

Available online at www.sciencedirect.com

ScienceDirect

www.compseconline.com/publications/prodclaw.htmComputer Law
&
Security Review

Preventing discrimination in the automated targeting of job advertisements

David Jacobus Dalenberg *

University of Amsterdam, Faculty of Law, International and European Law: Public International Law, Amsterdam, Netherlands

ABSTRACT

Keywords:

2000/43/EC
2000/78/EC
2006/54/EC
Algorithmic models
Artificial intelligence
Big data
Data mining
Employment equality
Machine learning
Non-discrimination directives

On the background of the increasing amount of discriminatory challenges facing artificial intelligence applications, this paper examines the requirements that are needed to comply with European non-discrimination law to prevent discrimination in the automated online job advertising business. This paper explains under which circumstance the automated targeting of job advertisements can amount to direct or indirect discrimination. The paper concludes with technical recommendations to dismantle the dangers of automated job advertising. Various options like influencing the pre-processing of big data and altering the algorithmic models are evaluated. This paper also examines the possibilities of using techniques like data mining and machine learning to actively battle direct and indirect discrimination. The European non-discrimination directives 2000/43/EC, 2000/78/EC, and 2006/54/EC which prohibit direct and indirect discrimination in the field of employment on the grounds of race or ethnic origin, sex, sexual orientation, religious belief, age and disability are used as a legal framework.

© 2017 David Jacobus Dalenberg. Published by Elsevier Ltd. All rights reserved.

1. Introduction

With the rise of artificial intelligence (AI) and the accompanying subfields of big data, data mining and machine learning, a lot of human tasks can be successfully performed by AI-driven software. A White House report on big data warns that such innovations can root discrimination deeply into society

and reinforce prejudice and bias.¹ An example of discriminatory AI is the computer program which is used in some jails to determine which prisoners are eligible for parole. The program generates a risk assessment score to determine which prisoners are likely to re-offend. According to a research done by ProPublica, the system is biased against prisoners of colour.² Such technologies are applied to automate decisions in multiple other fields such as online advertising and employment.³

* University of Amsterdam, Faculty of Law, International and European Law: Public International Law, Rombout Hogerbeetsstraat 52-3, 1052XG, Amsterdam, Netherlands.

E-mail address: David.Dalenberg@gmail.com.

¹ John Podesta and others, 'Big Data: Seizing Opportunities' (2014) <https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf> accessed on 23 July 2017.

² Julia Angwin, Surya Mattu and Lauren Kirchner, 'Machine Bias: There's Software Used across the Country to Predict Future Criminals. And It's Biased against Blacks.' ProPublica (23 May 2016) <<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>> accessed on 6 May 2017.

³ Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequity and Threatens Democracy* (1st edn, Crown 2016) 13. <https://doi.org/10.1016/j.clsr.2017.11.009>

Imagine that 50 years ago, a newspaper gave the option to advertise vacancies only in copies that went to male readers. Advertising like this belongs to the possibilities that platforms like Facebook provide nowadays. With countless targeting settings, people can be excluded until the advertiser has reached the perfect audience. When AI is used to control and apply these settings it is vital that it does not do so in a discriminatory way, especially in the field of employment. The chances for job seekers are seriously diminished when they are excluded from seeing job advertisements because this gives them a false start.⁴ This undermines the principle of equality from which follows that every individual should have the same opportunities, including equal access to employment.⁵ A good example of discriminatory job advertising can be found in a research that analysed advertising placements.⁶ It discovered that an advertisement for a high-paying executive position was shown almost six times more to men than to women.⁷ However, when used in the right way, AI can efficiently identify and reach the candidates possessing the required skills for a job while avoiding individual biases.⁸

On the background of the increasing amount of discriminatory challenges facing AI applications, this paper (or hereafter “we”) examines discrimination in the automated online job advertising business in Europe. Because this topic is so extensive, this paper examines how discrimination can be prevented in the automated online targeting of job advertisements. This question is answered in this paper in four steps. First, the technical elements that come into play and may cause discrimination when using AI to target advertisements are presented in Section 2. Secondly, the scope and effect of European non-discrimination law are established in Section 3. Thirdly, Section 4 examines in which ways the targeting of job advertisements can be discriminatory. Fourthly, technical recommendations on how to prevent discrimination when the targeting is done by AI are presented in Section 5. Various options like influencing the pre-processing of big data, altering the algorithmic models are evaluated. Section 5 also examines the possibilities of using techniques like data mining and machine learning to actively battle direct and indirect discrimination. Finally, Section 6 concludes the paper.

⁴ Paul Post and Rikki Holtmaat, ‘A False Start: Discrimination in Job Advertisements’ (2014) 2014/1 European Gender Equality Law Review 12, 12.

⁵ J Clifford, ‘Equality’ in D Shelton (ed), *The Oxford Handbook of International Human Rights Law* (1st edn, Oxford University Publishing 2013) 428.

⁶ Amit Datta, Michael Carl Tschantz and Anupam Datta, ‘Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination’ (2015) 2015 Proceedings on Privacy Enhancing Technologies 92.

⁷ Ibid 102.

⁸ Cecilia Munoz, Megan Smith and DJ Patil, ‘Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights’ (2016) 16 <https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf> accessed on 23 July 2017.

2. Using artificial intelligence in automated online advertising

In order to scrutinize automated online job advertising, an examination of the advertising process and a delimitation of the subject is needed. This section explains the practice of online advertising and which factors play a role in the outcome of an online advertising campaign.

2.1. Online job advertising

Traditionally, advertising took place through billboards, magazines, television or radio. Advertisers could play with what they would show, and more generally to whom, by changing the content, channel and timing of the advertisement. For this reason, toy commercials were not shown on MTV but rather on Nickelodeon. Since the invention of the internet, online advertising has become a booming business and experts estimate that by 2019 online advertising will become the biggest segment of advertising.⁹

The best known online advertising channels are Facebook and Google. They let other companies advertise to their users on their websites. Next to that, they provide advertising banner space on all kinds of websites or telephone apps. These platforms have their own advertising software. Facebook’s software is called Facebook Business Manager and Google has AdWords.¹⁰ These kinds of programs use human input to create an advertising campaign. Advertisers can set and change all kinds of variables and apply numerous advertising strategies in these programs to make sure their advertisements reach their target audience. For example, advertisers can target advertisements for dresses at women, diaper advertisements at soon-to-be moms and car advertisements at people with a driver’s license. This kind of targeting results in a big difference with traditional advertising since the audience reached can be precisely manipulated instead of generally influenced. The effectiveness of advertising increases because personalized advertisements reach the right person at the right time.¹¹

This paper focuses on the extensive targeting options rather than on the content of online advertisements since the rules concerning the content equally affect an online and a regular offline advertisement. This paper concentrates on Facebook Business Manager because it has the most comprehensive and far reaching targeting settings.

When creating a new campaign in Facebook Business Manager, there is one particular goal that has to be reached. In an employment opportunity case, the campaign needs to result in large numbers of applications by suitable candidates. The successful achievement of the goal can be influenced by two main factors: first, showing the advertisement to the right people and second, showing the right people what they

⁹ Shelly Rodger and Esther Thorson, *Digital Advertising: Theory and Research* (3rd edn, Routledge 2017) 345.

¹⁰ For more information see: <https://www.facebook.com/business/products/ads> and <https://adwords.google.com/home/how-it-works/> both accessed on 16 June 2017.

¹¹ Alexander Bleier and Maik Eisenbeiss, ‘Personalized Online Advertising Effectiveness: The Interplay of What, When, and Where’ (2015) 34 Marketing Science 669.

want to see. In other words, the advertisement needs to be targeted to the right people and the content needs to be appealing enough to click on. As said above, we focus on the former.

Facebook provides the option for advertisers to target their advertisements based on *inter alia* location, age, gender, sexual orientation, language, education, job, relationship status, financial status, home ownership, income, political interest and ethnic affinity.¹² These targeting settings are used to narrow down the audience and can be used to include or exclude people. Some of the targeting options have a potential for discriminatory usage. This can be the case when they are used to exclude people of a certain ethnic or racial affinity. In the case of job advertising, these settings can be used to target a job advertisement for marketers at people who studied marketing. However, they can also be used to further narrow down the audience within the population of marketers. Further narrowing down the audience is convenient, especially in large populations, to separate the more suitable candidates from the less suitable candidates. Ergo, targeting is the key element in reaching the right candidates for a job offer because it can be used to alter which people get to see the advertisement. The next section establishes how AI is involved in the targeting process.

