When Computers Say No: Towards a Legal Response to Algorithmic Discrimination in Europe

Raphaële Xenidis

When computers say no: towards a legal response to algorithmic discrimination in Europe

Raphaële Xenidis[1]

1. INTRODUCTION

Concerns over breaches of fundamental rights arising from the deployment of algorithmic systems have increased in recent years. In particular, research across the globe shows that algorithmic systems used in various decision-making processes can discriminate against legally protected groups. For instance, in a landmark decision the Italian *Tribunale di Bologna* found that the reputational ranking algorithm used by the delivery platform Deliveroo to give riders access to a system for booking working shifts was indirectly discriminatory.[2] In deciding which riders to prioritise, the system constructed a measure of their 'reliability' and 'participation' that did not take into account legally protected reasons for absence from work such as strike actions, illness, disability, personal beliefs, or care duties (still performed by women in majority). By treating all cancellations of work shifts indistinctly, the system unfairly limited riders' work opportunities. In Austria, the so-called 'AMS' algorithm was commissioned by the national employment agency to grant or withhold job seeker support based on a prediction of their chances of finding employment. Researchers showed that in some versions, the predictive system assigned a negative weight to female job candidates (in particular, when they had care duties[3]) and that it took into account features such as candidates' migration background, health impairments and age, thus making the system potentially discriminatory against legally protected groups (Kayser-Bril, 2019; Alhutter et al., 2020). Research has brought to light numerous other examples of algorithmic discrimination in Europe (for a recent overview, see Wulf, 2022).

To a certain extent, anti-discrimination laws in place in Europe can address algorithmic discrimination. Yet, thorny questions arise regarding the interpretation and the application of these laws. Existing legislation also exhibits gaps and shortcomings, especially in the context of machine learning systems. This chapter examines these problems and proposes reflections on how to enforce equality in the algorithmic society. To do so, it first scrutinises the roots and mechanics of algorithmic discrimination and proposes working definitions with the aim of disentangling existing semantic confusions. Second, this chapter investigates the shortcomings of the existing anti-discrimination law framework, distinguishing between regulatory, conceptual, doctrinal and procedural gaps. Finally, this chapter proposes some reflections on enforcing (algorithmic) equality. In so doing, this chapter reflects on the normative implications of different possible interpretations of the legal framework in light of the problem of algorithmic discrimination.

2. FROM ALGORITHMIC BIAS TO ALGORITHMIC DISCRIMINATION

2.1 How Does Algorithmic Discrimination Arise?

The well-known phrase 'garbage in, garbage out', recast by Mayson as 'bias in, bias out' (Mayson, 2018), places the focus on data as the origin of algorithmic discrimination. Because structural inequalities are ingrained in any social data, as the aggregated product of past discriminatory decisions, learning algorithms internalise and re-enact such patterns of inequality. Examples such as the now infamous Amazon CV screening prototype, which learnt from past hiring decisions to systematically discriminate against female job candidates, show that data-driven discrimination is a reality (Dastin, 2018). However, algorithmic discrimination can also originate elsewhere. As illustrated by the story of Dr Selby, a gym customer who could not access the women's changing room because the system associated the prefix 'Dr' with male rather than female clients, stereotypes also pervade problem definition and model design (Turk, 2015). At the operationalisation stage too, algorithmic discrimination can arise. Human agents display 'disparate interactions' with the output of algorithmic decision support systems, for example, overestimating the risks posed by racialised groups and underestimating those posed by majority groups in pretrial release decisions (Greene & Chen, 2019). Hence, the sources of algorithmic discrimination are multiple and difficult to disentangle. Ultimately, discrimination is likely to result from complex 'co-production' processes at the intersection of technological deployment, social practices and political objectives (Alhutter et al., 2020).

Because of narratives emphasising data as the source of algorithmic discrimination, policy discussions on how to address the problem have given particular attention to the accuracy of training datasets.[4] For example, Article 10 of the draft EU AI Act foresees quality requirements for data collection, data preparation and processing, and the identification of data gaps. Ensuring that training and validation data is representative is an important aspect of addressing algorithmic discrimination. So-called 'accuracy-affecting injustices' can indeed bias data collection and explain why some algorithmic systems underperform for, and underserve, certain population groups (Hellman, 2021). For instance, collecting data via users' internet access can lead to certain communities being under-represented in the data collected, i.e. older users or residents of rural or economically deprived areas where the internet infrastructure is underdeveloped. In turn, an algorithmic system trained with that data might not adequately account for the needs and behaviours of these communities, therefore potentially leading to injustices and disadvantages. Such harms can in part be addressed by improving the representativeness and accuracy of training and validation data.

