Recommended citation: Raphaële Xenidis and Linda Senden, 'EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination' in Ulf Bernitz et al (eds), *General Principles of EU law and the EU Digital Order* (Kluwer Law International, 2020)

EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination

Raphaële Xenidis[1] & Linda Senden[2]

The fast-growing use of machine-learning and artificial intelligence techniques opens up new horizons in terms of accuracy, efficiency and rapidity in everyday life applications. Artificial intelligence (AI) applications in fact play an increasingly important role in organising society at large, in the regulation of a wide variety of social systems and infrastructures and even in shaping human interactions and preferences.[3] Examples range from 'Internet of Things' applications to commercial nudging, from speech performance to face recognition software, from automated crime prediction to credit scoring systems, and from AI-powered cars to diagnosing tools in the healthcare sector, to cite just a few. Algorithmic applications therefore influence how opportunities and possibilities open up for, and are accessed by, individuals and entire groups of populations, or by contrast, how they close up. While AI can potentially provide society at large with a broader and more equal access to a wide range of goods and services, voices have emerged that also point at unprecedented risks of discrimination.[4] Many concrete examples relayed by the media show how grave, pervasive and multi-facetted the problem of algorithmic discrimination can be: face recognition applications perform much worse at recognising black women's faces than at white men's[5]; predictive policing algorithms prove to be racist[6]; and Tay.AI, a chatbot launched by Microsoft in 2016 had to be turned down after 24 hours because it had turned into a racist and sexist online hate speech machine[7]. Automated operations empowered by algorithms risk perpetuating, and thereby solidifying, a

[1] Raphaële Xenidis, European University Institute, Law Department & Utrecht University, International and European Law Department, raphaele.xenidis@eui.eu.
The authors would like to thank the organisers of the conference 'The Resilience of General Principles of EU Law and Fundamental Rights in the digital EU Legal order' held in November 2018 as well as Dr. Ad Feelders and Dr. Daniel Oberski from Utrecht University for their kind advice.

[2] Linda Senden, Professor of European Law at the Department of Law of Utrecht University, Co-Director of its RENFORCE research institute.

[3] See e.g. Iyad Rahwan and others, 'Machine behaviour' (2019) 568 Nature .

[4] See e.g. Virginia Eubanks, *Automating inequality : how high-tech tools profile, police, and punish the poor* (First edition., St. Martin's Press 2018); Safiya Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (NYU Press 2018); Cathy O'Neil, *Weapons of math destruction : how big data increases inequality and threatens democracy* (Crown 2016); Frank Pasquale, *The black box society : the secret algorithms that control money and information* (Harvard University Press 2015).

[5] See e.g. Joy Buolamwini and Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification* (Proceedings of Machine Learning Research 2018).

[6] See e.g. Julia Angwin and others, 'Machine Bias' (*ProPublica,* 23 May 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> accessed 22 September 2019.

[7] See e.g. Morgane Tual, 'A peine lancée, une intelligence artificielle de Microsoft dérape sur Twitter' (24 March 2016) *Le Monde,* available at https://www.lemonde.fr/pixels/article/2016/03/24/a-peine-lancee-une-intelligence-artificielle-de-microsoft-derape-sur-twitter_4889661_4408996.html accessed on 23 February 2019.

status quo that is deeply discriminatory. The consequences on citizens' daily life are tangible, endangering the exercise of their fundamental rights as well as reinforcing social hierarchies and material inequalities.

The way these digital technologies operate, the data they use and the way they process it as well as the output they produce thus need to be scrutinised and tested against the values and principles that organise the social contract and the legal rules that formalise it. Whole strands of literature on algorithmic fairness have rapidly developed and the question of 'algorithmic bias' has now become central to research in this area. As this book shows, lawyers too have started to grapple with the digital reality and its consequences for the rule of law, the established legal order and general principles of law. This chapter explores the question of the robustness of the legal protection framework in place in the EU in front of the risks of discrimination that arise from using these technologies. The aim is to map the main challenges that arise at the intersection of non-discrimination law and technological developments, with a particular focus on one area of artificial intelligence, namely machine learning algorithms.

While most studies on the topic of artificial intelligence, algorithms and bias have been conducted from the point of view of 'fairness' in the field of information technologies and computer science[8], this chapter explores the question of algorithmic discrimination — a category that does not neatly overlap with algorithmic bias — from the specific perspective of non-discrimination law. In particular and by contrast to the majority of current research on the question of algorithms and discrimination, which focuses on the US context, this chapter takes EU non-discrimination law as its object of inquiry. We pose the question of the resilience of the general principle of non-discrimination, that is the capacity for EU equality law to respond effectively to the specific challenges posed by algorithmic discrimination. Because EU law represents an overarching framework and sets minimum safeguards for the protection against discrimination at national level in EU member states, it is important to test out the protection against the risks posed by the pervasive and increasing use of artificial intelligence techniques in everyday life applications which this framework allows for. This chapter therefore focuses on the question on how technological developments in the field of artificial intelligence impact on the right not to be discriminated against, which is both a general principle and a fundamental right in EU law.