2.2. Using artificial intelligence

Targeting in the online advertising process is normally done by a human advertiser. In the automated process this is done by artificial intelligence. This section explains how this works by elaborating on the definitions of AI and some of its sub elements.

AI can be defined as ‘any device that perceives its environment and takes actions that maximize its chance of success at some goal’.¹³ In the automated advertising process AI is used to make targeting decisions and to judge results based on pre-programmed and self-learned strategies. The AI software is connected through multiple application programming interfaces (APIs) with third party software, like Facebook Business Manager. APIs are a set of function declarations that a software provides to handle requests made by other programs.¹⁴ In this way, APIs connect the output of the AI to the input of the advertising software and vice-versa. The use of APIs is very common and often AIs are connected with multiple other programs. In this way, they create a multi-layered web of information which is used to complete the goal. The goal in this case is to acquire as much relevant applications as possible for each job campaign. By using API’s, the AI can interact with multiple programs and it can take over the decision-making role of the human advertiser. But how does it learn how to make decisions?

It does so by a technique called machine learning. Machine learning addresses the question of how to build a computer

system that improves automatically through its experience.¹⁵ It uses methods like data mining that uncover structures within sets of data and it makes probability predictions based on the discovered patterns which can be used in a decision-making process.¹⁶ In other words, machine learning makes an AI better capable of making the right decision. The longer the AI learns, the more efficient and the higher the quality of the decisions will get. Simply said, the AI learns for each type of job which targeting settings are the most effective. These settings are saved into different function profiles or categories on which each new advertising campaign can be based.

Data mining uses algorithmic models to identify relationships between different attributes with the purpose of extracting useful information from big data.¹⁷ Big data is unstructured and the amount of data is extremely excessive.¹⁸ In our case, algorithmic models are used to find patterns that indicate that candidates are suited for or interested in the job. These patterns may seem completely random to humans because an algorithm looks for statistical correlation without the need for a logical causation between the data.¹⁹ Therefore, the use of big data and data mining can discover relevant effective targeting setting which would have never occurred to humans. This a very strong and useful advantage of the use of big data.

An example of the use of data mining with big data can be found with an employer who used these methods to discover that one predicting factor of strong coding skills is an affinity for a particular Japanese cartoon site.²⁰ It is evident that viewing a cartoon site does not cause or influence strong coding skills. This example shows that an appreciation for cartoons and coding skills are, in this case, correlated but presumably have no causal relation.²¹ Another hypothetical example, when the algorithms discover that the high performing law students share a common interest in tennis, the AI creates the following rule: “Target legal job vacancy at people who both studied law and like tennis.” But one can imagine that some learned rules may result in an unintentional adverse effect for a certain group of people. If an algorithm finds that people of a certain gender, age or ethnic background are worse coders than others, and the AI decides to exclude all those people from the targeting audience, the treatment possibly could be seen as discriminatory.

So, when targeting is done by AI, the factors that influence its targeting decisions are the big data sets, the algorithmic models that mine the data, the targeting rules that are learned from the found correlations and the use of those rules by the

¹² For a full list, see Business Manager or this website: <https://www.contentharmony.com/blog/facebook-ad-targeting/#facebook-ad-targeting-overview> accessed on 16 June 2017.

¹³ Stuart J Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach* (2003 Prentice Hall) 27.

¹⁴ Mehdi Khosrowpour, *Dictionary of Science and Technology* (2nd edn, Hershey 2013) 40.

¹⁵ MI Jordan and TM Mitchell, ‘Machine Learning: Trends, Perspectives, and Prospects’ (2015) 349 *Science* 255, 255.

¹⁶ Lina Zhou and others, ‘Machine Learning on Big Data: Opportunities and Challenges’ (2017) 237 *Neuroscience* 350, 350.

¹⁷ Charu C Aggarwal, *Data Mining: The Textbook* (1st edn, Springer 2015) 129–30.

¹⁸ Min Chen and others, *Big Data: Related Technologies, Challenges and Future Prospects* (1st edn, Springer 2014) 1–5.

¹⁹ Allen G King and Marko Mrkonich, ‘Big Data and the Risk of Employment Discrimination’ (2014) 68 *Oklahoma Law Review* 560–61.

²⁰ Don Peck, ‘They’re Watching You at Work’ *The Atlantic* (December 2013) <<http://www.theatlantic.com/magazine/archive/2013/12/theyre-watching-you-at-work/354681/>> accessed 16 May 2017.

²¹ King and Mrkonich (n 19) 560.

AI. Consequently, all these elements are also the factors that determine whether an AI driven job advertising campaign is discriminatory. So how do we control the ghost in the machine? The answer could lie in non-discrimination law which is set out in [Section 3](#) and applied in [Section 4](#).

3. Non-discrimination law

Provisions on non-discrimination and equality are strongly integrated in international law. The concept of equality has been expressed explicitly in most human rights instruments as a preambular objective, as an implicit descriptive function in the understanding of the scope and application of human rights, and it has been codified in substantive provision of human rights treaties.²² The implementation of these treaties in the domestic legal order is up to each state because of the sovereign character of states. However, human rights treaties are not always implemented in an effective way.²³ The European Union (EU) on the other hand has a supranational character and provides non-discrimination law through its legislative procedures which can penetrate effectively into the domestic legal order.²⁴

3.1. Equality and non-discrimination in European Union law

The focus in this paper lies on European Union non-discrimination law since that is the most specific and comprehensive non-discrimination law. Another reason for focussing on EU non-discrimination law is the fact that it has lots of elaborating case law and that it is applicable in twenty-eight countries. The wide reach of Union law is relevant because companies that use artificial intelligence to advertise online do business in many different states at once.

The concept of equality and non-discrimination can be found in multiple places in primary Union law. According to article 2 of the Treaty on European Union (TEU), the European Union is founded on the values of respect for human dignity, freedom, equality, the rule of law, and respect for human rights.²⁵ The article explicitly mentions the rights of minorities and claims that the principles of non-discrimination and equality between men and women must prevail in the Union's society. Article 3(3) TEU states that the Union will combat discrimination and promote equality between men and women. Article 10 of the Treaty on the Functioning of the European Union (TFEU) stipulates that the Union shall combat discrimination in defining and implementing its policies based on sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation.²⁶

²² Clifford (n 5) 431.

²³ Wade M Cole, 'Mind the Gap: State Capacity and the Implementation of Human Rights Treaties' (2015) 69 International Organization 405.

²⁴ Evelyn Ellis and Philippa Watson, *EU Anti-Discrimination Law* (Oxford Scholarship Online 2013) 43.

²⁵ Consolidated Version of the Treaty on European Union [2012] OJ C326/13.

²⁶ Consolidated Version of the Treaty on the Functioning of the European Union [2008] OJ C115/47.

The key provision is article 19 TFEU which confers the power to take legislative measures to combat discrimination to the Council after obtaining consent of the European Parliament. Based on this article the Union has adopted multiple non-discrimination directives which are addressed in [Section 3.2](#).

The characteristics of Union law have been set out in multiple cases. The scope and effect of Union law was first addressed in the *Van Gend & Loos* case.²⁷ The CJEU held that Union law enjoyed supremacy over domestic law and that it can be enforced by individuals in domestic courts in some cases.²⁸ The concept of enforceability by individuals of Union law is commonly referred to as 'direct effect'.²⁹ Provisions have direct effect when they are clear, unconditional, non-discretionary and when they do not require legislative intervention.³⁰ In *Costa v ENEL* the court elaborated that the spirit and nature of Union law include that it must be higher in rank than national law.³¹ The Union has exclusive competence in the area of battling discrimination when article 19 is read in conjunction with article 2(1) TFEU. From this exclusive competence it follows that primary and secondary Union law is superior to domestic law in the field of non-discrimination.³² From the above it follows that if the Union adopts non-discrimination legislation, the provisions form an integral part of the domestic legal order of the member states and that they enjoy supremacy over domestic provision. Furthermore, if the non-discrimination provisions have direct effect, they can be invoked by individuals or private parties before domestic courts.