However, when 'non-accuracy-affecting injustices' are responsible for algorithmic discrimination, measures improving data collection or quality are ineffective (Hellman, 2021). For instance, if an algorithmic system used to calculate workers' pay was trained on average pay data across Europe, the output would likely exhibit a difference between men's and women's pay. This difference corresponds to the gender pay gap, which is about 13% on average in Europe.[5] The data used to train the system is factually correct, but it reflects historical injustices. Addressing algorithmic discrimination in this case requires treating not only the symptoms (data representativeness) but also the roots (gender inequality) of such disadvantage. Policy and legal responses therefore need to go beyond requiring data accuracy, quality and transparency and to also make use of measures, such as positive action, that address the structural causes of algorithmic discrimination.

2.2 Clearing Some Semantic Confusions

Two main strands of disciplinary semantics coexist and can give rise to confusion. On the one hand, computer science and ethics literature mainly use the terms 'bias' and 'fairness' to capture the harms and injustices of algorithmic systems and the means available to address them. On the other hand, the legal literature qualifies unlawful algorithmic distinctions between groups as 'discrimination' and frames means of redress in terms of 'equal treatment'. Juxtaposing the two disciplinary frameworks raises difficult questions: How do notions of bias and fairness map onto equality law? In other terms, when does bias qualify as discrimination? What bias is unlawful? And what fairness metrics does equality law require?

The draft EU AI Act offers an interesting case study for interrogating the overlaps and differences between these different terms. The regulatory proposal implicitly equates addressing algorithmic bias with preventing algorithmic discrimination. At first sight, it gives the impression that algorithmic discrimination takes on an important place in the regulatory apparatus that it sets up. Both the explanatory memorandum and Recital 28 indicate that when classifying an AI system as high-risk, it is of particular relevance to consider '[t]he extent of the adverse impact caused by the AI system on the fundamental rights protected by the Charter' including 'non-discrimination' (Art. 21 EUCFR) and 'equality between women and men' (Art. 23 EUCFR). Recitals 35, 36 and 37 warn that AI systems used in core sectors such as education, employment and essential services are liable to 'violate […] the right not to be discriminated against' and 'perpetuate historical patterns of discrimination'.[6] Recital 44 explicitly refers to non-discrimination law when stressing the importance of high-quality data requirements to ensure that a high-risk AI system 'does not become the source of discrimination prohibited by Union law'.

Despite these numerous acknowledgements of the problem of algorithmic discrimination in the explanatory memorandum and the preamble, the binding part of the proposal mainly uses the term 'bias' and only mentions 'discrimination' twice.[7] Yet the polysemy and broad scope of the term 'bias' (e.g. Friedman & Nissenbaum, 1996) paves the way for legal uncertainty concerning the interpretation of the obligations falling on providers and users of algorithmic systems.[8] The EU's High-Level Expert Group on Artificial Intelligence defines algorithmic bias as 'systematic and repeatable errors in a computer system that create unfair outcomes, such as favouring one arbitrary group of users over others'.[9] This definition is much broader than the definition of discrimination, which, in EU law, captures unfair outcomes only when they harm protected groups within certain contexts that fall within the scope of non-discrimination law. In other words, anti-discrimination law does not address bias itself, but rather some of the harms which it creates for legally protected groups.

EU anti-discrimination law creates a harmonised set of minimum requirements for the 27 EU member states as well as EEA countries, candidates for EU membership and countries that have approximated their legislation to EU equality law. It sets three conditions for algorithmic bias to amount to discrimination. First, algorithmic bias has to create harm or disadvantage to a protected group or based on a protected ground. The personal scope of EU anti-discrimination law includes race or ethnic origin, sex or gender, disability, religion or belief, sexual orientation and age.[10] Second, EU law addresses algorithmic bias only in certain areas of life, including work and access to certain goods and services.[11] Third, to qualify as discrimination, algorithmic bias must fall within the central dichotomy of EU anti-discrimination law. If a protected group or category is treated differently from others, it qualifies as direct discrimination.

That would be the case for example when an algorithmic system used to screen CVs learns to use candidates' ethnic background as a predictor of lesser performance. Alternatively, bias creates a particular disadvantage to a protected group without using the protected characteristic as a decision-making factor and qualifies as indirect discrimination. That might be the case if, for example, an algorithmic credit scoring system used predictors such as income or employment history, which are facially neutral towards protected groups, but that still have a disadvantageous impact on women. While direct discrimination cannot be justified in principle, indirect discrimination comes with an open-ended justification regime. A *prima facie* indirectly discriminatory provision, criterion or practice can be objectively justified if it serves a legitimate aim and the means of achieving that aim are appropriate and necessary.[12]

Mapping algorithmic bias onto the EU anti-discrimination framework yields a complex picture where the harms deriving from algorithmic bias only qualify as legally prohibited discrimination when fulfilling the above three conditions linked to EU anti-discrimination law's personal, material and conceptual scope. Mapping fairness onto the legal framework is also a difficult task. The principle of equality is a contextually moving target, especially when courts are called to conduct a proportionality test to assess justifications. This raises the question of which definitions of fairness satisfy the requirements of the principle of equal treatment and under what conditions (Weerts, Xenidis, Tarissan, Palmer Olsen & Pechenizkiy, 2023). The case law of the ECJ shows that equal treatment cannot translate into a one-size-fits-all fairness formula because anti-discrimination law is polysemous and assumes different social and legal functions (Xenidis, 2021b).

