Colin Wolf v Stadt Frankfurt am Main*, Case C-229/08 EU:C:2010:3.

Judgement of 19th April 2012, *Galina Meister v Speechg Design Carrier Systems GmbH*, Case C415/10, EU:C:2012:217.

Judgment of 18 December 2014, *Fag og Arbejde (FOA) v Kommunernes Landsforening* (KL) C-354/13 EU:C:2014:2463

Judgment of 16 July 2015, '*CHEZ Razpredelenie Bulgaria' AD v Komisia za zashtita ot diskriminatsia* C-83/14 EU:C:2015:480.

Judgement of 6 April 2017, *Jyske Finans A/S v Ligebehandlingsnaevnet, acting on behalf of Ismar Huskic* C-668/15 EU:C:2017:278.

Judgement of 19th July 2017, *Abercombie & Fitch Italia Srl v Antonino Bordanoro*, Case C-143/16 EU:C:2017:566.

Judgement of 19th October 2017, *Elda Otero Ramos v Servicio Galego de Saude and Institutio Nacional de la Seguridad Social*, Case C-531-15 EU:C:2017:789.

Judgment of 15 November 2018, *Heiko Jonny Maniero v Studienstiftung des deutschen Volkes eV* C-457/17 EU:C:2018:912

Advocate General Opinion

Handels- og Kontorfunktionærernes Forbund I Danmark v Dansk Arbejdsgiverforening, acting on behalf of Danfoss, Case 109/88, Opinion of AG Lenz, delivered on 31 May 1989

US

McDonnel Douglas Corp. V. Green, 411 U.S. 792 (1973).

US Court of Appeals for the 7th Circuit, US v. Coscia, 866 F.3d 782 (2017).