3.2. The European non-discrimination directives

The Union has adopted three directives which address discrimination in multiple fields. In this part, their substance, interpretation, implementation and their direct effect are analysed. Directives 2000/43/EC on Racial Equality and 2000/78/EC on Employment Equality are both based upon article 19 TFEU.³³ Directive 2006/54/EC on Gender Equality, which is a recast, is adopted upon article 157(3) TFEU.³⁴ Both these treaty articles give the union competence to legislate in the area of non-discrimination and the directives are therefore legally binding on the member states.

²⁷ Case 26/62 *NV Algemene Transport- en Expeditie Onderneming van Gend & Loos v Netherlands Inland Revenue Administration* [1963] ECR 1.

²⁸ *Ibid* 12–13.

²⁹ Ellis and Watson (n 24) 54.

³⁰ *Ibid* 53.

³¹ Case 6/64 *Flaminio Costa v ENEL* [1964] ECR 585, 593–4.

³² Stephen Weatherill, *Law and Values in the European Union* (11th edn, Oxford University Press 2016) 159 & 163.

³³ Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin [2000] OJ L180/22 (Racial Equality Directive); Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation [2000] OJ L303/16 (Employment Equality Directive);

³⁴ Council Directive 2006/54/EC 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) [2006] OJ L204/23 (Gender Equality Directive).

The Directives prohibit discrimination based on multiple discrimination grounds and in multiple fields. The Racial Equality Directive prohibits discrimination based on race or ethnicity in the context of employment and occupation, social protection, social advantages, education and access to goods and services.³⁵ The Employment Equality Directive prohibits discrimination based on sexual orientation, religious belief, age and disability in the context of employment and occupation.³⁶ The Gender Equality Directive prohibits discrimination based on sex in the context of employment and social security.³⁷ Since this paper focuses on job advertising and thus on employment, not all aspects of the directives are relevant. Most important is that the substance of these three directives together covers discrimination in the field of employment based on race, ethnicity, sex, sexual orientation, religious belief, age and disability.

The interpretation of the directives is not a clear-cut case since the directives themselves lack a clarification on the scope and meaning of the grounds. The CJEU had to make up for this defect in its case law.³⁸ Next to interpretation through case law, international agreements to which the Union is a party can be relied on to interpret the directives. The interpretational value of international agreements follows from article 216(2) TFEU which states that Union institutions, including the CJEU, are bound by international agreements. This binding status was concluded in *HK Danmark* case.³⁹ In this case one of the directives was interpreted in the light of the United Nations Convention on the Rights of Persons with Disabilities.⁴⁰ The above means that CJEU must interpret all three directives in the light of the international agreements by which it is bound.

The directives affect the member states in two ways. First, the directives provide objectives which need to be achieved by all member states. These objectives are result-based obligations of the member states. States are free to determine which means they want to use to achieve these results.⁴¹ Hence, the national non-discrimination provisions in each state should reflect the EU directives. Second, national legislation must be interpreted in the light of the directives. This concept is called indirect effect.⁴² From indirect effect follows that the directives can be used before national courts to interpret domestic laws.

The directives are addressed to the member states so, in principle, they only bind the member states.⁴³ But because the provisions have direct effect, they can be invoked by individu-

als or private parties against the member states before national courts.⁴⁴ So when there is a dispute between a member state and a private party, the provision from the directives can be used to review the actions of the state. This direct vertical effect thus concerns the relationship between national law and Union law. In the case of a horizontal relationship or dispute between private parties, Union provisions can only be used as an autonomous ground for review if they are considered to have direct horizontal effect.⁴⁵ In the *Mangold* and *Kücükdeveci* cases the CJEU introduced the direct horizontal effect of the principle of non-discrimination.⁴⁶ The concept of direct horizontal effect is significant for job advertising companies since it entails that they need to comply with the directives and that individuals can use those provisions in national courts against job advertising companies.

The striking effect of the judgements in *Mangold* and *Kücükdeveci* is the *de facto* horizontal exclusion effect of the non-discrimination directives.⁴⁷ This is based upon two elements. First, the fact that the subject matter of the legislation fell within the scope of the particular directive sufficed to apply the general principle of non-discrimination based on age. This has as a consequence that all national legislation that falls within the scope of the directive would be subject to this principle. Second, the fact that the 'general principle of non-discrimination based on the grounds of age as expressed in Directive 2000/78'⁴⁸ is the basis for review.⁴⁹ Combined this means that in disputes between private parties, national legislation can be overruled by the principles contained in the directives. Ergo, the national provision will not be applied, as it is excluded on the basis of the EU's non-discrimination principle. This means that the directive is used to review national public acts that regulate private legal relationships.⁵⁰ The cases mentioned above only concerned the ground of age but by analogy the same approach would apply to the other discrimination grounds contained in the three directives combined.

Since all member states have to implement the Directives, all national law will be or at least should be in compliance with the Directives. If national law is not completely in line with the directives, it must be interpreted in the light of EU law. Even if it is not in line at all, it can be excluded since by analogy, all non-discrimination directives reflect general principles of Community law and thus all these principles have horizontal direct effect. This means that the Directives can be used autonomously to review horizontal private disputes

³⁵ Racial Equality Directive arts 1–3.

³⁶ Employment Equality Directive arts 1 and 3.

³⁷ Gender Equality Directive art 1.

³⁸ Dagmar Schiek, Lisa Waddington and Mark Bell, *Cases, Materials and Text on National, Supranational and International Non-Discrimination Law* (1st edn Hart 2007) 40.

³⁹ Joined cases C-335/11 and C-337/11, *HK Danmark, acting on behalf of Ring v Dansk Almennyttigt Boligselskab; HK Danmark, acting on behalf of Skouboe Werge v Dansk Arbejdsgiverforening, acting on behalf of Pro Display A/S* [2013], ECLI:EU:C:2013:222.

⁴⁰ *HK Danmark*, para 28–29; Convention on the Rights of Persons with Disabilities (24 January 2007) UN Doc A/RES/61/106.

⁴¹ Ellis and Watson (n 24) 55.

⁴² *Ibid* 73.

⁴³ Racial Equality Directive art 19; Employment Equality Directive art 21; Gender Equality Directive art 36.

⁴⁴ Lilla Farkas, *How to Present a Discrimination Claim: Handbook on Seeking Remedies under the EU Non-Discrimination Directives* (Publications Office of the European Union 2011) 90.

⁴⁵ Bruno De Witte, 'Direct Effect, Primacy and the Nature of the Legal Order' in Grainne De Burca and Paul P Craig (eds), *The Evolution of EU Law* (2nd edn, Oxford University Press 2011) 178 and 213.

⁴⁶ Case C-144/04, *Werner Mangold v Rüdiger Helm* [2005] ECR I-9981; Case C-555/07, *Seda Küçükdeveci v Swedex GmbH & Co. KG* [2010] ECR I-365.

⁴⁷ Mirjam De Mol, 'The Novel Approach of the CJEU on the Horizontal Direct Effect of the EU Principle of Non-Discrimination: (Unbridled) Expansionism of EU Law?' (2011) 18 *Maastricht Journal of European and Comparative Law* (MJ) 109, 123.

⁴⁸ *Kücükdeveci* (n 46) para. 27–33 and 43.

⁴⁹ De Mol (n 47) 123.

⁵⁰ De Mol (n 47) 111.

between job advertising companies and candidates. For these reasons, from the perspective of businesses using automated job advertising, it is most efficient to comply with the EU directives on non-discrimination.

4. Recognising discrimination in the targeting of job advertisements

This section aims to determine in which ways the targeting of a job advertisement can be discriminatory. The concepts of direct and indirect discrimination are set out and applied to the automated targeting of job advertisements in Sections 4.1 and 4.2.

In online job advertising, advertisements are shown based on the information the social media channel has collected about its users. Facebook creates an advertisement profile of its users based on information provided by them and their behaviour on Facebook.⁵¹ This profile does not necessarily represent the real characteristics of its users. In other words, Facebook assumes certain characteristics, and associates its users with certain groups. However, discrimination can be based upon an assumed or associated ground, whether true or not in reality, and, as such, is prohibited under the directives.⁵² Discrimination on associated grounds follows from the *Coleman* case in which a person was discriminated against because she got associated with disabled people without being disabled herself.⁵³ The Court stated in this case that discrimination is the less favourable treatment of someone based upon a protected ground, regardless of that person actually possessing that characteristic or belonging to a protected group.⁵⁴ Another example was the homophobic abuse of a heterosexual man; this constituted discrimination on an assumed ground.⁵⁵ So when Facebook determines that a person belongs to a certain group or possesses a certain characteristic, these associations or assumptions remain specifically profiled to that person. The argument that these characteristics might not be true in reality does not preclude discrimination from occurring.