3. THE EU EQUALITY PUZZLE: WHAT GROUPS ARE PROTECTED FROM ALGORITHMIC DISCRIMINATION AND WHEN?

How does EU anti-discrimination law apply to algorithmic discrimination? Answering this question brings to light a series of regulatory, conceptual, doctrinal and procedural gaps and challenges. Not only do the specific mechanics of algorithmic discrimination create uncertainty regarding the interpretation and application of EU anti-discrimination rules, often exacerbating existing tensions, but they also question the frontiers of EU anti-discrimination law. Thus, certain forms of algorithmic discrimination could fall into the cracks of EU antidiscrimination law.

3.1 A Patchy Material Scope

As hinted above, EU anti-discrimination law is a regulatory puzzle that combines provisions of EU primary and secondary law. Article 19 TFEU gives the EU power to adopt legislation prohibiting discrimination on grounds of sex, racial or ethnic origin, religion or belief, disability, age, and sexual orientation. Article 157 TFEU guarantees the equal treatment of men and women at work, especially with regard to pay. In the EU Charter of Fundamental Rights, Article 21 provides for a non-exhaustive list of protected criteria including, but also going beyond, those listed in Article 19 TFEU. In turn, Article 23 of the Charter ensures equality between men and women. In addition to these provisions

of primary law, four equality directives prohibit discrimination on grounds of sex, race or ethnic origin, religion or belief, disability, age, and sexual orientation.[13] Even though these directives pertain to discrimination in general, their application can be extended to algorithmic discrimination in particular.[14]

Despite a seemingly broad personal scope, the application of EU law is not equally extensive. With a broad brush, Directive 2000/43 applies to discrimination on grounds of race or ethnic origin in the areas of employment, goods and services, social protection, and education. Directive 2004/113 and Directive 2006/54 offer a similar protection in relation to sex discrimination.[15] However, discrimination on grounds of religion or belief, disability, age and sexual orientation is only prohibited in matters related to employment and occupation. This means that an algorithmic system that would exclude, for example, end users above a certain age or with a certain religious affiliation from accessing given services or from purchasing certain goods would not, in principle, be contrary to EU secondary anti-discrimination law as it stands.[16] In addition to these gaps in the material scope of EU anti-discrimination law, further exceptions exist, which might negatively impact EU law's grasp on algorithmic discrimination, such as the fact that the content of media and advertising is not covered by the ban on sex discrimination.[17] When considering the pervasiveness of algorithmic systems in the market for goods and services, these gaps seriously undermine the robustness of the existing framework. Importantly, however, EU law only foresees minimum requirements and member states are in principle free to adopt a higher level of protection against discrimination.

3.2 Personal Scope: Where to Draw the Boundaries?

When algorithmic systems are used to support decision-making, their primary function is often to *discriminate*. Yet, legally speaking, some forms of distinction are prohibited. This renders a societal consensus over the moral wrong of differentiating between certain social groups in certain contexts. EU secondary law only prohibits a finite number of such distinctions, namely when they are based on sex or gender, race or ethnic origin, disability, religion or belief, age and sexual orientation. Should the law extend to algorithmic distinctions that unfairly and systematically exclude given social groups from accessing valuable social goods? This raises deep-seated normative questions concerning the social function and the mandate of non-discrimination law. For example, should anti-discrimination law apply to algorithmic decision-making systems that rely on behavioural data (such as e.g. eating, sleeping, or sports habits) to exclude consumers from given insurance policies or to personalise the prices of such services in exclusionary ways? With the pervasive deployment of algorithmic systems, there is a risk that new patterns of systemic discrimination emerge based on aggregated social sorting performed by predictive analytics, algorithmic profiling and decision-making. This could bear grave socio-economic consequences for social groups that are not protected under EU equality law (Gerards & Borgesius, 2022; Wachter, 2022). In addition, some scholars have drawn attention to the fact that 'emergent' forms of algorithmic distinctions might not always correspond to socially salient features (Mann & Matzner, 2019; Leese, 2014). While it is clear that such algorithmic distinctions are morally unfair, is it the role of anti-discrimination law to address them? And if so, how? In responding to these questions, there is scope to consider how the open-ended list of protected grounds provided in Article 21 of the Charter, similar to Article 14 of the European Convention of Human Rights, could be used to address algorithmic

discrimination beyond the categories protected by EU secondary law. This situation would nevertheless be limited to the subsidiary space where member states are implementing EU law but in situations falling outside the scope of the equality directives (Kilpatrick, 2014).