There is another relevant concept that applies to the automated online advertising of jobs. This concept entails that there does not necessarily have to be an identifiable victim or claimant for sanctions to be applied. This concept was introduced in the *Feryn* case.⁵⁶ The court stated that the existence of direct discrimination does not depend on the identification of a claimant who states to be a victim of discrimination.⁵⁷ The judgement of the *Feryn* case is important because of the

nature of targeting. With certain settings, groups of people can be excluded from seeing an advertisement at all. It is therefore extremely unlikely that they will know that they are being discriminated against and even more unlikely that they will present a claim. A company could argue that for this reason it should not take efforts to comply to non-discrimination rules. This argument is invalid since the *Feryn* case stipulates that discrimination is punishable even without an identifiable victim or claimant.⁵⁸ The fact that assumed or associated discrimination is covered by the directives and the fact that identification of a victim is not needed for discrimination to occur are two important reasons to comply to the non-discrimination directives when targeting job advertisements online.

4.1. Direct discrimination

4.1.1. Establishing direct discrimination

In order to prevent discrimination in online job advertising it is important to first establish in which ways it can be discriminatory. Targeting can be directly and indirectly discriminating. But what does direct discrimination exactly entail? The Gender Equality Directive defines direct discrimination as follows: “direct discrimination”: where one person is treated less favourably on grounds of sex than another is, has been or would be treated in a comparable situation”.⁵⁹ The prohibition on direct discrimination prohibits less favourable treatment based upon a protected ground. An important aspect of direct discrimination is that it is most often linked to overt practice.⁶⁰ When certain conduct amounts to direct discrimination, it will be obvious since the practice will manifest itself as discrimination. The discriminatory character immediately stands out when classic examples are presented. Paying women lower salaries for the same jobs or not hiring someone because of his or her sexual orientation are clear cases of direct discrimination.

While this interpretation seems straightforward, there are two issues that are connected with direct discrimination. First, the question whether direct discrimination addresses individuals or groups and second the question whether it should be seen from the perspective of the alleged victim or from the perspective of the alleged perpetrator.⁶¹ The answers to these questions can be found in the *CHEZ* case where the CJEU addressed the meaning of direct discrimination.⁶² The CJEU stated that the purpose of the directive is not only to protect individuals or groups who suffer from discrimination but also to end discriminatory practice based upon the protected grounds.⁶³ From *CHEZ* we see that the focus lies on the practice itself and reference to the grounds connected to the practice. The actual

⁵¹ See for more information: <https://www.facebook.com/about/privacy/update#what-kinds-of-information-do-we-collect> and <https://www.facebook.com/help/562973647153813/> both accessed on 22 June 2017.

⁵² Farkas (n 44) 21.

⁵³ Case C-303/06 *S Coleman v Attridge Law and Steve Law* [2008] ECR I-5603.

⁵⁴ *Coleman* (n 53) para 50.

⁵⁵ *English v Thomas Sanderson Blinds Ltd* [2009] 2 CMLR 18 <<http://www.bailii.org/ew/cases/EWCA/Civ/2008/1421.html>> accessed on 17 July 2017.

⁵⁶ Case C-54/07 *Centrum voor gelijkheid van kansen en voor racismebestrijding v Firma Feryn NV* [2008] ECR I-5187.

⁵⁷ *Ibid* para 25.

⁵⁸ *Ibid*.

⁵⁹ Gender Equality Directive art 2(1)(a); cf Racial Equality Directive art 2(2)(a); cf Employment Equality Directive art 2(2)(a).

⁶⁰ *Ellis and Watson* (n 24) 144.

⁶¹ Christopher McCrudden, ‘The New Architecture of EU Equality Law after *CHEZ*: Did the Court of Justice Reconceptualise Direct and Indirect Discrimination?’ [2016] 2016(1) European Equality Law Review 4.

⁶² Case C-83/14 *CHEZ Razpredelenie Bulgaria AD v Komisia za zashtita ot diskriminatsia* [2015].

⁶³ *Ibid* para 56.

effect or the existence of a victim does not influence the existence of direct discrimination.

In principle, a neutral practice will never amount to direct discrimination but it can possibly amount to indirect discrimination.⁶⁴ However, there is a deviation from this rule when ‘a measure is apparently neutral, but actually affects or is capable of affecting only persons possessing a protected characteristic.’⁶⁵ So when the distinction is based upon a characteristic which is inextricably linked to a protected characteristic, direct discrimination occurs.⁶⁶ A distinction based on pregnancy, for instance, constitutes direct discrimination based on the ground of sex since only women can become pregnant.⁶⁷ Because the practice does not necessarily have to refer to the protected ground to amount to direct discrimination, the following question should be asked: Would the person have been treated differently if he/she were of another sex, race, age or any other protected characteristic?⁶⁸ If the answer is affirmative, the less favourable treatment is caused by a discrimination ground because the treatment is based upon a protected characteristic.⁶⁹ The following question can be inferred: Would certain persons be included in the target audience and thus be able to see the advertisement if they did not possess a protected characteristic? If the answer is affirmative, direct discrimination occurs. This does not imply that targeting of all types of advertisements on discrimination grounds is illegal. However, in the case of job advertisements it is because access to employment is protected by the Directives.

So, in principle, targeting based on the protected grounds constitutes direct discrimination with regard to access to employment. Direct discrimination also occurs when a *prima facie* neutral targeting setting is inextricably linked to a protected ground and is only capable of affecting persons of that protected group. So, targeting settings which narrow down an audience by excluding people based upon protected characteristics amount to direct discrimination. For instance, if a job advertising campaign is set to only reach Caucasian people, Christian people or straight people, discrimination is a fact.

4.1.2. Justifying direct discrimination

Now that it is clear how targeting can be directly discriminating, it is time to delve into the justification methods provided by the directives. The directives provide limited and only specific defences to direct discrimination. This means that direct discrimination can only be justified when it is in pursuit of the particular aims which are set out in the directives.⁷⁰ The ‘genuine and determining occupational requirement’ (GOR) justification is found in all of the directives and is discussed here.⁷¹

To look into the implication of GORs for the automated targeting of job advertisement, an examination of the concept is required. The CJEU has a strict approach when interpreting defences to differential treatment. This would mean that exception based on a GOR should be interpreted narrowly.⁷² All three directives define the exception as follows:

*a difference of treatment which is based on a characteristic related to sex shall not constitute discrimination where, by reason of the nature of the particular occupational activities concerned or of the context in which they are carried out, such a characteristic constitutes a genuine and determining occupational requirement, provided that its objective is legitimate and the requirement is proportionate.*⁷³

The Directives provide that a difference of treatment shall not constitute discrimination when the characteristic is a genuine and determining occupational requirement, provided that the objective is legitimate and the requirement is proportionate. This means that it is allowed to differentiate based on a protected characteristic when such characteristic is directly related to the competence or suitability to perform the duties of that particular job.⁷⁴ An example of such exception is targeting advertisements solely at women if it concerns a female fashion modelling job.

Two things to note about the scope of GORs are relevant for automated job advertising. First, the GOR only needs to be related to a protected characteristic and does not have to be a protected characteristic itself.⁷⁵ Second, both the nature of the occupational activities and the context in which they are carried out may constitute a GOR.⁷⁶ In the *Johnston* case the Court stated that the nature of the work, reserve time police officer, did not constitute a GOR but the context, the possibility that women are more often an object of attack, did.⁷⁷ Targeting a job advertisement for fitness instructor in a women’s only gym at women can be justified because the context of the work constitutes a GOR even while the nature of the work itself entails that it can be done by both genders. On the other hand, the nature of the work results in a GOR in the case that a job advertisement for the role of Othello is targeted at black men.⁷⁸ GORs based on nature of the work often occur in artistic jobs like acting or modelling. GORs based on the context will mostly occur in institutions where one of the discrimination grounds plays a relevant role like in a female prison.⁷⁹

⁶⁴ See chapter 3.2.

⁶⁵ *CHEZ* (n 62), opinion of AG Kokott, para 82.