Another friction that arises when attempting to apply EU anti-discrimination law to algorithmic discrimination relates to intersectionality. Algorithmic profiling powered by big data affects the granularity of the classifications underpinning decision-making. In other words, algorithmic distinctions are very likely to compound numerous data points, potentially at the intersection of several protected groups. This could give rise to so-called intersectional forms of discrimination, i.e. discrimination originating in several inextricably linked vectors of disadvantage. The problem is that the Court of Justice of the European Union has not recognised intersectional discrimination as a prohibited form of discrimination so far. In *Parris*, it stated that 'there is […] no new category of discrimination resulting from the combination of more than one [protected] groun[d] […] that may be found to exist where discrimination on the basis of those grounds taken in isolation has not been established'.[18] Thus discrimination induced by the use of algorithmic systems conceptually challenges the unidimensional or 'single axis' understanding of discrimination prevalent in EU law. Since intersectional discrimination is already pervasive in society but not legally recognised as such, it also risks being amplified through feedback loops while at the same time still remaining invisible. Indeed, the lack of participation and representation opportunities for intersectionally marginalised groups in society leads to increased algorithmic invisibility for these groups. The Gender Shades study has shown, for example, that face recognition systems display high rates of nonor misrecognition of the faces of women of colour (Buolamwini & Gebru, 2018). To effectively capture algorithmic discrimination, EU law should evolve towards a more complex conceptualisation of discriminatory harms, for instance by extending its scope to intersectional and systemic discrimination.

4. LEGALLY QUALIFYING ALGORITHMIC HARMS: REVISITING THE DICHOTOMY BETWEEN DIRECT AND INDIRECT DISCRIMINATION

The third type of difficulty encountered when applying EU anti-discrimination law to algorithmic harms is that of shoehorning algorithmic discrimination into the direct/ indirect bifurcated framework. The definitions of direct and indirect discrimination provided in the EU equality directives highlight three main legal criteria that in principle serve to distinguish between direct and indirect discrimination: the (absence of) *neutrality* of a given measure or practice, the existence of a discriminatory '*treatment*' vs. discriminatory '*effects*', and the presence of *group* vs. *individual* harm.[19] Applying such distinguishing criteria to algorithmic harms to determine whether they qualify as direct or indirect discrimination raises difficult normative questions. Qualifying algorithmic unfairness as direct or indirect discrimination is crucial because it leads to, respectively, a closed or an open-ended regime of justifications. This regime determines whether and how users of algorithmic technologies can lawfully use a system that is biased against a protected group. Hence, rather than simply responding to a legal technicality, the choice of qualifying algorithmic harms as direct or indirect discrimination directly shapes liability for algorithmic discrimination and thus amounts to deciding how the burdens of inequality are allocated among users of algorithmic technologies, potential victims and society at large.

First, how to qualify the neutrality of algorithmic operations? In other terms, what is a neutral criterion or practice in the context of algorithmic decision-making systems? Or else, in the absence of bias mitigation practices, can data-driven systems ever be conceptualised as neutral towards protected grounds? This is a thorny question because untreated data is very likely to reflect past discrimination, which algorithmic systems are then likely to treat as relevant factors for future predictions. One might even go so far as to call algorithmic discrimination a self-fulfilling prophecy. Hence, these systems can hardly qualify as neutral. At the same time, as argued elsewhere, algorithmic discrimination mostly takes the form of proxy discrimination (Prince & Schwarcz, 2019). Machine learning algorithms are trained to recognise patterns in large datasets. As a result, even if developers remove labels like sex, race or age, these systems still identify related patterns through correlated variables and can therefore unlawfully discriminate. For instance, an algorithm that used the distance between workers' home and the workplace as a predictor for job tenure was found to be discriminatory by proxy because it inferred workers' membership of an ethnic group based on zip code data (Williams et al., 2018). In the same vein, we could imagine that sports data collected through the use of apps influence the price paid by end users for loans or insurance. Or the content watched on media platforms such as YouTube or Netflix might reveal one's cultural affiliation and perhaps correlate with one's ethnic background, age and even socio-economic background. In *Dekker*, the ECJ recognised that when such proxies are inextricably linked to protected characteristics (e.g. pregnancy and sex), they cannot be considered neutral and such proxy discrimination qualifies as direct discrimination.[20] However, the Court's approach to what constitutes a proxy 'inextricably linked' to a protected ground is not entirely clear. This was illustrated in *Jyske Finans*, where the Court found that an applicant's country of birth did not suffice 'in itself, [to] justify a general presumption that that person is a member of a given ethnic group'.[21] It is thus difficult to predict whether the Court will treat algorithmic proxy discrimination as direct or indirect on this basis. Moreover, the notion of neutrality is easily manipulated depending on how the comparator group – or in Westen's terms the desirable level of equality in society (Westen, 1982) – is defined, as demonstrated in cases like *Achbita*, *WABE*, and *VL*.[22] Deciding whether an algorithmic system can qualify as neutral is key because it impacts the finding of direct or indirect discrimination and thus corresponding justification routes, with consequences on users' liability. Yet, data-driven decision-making calls for re-assessing the contours of the neutrality criterion.