⁶⁶ *Ibid* para 86.

⁶⁷ Case C-177/88 *Elisabeth Johanna Pacifica Dekker v Stichting Vormingscentrum voor Jong Volwassenen (VJV-Centrum) Plus* [1990] ECR I-3941 para 12.

⁶⁸ FRA, *Handbook on European Non-Discrimination Law* (Publications Office of the European Union 2011) 26.

⁶⁹ *Ibid*.

⁷⁰ FRA (n 68) 43.

⁷¹ Ground specific justifications based on age religion are also provided by the directives but are not discussed due to the limited size of this paper.

⁷² FRA (n 68) 46.

⁷³ Gender Equality Directive art 14(2); cf Racial Equality Directive art 4(1); cf Employment Equality Directive art 4(1).

⁷⁴ FRA (n 68) 46.

⁷⁵ Gwyneth Pitt, ‘Genuine Occupational Requirements’ [2009] *Academy of European Law* 3–4 <http://www.era-comm.eu/oldoku/Adiskri/05_Occupational_requirements/2009_Pitt_EN.pdf> accessed on 18 June 2017.

⁷⁶ *Ibid* 4.

⁷⁷ Case 222/84 *Johnston v Chief Constable of the RUC* [1986] ECR 165.

⁷⁸ Othello is a male general of Moorish origin in the tragedy by William Shakespeare.

⁷⁹ Case 248/83 *Commission v Germany* [1985] ECR 1459 para 21.

From the preambles of the directives it becomes clear that exceptions only apply in ‘very limited circumstances’.⁸⁰ These very limited circumstances are met when the objective of the GOR is legitimate and proportionate. A legitimate aim is, for instance, the proper functioning and operational capacity of the fire service. The Court held that such an aim can be achieved by only recruiting persons under the age of thirty.⁸¹ The principle of proportionality demands that the derogation is appropriate and necessary to achieve the aim.⁸² The principle requires that there are no other measures that would achieve the same result while being less harmful. Proportionality furthermore entails that the exception needs to be transparent in nature.⁸³ If these criteria are met, the exception applies. Recruiting only persons under the age of twenty for the fire department would achieve the same legitimate aim. However, such age limitation is not proportionate since setting the age limit at thirty could achieve the same aim while being less discriminatory.

There is an important difference in the way GORs and targeting settings are linked to the job. GORs are inseparable from the job, without them the work cannot be performed the way it should be. Targeting settings are used on the one hand to filter out the people who cannot comply to those genuine requirements of the job. On the other hand, targeting is used to further narrow down the audience of suited people in order to reach only the best performing candidates. Reaching the best performing candidates is an economically efficient way of spending advertising budget and the employer will be more satisfied when he/she only receives applications from high quality candidates.

4.2. Indirect discrimination

4.2.1. Establishing indirect discrimination

Besides direct discrimination, the Directives include the concept of indirect discrimination which is not aimed at the practice itself but at the outcome. This section describes the concept and its justifications. The definition set out below is the same in all three directives.

*‘indirect discrimination’: where an apparently neutral provision, criterion or practice would put persons of one sex at a particular disadvantage compared with persons of the other sex, unless that provision, criterion or practice is objectively justified by a legitimate aim, and the means of achieving that aim are appropriate and necessary*⁸⁴

Hence, indirect discrimination requires an apparently neutral provision, criterion or practice that puts a certain protected group at a particular disadvantage unless it can be objectively justified. The terms ‘provision, criterion or practice’ show

that measures in the broadest meaning of the word fall within the scope of the definition.⁸⁵ The term an ‘apparently’ neutral practice should be understood as an ostensible, or *prima facie*, neutral practice, so having regard to factors which are not equivalent to the protected characteristics.⁸⁶ This is different from an apparently neutral practice that is inextricably linked to a protected characteristic which amounts to direct discrimination. Furthermore, the concept of ‘particular disadvantage’ does not refer to significant inequality but only to cases in which members of a protected group are affected more adversely than others.⁸⁷ Not being able to see the advertisement and thus not having the opportunity or at best having a smaller opportunity to apply to a job are adverse effects. From the *CHEZ* case follows that there is no need to assess the seriousness of not being able to see a job advertisement. However, the detrimental effect or disparate impact of the practice needs to be assessed.⁸⁸ It would be hard to argue that indirect discrimination occurs when there is no detrimental effect since an equal amount of female and male candidates in an audience are excluded by a neutral rule.

To establish indirect discrimination, a comparison between the effect on the protected group and on others needs to be made. It is vital that meaningful statistics are available for making this comparison.⁸⁹ The suited audience is determined with justifiable job requirements. Neutral targeting settings are used to further narrow the audience down to reach the perfect candidates. When a neutral setting has the effect of eliminating more people from a protected group than others from an audience of candidates, it has an indirectly discriminatory effect. But when is ‘more’ enough to amount to indirect discrimination?

The *Seymour Smith* case elaborates when an adverse impact is present in the case of sex discrimination.⁹⁰ According to the court this is the case when there is ‘a more unfavourable impact on women than on men’ and when ‘a considerably smaller percentage of women than men’ are affected by the, in this case beneficial, practice.⁹¹ So, the remaining percentage of people from a protected group must be considerably smaller than the percentage of people that remain from other groups. Or in other words, the exclusion effect that the targeting setting has on a protected group must be considerably larger than the effect it has on others.

A practical example: Imagine if in an audience of waiters, a certain neighbourhood which contains a lot of people of colour is excluded, the neutral rule has a discriminating effect when it removes a considerable percentage of the protected group.

⁸⁰ Employment Equality Directive recital 23; Racial Equality Directive recital 18; cf Gender Equality Directive recital 19.

⁸¹ Case C-229/08 *Wolf v Stadt Frankfurt am Main* [2010] ECR I-1, para 39–39.

⁸² *Johnston* (n 77) para 38.

⁸³ Case 318/86 *Commission v France* [1998] ECR 3559, para 25–27.

⁸⁴ Gender Equality Directive art 2(1)(b); cf Racial Equality Directive art 2(2)(b); cf Employment Equality Directive art 2(2)(b).

⁸⁵ Timo Makkonen, *Measuring Discrimination Data Collection and EU Equality Law* (Office for Official Publications of the European Communities 2007) 33.

⁸⁶ *CHEZ* (n 62) para 109; *CHEZ* (n 62), opinion of AG Kokott, para 93.

⁸⁷ *CHEZ* (n 62) para 109.

⁸⁸ Christa Tobler, *Limits and Potential of the Concept of Indirect Discrimination* (Office for Official Publications of the European Communities 2008) 30.

⁸⁹ *Ellis and Watson* (n 24) 155.

⁹⁰ Case C-167/97 *The Queen v Secretary of State for Employment, ex parte Nicole Seymour-Smith and Laura Perez* [1999] ECR I-623.

⁹¹ *Ibid* para 58–60.

However, the term considerable percentage needs to be defined in order to draw a practical conclusion.

The percentage of people removed should be compared to the effects of the neutral rule to the other protected groups. If 95% of the women are excluded but only 30% of the men, the neutral rule obviously has a discriminating effect. On the contrary, if a targeting setting removes 80% of the women, but also 80% of the men, it narrows the reached audience in an equal way. But what happens when 80% of the women are removed and 79% of the men? Is there a considerably smaller percentage of women that is left in this case? In order to answer this, we need to know when the scale tips from considerable to inconsiderable and vice-versa. One indication is that the degree in disparate impact must be quite high. However, numbers that define “quite high” are not available.⁹²

From the *Seymour Smith* case a statistical guideline can be inferred. This case however does not elaborate on when indirect discrimination occurs. To the contrary, it determines when indirect discrimination does not occur. In *Seymour Smith* the CJEU ruled that if a requirement could be fulfilled by 77.4% of the men and 68.9% of the women, it did not suffice for a disparate impact and that indirect discrimination did not occur as a result.⁹³ So, in *Seymour Smith* the requirement could be fulfilled by more men than women. Or in other words, more women than men were excluded by the requirement. To be precise, the requirement could be fulfilled by 12.3% more men than women.⁹⁴ From *Seymour Smith*, it can be inferred that this difference is not considerable enough to constitute a disparate impact. By analogy this can also be applied to the other grounds. So, when the difference is less than 12.3% it is certain that indirect discrimination does not occur. This does not mean that a difference of 12.4% does amount to discrimination. It is still not clear where the tipping point lies. However, using 12.3% as the maximum percentage of disparate impact would be the safest choice because it would be certain that discrimination would not occur.