Second, as the Advocate General recalled in *VL*, '[t]here is "indirect" discrimination where the difference resides not so much in the treatment as in the effects which it produces'. Yet qualifying algorithmic operations in terms of treatment or effect raises questions about how to qualify the entangled forms of agency that exist in human-machine relationships and sociotechnical systems. Algorithmic discrimination is technologically mediated by machines that can learn to discriminate. In this context and in light of automation biases, how to conceive of the agency of the humans in the loop? The notion of direct discrimination would capture algorithmic discrimination as a form of differential treatment consisting of a human decision interpreting an algorithmic recommendation.[23] In this legal construct, machine support would not detract from the integrity of human agency. Human agents would not be able to justify algorithmic discrimination invoking their lack of intent or awareness that the algorithmic recommendation was biased. Conversely, the notion of indirect discrimination would construct algorithmic discrimination as the effect of multiple conjugated causes among which a (set of) human practice(s). While strategically assigning

responsibility to the human agent, it casts a looser net around liability by permitting escape routes via justifications.

The same kind of dilemma arises when considering the question of causation. In direct discrimination, the notion of treatment 'on grounds of' a protected category and its interpretation by the ECJ have often been said to amount to a causation requirement. In other words, the difference in treatment arises 'because of' a protected ground. How to qualify causation when machine learning systems operate on the basis of correlations? At first sight, this would speak for conceptualising algorithmic discrimination as indirect discrimination. Yet, the causation requirement has often been relaxed by the ECJ, for instance in its case law on discrimination by association or direct proxy discrimination.[24] Again, it appears that both the direct and the indirect discrimination framework can be constructed to fit algorithmic discrimination, but choosing one over the other pertains more to a strategic rather than to a technical interpretation of the legal framework. *In fine*, this choice pertains to how legal rules distribute the costs of inequality among applicants, defendants and society.

Third, even though the ECJ has departed from this distinction, in principle, direct discrimination is related to individual harm and indirect discrimination to group harm. The problem is that algorithmic systems upset this distinction. They do not treat subjects *qua* individuals but rather based on algorithmic representations inferred from aggregated data and clustering. Hence these systems compound individual and collective harm by letting structural discrimination feed into individualised assessments. Even if a given algorithmic classification is not accurate for a given individual, its inclusion in a given algorithmic cluster will determine the treatment applicable to that person. Conceptualising algorithmic discrimination as individual or group harm is once again a strategic, rather than a technical, choice.

If, as has been argued in the literature so far (e.g. Hacker, 2018; Borgesius, 2020; Kelly-Lyth, 2021),[25] courts go down the indirect discrimination route to qualify algorithmic discrimination, this will create important challenges. Indirect discrimination raises essential questions such as 'how much' disadvantage amounts to a 'particular disadvantage' prohibited under EU anti-discrimination law and what ground truth to use as a baseline for comparison. It also entails an open pool of justifications and a proportionality test that is difficult to 'translate' in the context of algorithmic systems. Attempting to assess whether algorithmic discrimination can be justified leads to assessing decisions pertaining to technical trade-offs between accuracy and fairness.

5. CONCLUSIONS: SOME REFLECTIONS ON ALGORITHMIC EQUALITY

This chapter has shown how algorithmic technologies disrupt the application of existing legal constructs and thereby destabilise the justice arrangements entrenched in the law. Our task as legal scholars is to reflect on the normative implications of different possible interpretations of the legal framework in light of problems such as algorithmic discrimination. Such a reflection demands articulating the normative equilibria underpinning existing legal constructs and revisiting the law 'strategically' to safeguard or restore (and in some cases perhaps even alter) these value frameworks. In particular, national and EU non-discrimination laws need to be applied in a teleological and instrumental manner to guarantee that technological evolutions do not jeopardise fundamental rights. Anti-

discrimination law is technology-neutral and its application should not stop at the frontiers of the digital world. Possibilities for legal resilience include harnessing the principles of effectiveness and purposive interpretation. At the same time, reflecting on the strategic nature of legal interpretation and the deep-seated normative questions it raises invites us to look ahead. Thinking about how to redress algorithmic discrimination is an invaluable opportunity to think about the transformative potential of anti-discrimination law. Specifically, fixing algorithmic bias (imagining that it is even possible) will not solve the problem of algorithmic discrimination. Instead, preventing discriminatory algorithms from becoming a self-fulfilling prophecy requires transforming the *status quo*. Legally speaking, this invites us to think about carving a greater role for positive action measures in EU anti-discrimination law and to reflect on how to best tap into their transformative potential in the algorithmic society.