4.2.2. “Justifying” indirect discrimination

Not in all cases where there is an adverse effect or disparate impact, can indirect discrimination be established. Since the justification of indirect discrimination is embedded in its definition, indirect discrimination does not occur at all when the practice can be justified. Unlike direct discrimination, indirect discrimination has an open-ended justification. The non-discrimination directives stipulate in the definition of indirect discrimination that it does not occur when the practice can be ‘objectively justified by a legitimate aim, and the means of achieving that aim are appropriate and necessary’.⁹⁵

Objective justification can be seen as ‘justification proper’, justification in the strict sense. Justification proper involves balancing the general goal to avoid discrimination and conflicting goals without a strict application of the principle of

proportionality.⁹⁶ This type of justification can be used in the same way as the genuine and determining occupational requirement justification can be used with direct discrimination.⁹⁷ Indirect discrimination therefore does not occur when a targeting setting serves the business need of hiring people with a certain educational or professional background. So, it would not amount to indirect discrimination to target a medical job vacancy at people who went to medical school, regardless of any discriminatory effects of the setting.

On the other hand, objective justification can be seen as a causal link test. This causation approach can be used to rebut the assumption that there is a link between the practice and the detrimental effect.⁹⁸ The existence of detrimental effect or factual disadvantage only creates the presumption of indirect discrimination and this presumption can be rebutted arguing a missing causal link between the detrimental effect and the practice. In the case that an audience of mechanics consists for a considerably larger percentage out of men, a causal link defence would state that this is not because of the used targeting settings but because there are in fact more male than female mechanics. In other words, the cause lies somewhere else, not with the targeting setting.

According to the non-discrimination directive, a practice can be objectively justified if it has a legitimate aim and at the same time is necessary and appropriate. So, the first question is: when does a practice have a legitimate aim? General indications can be found in the CJEU’s case law on sex discrimination.⁹⁹ The CJEU stated that in order to be objectively justified, the difference in treatment must be based upon factors which are unrelated to any discrimination on grounds of sex.¹⁰⁰ The factors must also correspond to a real need on the part of the undertaking.¹⁰¹ But since objective justification is open ended, there are lots of possible justification grounds. On the other hand, there are important limits. Considerations that are purely budgetary cannot serve as an objective justification.¹⁰² So, even though budgetary considerations may influence a particular practice, they cannot by themselves justify indirect discrimination. For instance, excluding older people from the audience only because they are more expensive to employ cannot be justified. Furthermore, a mere generalisation is insufficient to prove that the practice is unrelated to such ground.¹⁰³ This means that there needs to be assessed whether the legitimate aim is actually being pursued. How this assessment needs to be done becomes clear from the judgement in the *CHEZ* case. A company needs to factually establish the problem and establish that the problem will continue if the measure is not taken.¹⁰⁴ With the targeting of job advertisements, the problem would be reaching people who are not suited for the job. It would be

⁹² Tobler (n 88) 31.

⁹³ *Seymour Smith* (n 90) para 63.

⁹⁴ $\frac{77.4\%}{68.9\%} = 112.3\%$, so 12.3% more.

⁹⁵ Gender Equality Directive art 2(1)(b); cf Racial Equality Directive art 2(2)(b); cf Employment Equality Directive art 2(2)(b).

⁹⁶ Schiek, Waddington and Bell (n 38) 436.

⁹⁷ Ibid 437.

⁹⁸ Ibid 439.

⁹⁹ Tobler (n 88) 32.

¹⁰⁰ Case 170/84 *Bilka-Kaufhaus GmbH v Karin Weber von Hartz* [1986] ECR 1607, para 30; Case C-196/02 *Vasiliki Nikoloudi v Organismos Tilepikoinonion Ellados AE* [2005] ECR I-1789, para 47.

¹⁰¹ Tobler (n 88) 32.

¹⁰² *Nikoloudi* (n 100) para 53.

¹⁰³ *Seymour-Smith* (n 90) para 76.

¹⁰⁴ *CHEZ* (n 62) para 116–122.

a waste of effort and advertising budget to target a medical vacancy at people who have had no medical education.

After it is confirmed that the practice has a legitimate aim, it needs to be examined if the practice is necessary and appropriate, ergo the proportionality test has to be done.¹⁰⁵ If this test fails, the practice cannot be justified and it will amount to indirect discrimination. In the *Mangold* case the court stated that ‘observance of the principle of proportionality requires every derogation from an individual right to reconcile, so far as is possible, the requirements of the principle of equal treatment with those of the aim pursued’.¹⁰⁶ This means that there should be no feasible alternatives that protect the principle of equality better. If we focus on the conduct that should justify the discriminatory effect we must note that the fact that a practice or measure is desirable or convenient is not sufficient, it needs to be appropriate. Hence, it must be suitable to achieve the particular aim.¹⁰⁷ A measure is not suitable to achieve the particular aim and thus not appropriate when the practice excessively prejudices the legitimate interest of the persons concerned.¹⁰⁸ The measure also needs to be necessary, this means that there should be no other available measures which can achieve the same result and cause a less discriminatory effect.¹⁰⁹ When a low level administrative vacancy, targeted at people who have studied at university level in order to reach high performing candidates, has a discriminatory effect, the settings cannot be justified because the measure is not proportionate since there are other less discriminatory targeting settings available which would achieve the same aim.

Targeting settings that reflect occupational requirements serve a legitimate aim since they are necessary to perform the job. This means that targeting settings based on educational level, educational subject, professional background, job title or work industry cannot amount to indirect discrimination. So, a targeting setting which narrows an audience down to people who have studied economics at university level and who work in management in the steel industry can always be used. Any discriminatory effect of such setting is justified by the needs of the employer.

The conclusion is that even with the use of neutral targeting settings, there still is a risk of discrimination. However, if this neutral setting is objectively justified, the practice will not constitute indirect discrimination. This is the case when the targeting setting has a legitimate aim, is appropriate and necessary. This means that the setting does not excessively prejudice the legitimate interests of the persons concerned, in this case the opportunity to apply to jobs. Furthermore, there should be no other targeting settings available that cause a less discriminatory effect and achieve the same result.

¹⁰⁵ Álvaro Oliveira and Sarah-Jane King, ‘In Its First Judgment on Discrimination against Roma People, the EU Court of Justice Sheds New Light on How to Interpret the Concepts of Direct and Indirect Discrimination Based on Ethnic Origin under’ (2017) 41 European Law Review 865, 881.

¹⁰⁶ *Mangold* (n 46) para 65.

¹⁰⁷ *Tobler* (n 88) 35.

¹⁰⁸ Oliveira and King (n 105) 881.

¹⁰⁹ Tamara K Hervey, *EC Law on Justifications for Sex Discrimination in Working Life* (Conference Paper 2002) 121 <http://www.juridicum.su.se/stockholmcongress2002/hervey_english.pdf> accessed on 5 July 2017.

5. Preventing discrimination in the automated targeting of job advertisements

This section provides recommendations on how direct and indirect discrimination can be prevented when using AI in the automated job advertising process and it builds on the concepts discussed in sections two and four. When the targeting is done by AI, the factors that influence its targeting decisions are the data contained in the big data sets, the algorithms used for the data mining, the rules that are learned from this and the use of those rules by the AI. Consequently, these are also the factors that determine if an AI-driven job advertising campaign is discriminatory. By manipulating these factors discrimination can be prevented. It is important to prevent discrimination and to allow for the exceptions that justify discrimination at the same time.

5.1. Preventing direct discrimination

With regard to direct discrimination, this means that the use of the protected grounds as targeting settings must be prevented while allowing the use of protected grounds that are genuine occupational requirements. So, if a protected ground is not explicitly required to fulfil the job, it should not be used as a targeting setting. Furthermore, the use of characteristics which are inextricably linked to a protected ground must be prevented. Consequently, the AI must recognize these characteristics in order to prevent them from being used.