Endnotes

[1] Assistant Professor in European Law, Sciences Po, Law School. This research is linked to a Marie Skłodowska-Curie Fellowship project conducted at iCourts at the University of Copenhagen and the University of Edinburgh, School of Law. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 898937.

[2] Tribunal of Bologna, Order no. 2949/2019, 31 December 2020. Retrieved from tinoadapt.it/wp-content/uploads/2021/01/Ordinanza-Bologna.pdf

[3] By contrast, care duties did not influence men's score negatively.

[4] For example, accuracy is a legal requirement for high-risk systems in the draft EU AI Act. See inter alia Recitals 43 and 49, Art. 13(3)(b)(ii), Art. 15(1)(2) in 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, (2021).

[5] European Commission. (7 November 2022). The gender pay gap situation in the European Union. Retrieved from https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental
-rights/gender-equality/equal-pay/gender-pay-gap-situation-eu_en

[6] The wording changed slightly in the compromise version of November 2022: Permanent Representatives Committee (25 November 2022). *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: General approach*, Brussels, 14954/22. Retrieved from https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf

[7] By contrast to previous versions that did not mention discrimination in the binding part of the text, the compromise text adopted in November 2022 now explicitly acknowledges that bias can lead to 'discrimination prohibited by Union law' in Art. 10(2)(f) and recognises the right of national public authorities or bodies in charge of supervising or enforcing the respect of *inter alia* the right to nondiscrimination in relation to the use of high-risk AI systems to request or access information in Art.
64(3). See also Art. 10(5), Art. 14(4)(b) and Art 15(3) (European Commission, 2021a). Interestingly, the term 'fairness' is entirely absent from the binding part of the proposal.

[8] Amendment 78 of the European Parliament aims to mitigate this uncertainty in the context of the processing of special categories of personal data to ensure the detection and correction of 'negative bias' in relation to high-risk AI systems by adding to Recital 44 that '[n]egative bias should be understood as bias that create[s] direct or indirect discriminatory effect against a natural person'. See European Parliament. (14 June 2023). Amendments on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) A9-0188/2023 available at: https://www.europarl.europa.eu/doceo/document/TA-9
-2023-0236_EN.html

[9] European Commission. *Impact Assessment accompanying the 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*, Brussels, SWD(2021) 84 final (p. 91).

[10] See Directive 2000/43/EC (2000). *Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin*, Brussels, OJ L 180/22; Directive 2000/78/EC (2000). *Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation*, Brussels, OJ L 303/16; Directive 2004/113/EC (2004). *Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services*, Brussels, OJ L 373/37; Directive 2006/54/EC (2006). *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)*, Brussels, OJ L 204/23. In EU primary law, and in particular Art. 21 of the EU Charter of Fundamental Rights, the personal scope of protection is broader and non-exhaustive, but he CJEU has ruled that it could not be relied on to extend the scope of EU secondary law (C-354/13). See Case C-354/13, *Judgment of 18 December 2014: Fag og Arbejde (FOA) v Kommunernes Landsforening (KL)*. Court of Justice of the European Union. EU:C:2014:2463.

[11] As explained below, the protection is not even for all protected grounds across these areas.

[12] Note that the qualification of direct and indirect discrimination is different from US law. In EU law, direct discrimination does not require showing intent.

[13] See Directive 2000/78/EC. *Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation*, Brussels, OJ L 303/16; Directive 2004/113/EC. *Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services*, Brussels, OJ L 373/37; Directive 2006/54/EC. *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)*, Brussels, OJ L 204/23; Directive 2000/43/ EC. *Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin*, Brussels, OJ L 180/22.

[14] As highlighted in previous work, this 'translation' to the algorithmic context comes with conceptual, doctrinal and procedural challenges (Xenidis & Senden, 2019; Xenidis, 2021a; Gerards & Xenidis, 2021).

[15] Except in relation to the content of media or advertising and to education, see below.

[16] Of course, individual member states can opt for a more protective legal framework as long as it does not breach the EU treaties. A proposal for evening out the material scope of EU anti-discrimination law across protected grounds is pending since 2008. See European Commission. *Proposal for a Council Directive on implementing the principle of equal treatment between persons irrespective of religion or belief, disability, age or sexual orientation*, Brussels, COM(2008) 426 final.. In addition, Art. 21 of the EU Charter of Fundamental Rights might apply when member states are implementing EU law in matters falling outside the scope of the equality directives.

[17] See Recital 13, Directive 2004/113/EC. *Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services*, Brussels, OJ L 373/37.

[18] See Case C-443/15. *Judgment of 24 November 2016: David L. Parris v. Trinity College Dublin and Others*. Court of Justice of the European Union. EU:C:2016:897.

[19] Direct discrimination is defined in EU law as a situation in which '*one person* is *treated* less favourably than another [… is …] *on grounds of* [a protected characteristic]'. Indirect discrimination refers to situations 'where an *apparently neutral* provision, criterion or practice would put *persons* [who are members of a protected group] at a *particular disadvantage* compared with *other persons*' unless it can be objectively justified. (emphasis added).

[20] Case C-177/88. *Judgment of 8 November 1990: Elisabeth Johanna Pacifica Dekker v Stichting Vormingscentrum voor Jong Volwassenen (VJV-Centrum) Plus*. Court of Justice of the European Union. EU:C:1990:383.

[21] See Case C-668/15. *Judgment of 6 April 2017: Jyske Finans A/S v Ligebehandlingsnævnet, acting on behalf*

[22] See Case C-157/15. *Judgment of 14 March 2017: Samira Achbita and Centrum voor gelijkheid van kansen en voor racismebestrijding v G4S Secure Solutions NV*; Cases C-804/18 and C-341/19 (joined). *Judgment of 15 July 2021: IX v WABE eV and MH Müller Handels GmbH v MJ. Court of Justice of the European Union.* EU:C:2021:594; Case C-16/19. *Judgment of 26 January 2021: VL v Szpital Kliniczny im. dra J. Babińskiego Samodzielny Publiczny Zakład Opieki Zdrowotnej w Krakowie.* Court of Justice of the European Union. EU:C:2021:64. Court of Justice of the European Union. EU:C:2017:203. For instance, it has been argued that comparing religious and non-religious employees, as opposed to employees whose religious beliefs mandate the wearing of religious garment and employees whose (absence of) religious beliefs do(es) not, yields a different understanding of the neutrality of a practice (Cloots, 2018; Sharpston, 2021).

[23] For example, the outputs of the AMS algorithm were framed as 'second opinions' and the system itself as a 'mere support system' while decisions were 'delegated to case workers' (Alhutter et al., 2020).

[24] See Case C-83/14. *Judgment of 16 July 2015, 'CHEZ Razpredelenie Bulgaria' AD v Komisia za zashtita ot diskriminatsia*. Court of Justice of the European Union. EU:C:2015:480.

[25] More recently, it has been argued that algorithmic discrimination can amount to direct discrimination (Adams-Prassl, Binns & Kelly-Lyth, 2023).

References

Adams-Prassl, J., Binns, R. & Kelly-Lyth, A. (2023). Directly discriminatory algorithms. *The Modern Law Review*, *86*(1), 144–175. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-2230.12759.

Allhutter, D., Cech, F., Fischer, F., Grill, G. & Mager, A. (2020). Algorithmic profiling of job seekers in Austria: How austerity politics are made effective. *Frontiers in Big Data*, *3*, 1–17. Retrieved from https://doi.org/10.3389/fdata.2020.00005.

Borgesius, F.J.Z. (2020). Strengthening legal protection against discrimination by algorithms and artificial intelligence. *The International Journal of Human Rights*, *24*(10), 1572–1593. Retrieved from https://doi.org/10.1080/13642987.2020.1743976.

Buolamwini, J. & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, Proceedings of Machine Learning Research*, *81*, 77–91. Retrieved from https://proceedings.mlr.press/v81/buolamwini18a.html.

Cloots, E. (2018). Safe harbour or open sea for corporate headscarf bans? Achbita and Bougnaoui. *Common Market Law Review*, *55*(2), 589–624.

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. Retrieved from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-airecruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

Friedman, B. & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems (TOIS)*, *14*(3), 330–347.

Gerards, J. & Xenidis, R. (2021). *Algorithmic Discrimination in Europe: Challenges and Opportunities for EU Gender Equality and Non-Discrimination Law*. Brussels: Publications Office of the European Union. Retrieved from https://op.europa.eu/en/publication-detail/-/publication/082f1dbc-821d-11eb-9ac9-01aa75ed71a1.

Gerards, J. & Borgesius, F.Z. (2022). Protected grounds and the system of non-discrimination law in the context of algorithmic decision-making and artificial intelligence. *Colorado Technology Law Journal*, *20*, 1. Retrieved from https://ctlj.colorado.edu/?p=860.