The first factor that could be influenced is big data. If the discrimination grounds are excluded from the big data, the algorithms can never pick them up.¹¹⁰ If the big data does not contain the protected characteristics, correlations which include those characteristics will never be found and thus no discriminating rules would be created for the AI to use. However, excluding information from big data requires data pre-processing, which is costly.¹¹¹

Another solution would be the use of algorithmic models which do not contain the protected characteristics. In this way, a blind spot is created on purpose and the protected grounds are not used by the algorithm when it searches for correlations.¹¹² Eliminating the grounds from the big data or the algorithm would work perfectly for the grounds race, ethnicity, sex, sexual orientation, religious belief, age and disability which are set out in the directives. However, a problem arises since AI would still be able to use targeting settings that are apparently neutral but are only capable of affecting persons possessing a protected characteristic. Such settings, those that are inextricably linked to the protected grounds, also amount to direct discrimination.¹¹³ Manipulating the big data or the algorithms thus will not result in a sufficient solution.

¹¹⁰ Sara Hajian and Josep Domingo-Ferrer, ‘A Methodology for Direct and Indirect Discrimination Prevention in Data Mining’ (2013) 25 IEEE Transactions on Knowledge and Data Engineering 1445.

¹¹¹ Zhou and others (n 16) 352.

¹¹² O’Neil (n 3) 29.

¹¹³ *CHEZ* (n 62), opinion of AG Kokott, para 82; *Dekker v VJV* (n 67) para 12.

The solution to the problem of recognising characteristics which are inextricably linked to the protected grounds can be found in the use of data mining. When using the big data for more than the single purpose of discovering correlations which result in the most successful targeting rules, other correlations which are useful for the prevention of discrimination can be found. The big data can be used to determine which neutral grounds are inextricably linked to protected grounds. In order to do so, both the big data and the algorithms must include the protected grounds. When this is done, the algorithms could find, for instance, that 100% of the pregnancies are correlated to women, that 100% of the people with facial hair are men or that 100% of the people that wear burkas are Muslim. The fact that something correlates with a protected ground for 100% proves that it is inextricably linked to that ground because it is connected to the ground in every case. With this information, a list of prohibited targeting settings can be established, e.g., a blacklist. This blacklist contains the grounds protected by the directives and all characteristics which are inextricably linked to it. This blacklist must be used as a filter between the established rules or correlations and the output of the AI. In this way, the AI is able to identify discriminatory targeting settings and it will not use settings that are directly discriminating.

The exception of GORs should also be added. This means that the AI must know in what circumstances it should allow the use of an otherwise directly discriminating targeting setting. For the AI to determine if the use of a protected ground can be justified because it constitutes a genuine occupational requirement, it must have knowledge of the context and the nature of the work that has to be performed. Since every job vacancy is different it is hard to draft general rules to assess when a protected characteristic is a GOR. Even if this is technically feasible it would still require the AI to have access to each employer's system to gather information about the context and nature of the work. It is unlikely that every employer will give this access and if they would, it would be a huge operation to connect the AI to every employer's system and to preprocess all their data to the right format to be used.¹¹⁴

Another option would be to exclude the justification of GORs at all. Since they are an exceptional justification and require a narrow interpretation, they will not be applicable in a lot of cases.¹¹⁵ So, excluding them will not have a big impact on the efficiency of the AI. In this case a bit of efficiency is sacrificed in order to limit the chance of discrimination.

The third option is the use of machine learning and rule induction. With these techniques, the AI can learn to recognise real life situations and predict their outcomes based on observational data from previous events.¹¹⁶ The AI is able to recognise when the employer provides a job requirement which would be directly discriminating. If this is the case a human must intervene to determine if the asked requirement con-

stitutes a GOR. The AI could learn by induction to recognize which GORs are connected to which jobs based on the decisions made by the human. In this way, it can label certain job profiles as allowing certain discrimination grounds. It could learn for instance that there are male and female fashion models and allow for the targeting based on gender in these cases.

5.2. Preventing indirect discrimination

As for indirect discrimination, the use of neutral targeting settings which have an indirectly discriminating effect must be prevented while allowing the use of such settings if they can be objectively justified. The definition of indirect discrimination entails that neutral targeting settings that are objectively justifiable cannot amount to indirect discrimination. When this rule is applied to targeting, a distinction between two different types of targeting needs to be made.

First, there are targeting settings that entail occupational requirements which are determined by the employer. This is a different concept than the GORs connected to the justification of direct discrimination. In the context of indirect discrimination, targeting based upon occupational requirements should be seen as neutral rules that reflect the genuine needs of the employer. Occupational requirements are objectively justifiable because they serve the legitimate aim of employing a qualified person. So, as long as these requirements are necessary and appropriate, targeting settings based on educational level, educational subject, professional background, job title or work industry cannot amount to indirect discrimination. It is therefore the responsibility of the employer to provide the advertising company that uses the AI with legitimate requirements. Facebook provides categories of targeting settings in the exact categories as mentioned above. These categories should be put on a green list and should always be allowed. So, when the rule picking element of the AI is manipulated in this way, there is no need to assess the effect of these settings since discrimination cannot occur either way. For instance, an advertisement for a programmer targeted at people who studied at university level and have programmer as current job title should always be allowed.

The second type of targeting settings are the additional settings that further narrow down the qualified audience established with the use of occupational targeting setting. The second type of settings is determined by the AI based on correlations found in the big data. The big advantage of this is that the AI will come up with effective targeting settings which humans would never think of. For instance, AI found through data mining that an affinity for a Japanese cartoon site is a solid predictor of good coding skills.¹¹⁷ So the AI could target the programmer advertisement at people who studied at university level and have programmer as current job title and like the particular Japanese cartoon site. In this situation, the first two settings are occupational requirements which are necessary to perform the job, the last setting is not. Consequently, this last targeting setting cannot be objectively justified since it is not necessary to perform the job and thus not objectively justifiable. Therefore, it can only be used if it is not indirectly

¹¹⁴ Zhou and others (n 16) 352.

¹¹⁵ Employment Equality Directive recital 23; Racial Equality Directive recital 18; cf Gender Equality Directive recital 19; FRA (n 68) 46.

¹¹⁶ Pat Langley and Herbert a Simon, 'Applications of Machine Learning and Rule Induction' (1995) 38 Communications of the ACM 54, 55–57 and 62.

¹¹⁷ Peck (n 20).

discriminating. So, with regard to the settings which are not on the green list, the discriminatory effect needs to be predicted.

To predict the discriminatory effects of a setting, the AI needs to understand the concept of detrimental effect or disparate impact. Such effect or impact exists when a considerably larger percentage of a protected group is excluded than others by the targeting setting. From the *Seymour Smith* case, it can be inferred that indirect discrimination does not occur when the relative difference in exclusion is 12.3% or less. So, the AI should compare the predicted impact of the targeting settings on protected groups and others and use this as a rule when comparing the effects. Through its APIs it can see what kind of effect each targeting setting will have on the audience. If the setting has an adverse effect on a protected group, it may not be used.

The following simplified examples will make this clear. If an advertisement for a management function is targeted at people who work as managers in the Netherlands, 92.000 women and 108.000 men are reached.¹¹⁸ Let us assume, for the sake of the example, that it follows from the big data that an affinity for entertainment correlates with strong managing skills. If we add this as a targeting setting 70.500 women are reached and 74.500 men are reached. So, 76.7% of the women can fulfil the requirement and 69.0% of the men. The relative difference in exclusion is 11.2% which is less than the *Seymour Smith* rule of 12.3% so no indirect discrimination occurs. If from the big data follows that an affinity for beer correlates with strong managing skills the audience is distributed in a more unequal way. The audience now consists out of 15.750 women and 29.250 men. So, 17.1% of the women are able to fulfil this requirement and 27.1% of the men. In this case 58.5% more men than women can fulfil the requirement which is more than the *Seymour Smith* rule. In the last case, indirect discrimination does occur and the targeting setting should not be used.

6. Conclusion

On the background of the increasing amount of discriminatory challenges facing AI applications, this paper examined the requirements needed in order to comply to European non-discrimination law to prevent discrimination in the automated online job advertising business in Europe. The factors that influence the occurrence of discrimination are the big data, the algorithms that mine the big data, the correlations that are found and the accompanying targeting rules, and the way AI uses these rules. Job advertising companies using artificial intelligence to target their job advertisement should comply to the European non-discrimination directives since member states have to implement the directives, national courts must interpret national law in light of the directives and individuals can directly invoke the principle of non-discrimination. From the directives follows that targeting may not be directly or indirectly discriminating on the grounds of race, ethnicity, sex, sexual orientation, religious belief, age and disability. Direct dis-

crimination occurs when a targeting setting is used that contains such ground or a characteristic that is inextricably linked to it. Indirect discrimination occurs when a neutral setting excludes a considerable larger percentage of a protected group compared to others. Direct discrimination can only be justified by a genuine occupational requirement while indirect discrimination has an open justification. So, targeting settings which reflect occupational requirements cannot amount to discrimination. These justifications can be implemented within an AI by establishing a black list with targeting settings that cannot be used at all to prevent direct discrimination. A green list with settings which can always be used needs to be established to allow for neutral settings which are justified. To prevent indirect discrimination the effect of a setting that is not justifiable needs to be predicted. Only if the setting does not exclude a considerably bigger percentage of a protected group, it cannot amount to indirect discrimination and it can be used. When these requirements as guided by the EU directives and with the proved conditions of either the black or green lists are implemented in the AI that makes the targeting decision, discrimination in the automated online targeting of job advertisements can be prevented.

Acknowledgements

I would like to thank Kevin Jon Heller, Maarten den Heijer and Frederik Zuiderveen Borgesius for proofreading the paper, for suggestions and for advice. I would also like to thank Simone van Beek for language help. Any errors are mine.

REFERENCES

Literature

- Aggarwal CC. Data mining: the textbook. 1st ed. Springer; 2015.
- Angwin J, Mattu S, Kirchner L. Machine Bias: There's Software Used across the Country to Predict Future Criminals. And It's Biased against Blacks. *ProPublica* (2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- Bleier A, Eisenbeiss M. Personalized online advertising effectiveness: the interplay of what, when, and where. *Market Sci* 2015;34:669.
- Chen M, Mao S, Leung VCM, Zhang Y. Big data: related technologies, challenges and future prospects. 1st ed. Springer; 2014.
- Clifford J. Equality. In: Shelton D, editor. *The Oxford handbook of international human rights law*. 1st ed. Oxford University Publishing. 2013.
- Cole WM. Mind the Gap: state Capacity and the Implementation of Human Rights Treaties. *Int Organ* 2015;69:405.
- Datta A, Tschantz MC, Datta A. Automated experiments on ad privacy settings: a tale of opacity, choice, and discrimination. *Proc Privacy Enhanc Technol* 2015;2015:92.
- De Mol M. The novel approach of the CJEU on the horizontal direct effect of the EU principle of non-discrimination: (Unbridled) Expansionism of EU law? *Maastricht J Eur Comp Law* 2011;18:109.

¹¹⁸ These numbers are generated by Facebook's Audience Insights prediction tool on 24 June 2017.

- De Witte B. Direct effect, primacy and the nature of the legal order. In: De Burca G, Craig PP, editors. *The evolution of EU law*. 2nd ed. Oxford University Press; 2011.
- Ellis E, Watson P. *EU anti-discrimination law*. Oxford Scholarship Online; 2013.
- Farkas L. How to present a discrimination claim: handbook on seeking remedies under the EU non-discrimination directives. Publications Office of the European Union; 2011.
- FRA. *Handbook on European non-discrimination law*. Publications Office of the European Union; 2011.
- Hajian S, Domingo-Ferrer J. A methodology for direct and indirect discrimination prevention in data mining. *IEEE Trans Knowl Data Eng* 2013;25:1445.
- Hervey TK. *EC Law on Justifications for Sex Discrimination in Working Life* (Conference Paper 2002) Available from: http://www.juridicum.su.se/stockholmcongress2002/hervey_english.pdf.
- Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science* 2015;349:255.
- Khosrowpour M. *Dictionary of science and technology*. 2nd ed. Hershey; 2013.
- King AG, Mrkonich M. Big data and the risk of employment discrimination. *Oklahoma Law Rev* 2014;68:555.
- Langley P, Simon H. Applications of machine learning and rule induction. *Commun ACM* 1995;38:54.
- Makkonen T. Measuring discrimination data collection and EU Equality law. Office for Official Publications of the European Communities; 2007.
- McCrudden C. The New Architecture of EU Equality Law after CHEZ: Did the Court of Justice Reconceptualise Direct and Indirect Discrimination? [2016] *European Equality Law Review*.
- Munoz C, Smith M, Patil D. *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights* (2016) Available from: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf.
- Oliveira Á, King S-J. In Its First Judgment on Discrimination against Roma People, the EU Court of Justice Sheds New Light on How to Interpret the Concepts of Direct and Indirect Discrimination Based on Ethnic Origin under. *Eur Law Rev* 2017;41:865.
- O'Neil C. *Weapons of math destruction: how big data increases inequity and threatens democracy*. 1st ed. Crown; 2016.
- Peck D. They're Watching You at Work. *The Atlantic* (2013) Available from: <http://www.theatlantic.com/magazine/archive/2013/12/theyre-watching-you-at-work/354681/>.
- Pitt G. Genuine Occupational Requirements' [2009] *Academy of European Law*. Available from: http://www.era-comm.eu/oldoku/Adiskri/05_Occupational_requirements/2009_Pitt_EN.pdf.
- Podesta J, Pritzker P, Moniz EJ, Holdren JP, Zients J. *Big Data: Seizing Opportunities* (2014; Available from: https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf.
- Post P, Holtmaat R. A false start: discrimination in job advertisements. *Eur Gender Equality Law Rev* 2014;2014/1:12.
- Rodger S, Thorson E. *Digital advertising: theory and research*. 3rd ed. Routledge; 2017.
- Russell SJ, Norvig P. *Artificial intelligence: a modern approach*. Prentice Hall; 2003.
- Schiek D, Waddington L, Bell M. *Cases, materials and text on National, Supranational and International Non-Discrimination Law*. 1st ed. Hart; 2007.
- Tobler C. Limits and potential of the concept of indirect discrimination. Office for Official Publications of the European Communities; 2008.
- Weatherill S. *Law and values in the European Union*. 11th ed. Oxford University Press; 2016.
- Zhou L, Pan S, Wang J, Vasilakos AV. Machine learning on big data: opportunities and challenges. *Neuroscience* 2017;237:350.

Cases

- Case 26/62 *NV Algemene Transport- en Expeditie Onderneming van Gend & Loos v Netherlands Inland Revenue Administration* [1963] ECR 1.
- Case 6/64 *Flaminio Costa v ENEL* [1964] ECR 585.
- Case 248/83 *Commission v Germany* [1985] ECR 1459.
- Case 170/84 *Bilka-Kaufhaus GmbH v Karin Weber von Hartz* [1986] ECR 1607.
- Case 222/84 *Johnston v Chief Constable of the RUC* [1986] ECR 165.
- Case 318/86 *Commission v France* [1998] ECR 3559.
- Case C-177/88 *Elisabeth Johanna Pacifica Dekker v Stichting Vormingscentrum voor Jong Volwassenen (VJV-Centrum) Plus* [1990] ECR I-3941.
- Case C-167/97 *The Queen v Secretary of State for Employment, ex parte Nicole Seymour-Smith and Laura Perez* [1999] ECR I-623.
- Case C-196/02 *Vasiliki Nikoloudi v Organismos Tilepikoinonion Ellados AE* [2005] ECR I-1789.
- Case C-144/04 *Werner Mangold v Rüdiger Helm* [2005] ECR I-9981.
- Case C-54/07 *Centrum voor gelijkheid van kansen en voor racismebestrijding v Firma Feryn NV* [2008] ECR I-5187.
- Case C-555/07 *Seda Küçükdeveci v Swedex GmbH & Co. KG* [2010] ECR I-365.
- Case C-229/08 *Wolf v Stadt Frankfurt am Main* [2010] ECR I-1.
- English v Thomas Sanderson Blinds Ltd* [2009] 2 CMLR 18. Available from: <http://www.bailii.org/ew/cases/EWCA/Civ/2008/1421.html>.
- Joined cases C-335/11 and C-337/11 *HK Danmark, acting on behalf of Ring v Dansk Almennyttigt Boligselskab; HK Danmark, acting on behalf of Skouboe Werge v Dansk Arbejdsgiverforening, acting on behalf of Pro Display A/S* [2013], ECLI:EU:C:2013:222.
- Case C-83/14 *CHEZ Razpredelenie Bulgaria AD v. Komisija za zashtita ot diskriminatsia* [2015].