Green, B. & Chen, Y. (2019). Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments. *ACM FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency*, 90–99. Retrieved from https://doi.org/10.1145/3287560.3287563.

Hacker, P. (2018). Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law. *Common Market Law Review*, *55*(4), 1143–1185. Retrieved from https://doi.org/10.54648/cola2018095.

Hellman, D. (2021). Big data and compounding injustice. *Virginia Public Law and Legal Theory Research Paper, No. 2021-27*. Retrieved from https://ssrn.com/abstract=3840175.

Kayser-Bril, N. (2019). Austria's employment agency rolls out discriminatory algorithm, sees no problem. *Algorithm Watch*. Retrieved from https://algorithmwatch.org/en/austrias-employment
-agency-ams-rolls-out-discriminatory-algorithm/.

Kelly-Lyth, A. (2021). Challenging biased hiring algorithms. *Oxford Journal of Legal Studies*, *41*(4), 899–928. Retrieved from https://doi.org/10.1093/ojls/gqab006.

Kilpatrick, C. (2014). Article 21 – non-discrimination. In S. Peers, T. Hervey, J. Kenner & A. Ward (Eds.). *The EU Charter of Fundamental Rights: A Commentary* (pp. 579–604). London: Hart Publishing.

Leese, M. (2014). The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union. *Security Dialogue*, *45*(5), 494–511. Retrieved from https://doi.org
/10.1177/0967010614544204.

Mann, M. & Matzner, T. (2019). Challenging algorithmic profiling: The limits of data protection and anti-discrimination in responding to emergent discrimination. *Big Data Society*, *6*(2), 1–11. Retrieved from https://doi.org/10.1177/20539517198958.

Mayson, S.G. (2018). Bias In, Bias Out. *University of Georgia School of Law Legal Studies Research Paper No. 2018-35*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3257004.

Prince, A.E.R. & Schwarcz, D. (2019). Proxy Discrimination in the age of artificial intelligence and big data. *Iowa Law Review*, *105*, 1257–1318. Retrieved from https://ilr.law.uiowa.edu/print/volume-105
-issue-3/proxy-discrimination-in-the-age-of-artificial-intelligence-and-big-data.

Sharpston, E. (2021). *Shadow Opinion in Joined Cases C-804/18 and C-341/19 IX v WABE e.V and MH Müller Handels GmbH v MJ*. Retrieved from http://eulawanalysis.blogspot.com/2021/03/shadow
-opinion-of-former-advocate.html.

Turk, V. (2015). When algorithms are sexist. *Vice*. Retrieved from http://www.vice.com/en_us/article/
ezvkee/when-algorithms-are-sexist.

Wachter, S. (2022). The theory of artificial immutability: Protecting algorithmic groups under antidiscrimination law. *Tulane Law Review*, *97*(2), 149–204. Retrieved from
https://www.tulanelawreview
.org/pub/artificial-immutability.

Weerts, H., Xenidis, R., Tarissan, F., Olsen, H.P. & Pechenizkiy, M. (2023). Algorithmic unfairness through the lens of EU non-discrimination law: Or why the law is not a decision tree. *2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Retrieved from https://arxiv
.org/abs/2305.13938.

Westen, P. (1982). The empty idea of equality. *Harvard Law Review*, *95*(3), 537–596. Retrieved from https://doi.org/10.2307/1340593.

Williams, B.A., Brooks, C.F. & Shmargad, Y. (2018). How algorithms discriminate based on data they lack: Challenges, solutions, and policy implications. *Journal of Information Policy*, *8*, 78–115. Retrieved from https://doi.org/10.5325/jinfopoli.8.2018.0078.

Wulf, J. (2022). *Automated Decision-Making Systems and Discrimination: Understanding Causes, Recognizing Cases, Supporting Those Affected*. Berlin: Algorithm Watch. Retrieved from https://
algorithmwatch.org/en/wp-content/uploads/2022/06/AutoCheck-Guidebook_ADM_Discrimination
_EN-AlgorithmWatch_June_2022.pdf.

Xenidis, R. & Senden, L. (2019). EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination. In U. Bernitz, X. Groussot, J. Paju & S.A. de Vries (Eds.). *General Principles of EU Law and the EU Digital Order*. Alphen aan den Rijn: Wolters Kluwer.

Xenidis, R. (2021a). Tuning EU equality law to algorithmic discrimination: Three pathways to resilience. *Maastricht Journal of European and Comparative Law*, *27*(6), 736–758. Retrieved from https://journals.sagepub.com/doi/full/10.1177/1023263X20982173.

Xenidis, R. (2021b). The polysemy of anti-discrimination law: The interpretation architecture of the Framework Employment Directive at the Court of Justice. *Common Market Law Review*, *58*(6), 1649–1696. Retrieved from https://doi.org/10.54648/cola2021108.