The central purpose of this chapter is thus to map the challenges which these digital developments entail in the light of the EU principle of non-discrimination. In so doing, we hope to lay the foundations of a new research agenda at the intersection of EU non-discrimination law and technology. In particular, we propose a mapping of the frictions, mismatches and difficulties arising at the contact points between the growing use of AI, and notably machine

[8] See e.g. Solon Barocas and others, *Fairness and machine learning: Limitations and Opportunities* (2019) https://fairmlbook.org accessed 10 September 2019; Sam Corbett-Davies and others, 'Algorithmic decision making and the cost of fairness' (2017) Part F129685 Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 797; Dwork Cynthia and others, *Fairness through awareness* (2012); Matt Kusner and others, 'Counterfactual Fairness' (31st Conference on Neural Information Processing Systems); Richard Zemel and others, *Learning Fair Representations* (Journal of Machine Learning Research 2013).

learning algorithms, in decision-making procedures and the protective framework offered by EU non-discrimination law. The picture we paint is not monochrome and we contend that AI also offers pathways towards less discriminatory decision-making and opens up opportunities to increase awareness of, and tackle, the current discriminatory status quo. Likewise, we also contend that EU non-discrimination law offers potential pathways to deal with algorithmic discrimination which are so far little used and worthwhile exploring further. We proceed in four steps. First, we identify the specific risks of discrimination that AI-based decision-making, and in particular machine-learning algorithms, pose. Second, we review how EU non-discrimination law can capture algorithmic discrimination in terms of its substantive scope. Third, we conduct this review from a conceptual perspective, mapping the friction points that emerge from the viewpoint of the EU concepts of direct and indirect discrimination, as developed by the Court of Justice of the EU (CJEU). In the final step, we identify the core challenges algorithmic discrimination poses at the enforcement level, and propose potential ways forward.

I The mechanics of algorithmic discrimination

I. 1. Artificial intelligence, machine-learning and algorithms

Even though artificial intelligence is a very broad field, it can be roughly defined as encompassing a set of technologies seeking to “make computers do the sorts of things that [human] minds can do”.[9] To do so, machines rely on sets of rules called algorithms. Algorithms are mathematical instructions that aim to solve problems, answer questions or perform specific tasks. They can be very simple, for instance specifying that if a condition A is met an outcome B should follow, or very complex, combining multiple sets of rules towards building a whole system. One complex class of algorithms perform what is called machine-learning, that is the “ability [for machines] to acquire their own knowledge, by extracting patterns from raw data”.[10] While automation can pose challenges in themselves by leading to the generalisation of discriminatory rules on a larger scale than would be the case with human beings, the operation of machine-learning algorithms is, however, of much greater concern.

Because machine-learning has the ability to learn and discover correlations in any given dataset, the risk of discrimination is greater. Even when not instructed to discriminate, machine-learning algorithms have been found producing discriminatory outcomes. Besides just automation of discrimination, machine-learning techniques pose the additional risk of reproducing existing patterns of inequality in ways unintended by their designers. By way of illustration, a machine-learning algorithm designed to decide about workplace promotions and relying on past successful applicants’ data could be made blind to applicants’ gender and still be discriminatory against female applicants, for instance if asked to select candidates based on an assessment of

[9] Margaret A. Boden, *Artificial intelligence: a very short introduction* (Oxford University Press 2018), 1. See also Ian Goodfellow, Yoshua Bengio and Aaron Courville, *Deep Learning* (MIT Press 2016), 9.
[10] Goodfellow, Bengio and Courville, *Deep Learning*, 2-3.

their performance at work that would include their average working hours.[11] Women would for instance be disadvantaged because they disproportionately carry the burden of caregiving according to typical gender roles in society, thus working less hours or more part-time.[12] The challenge with machine learning is that algorithms are not only at risk of conveying human biases in their operation – they are also a vehicle for reproducing inequality patterns that are structurally engrained in the data they process. In view of these challenges, machine-learning algorithms are the point of focus of this chapter. While a great variety of such algorithms exist as well as different learning procedures, the scope of this chapter only allows us to adopt a general focus.[13]

I. 2. What risks of discrimination do machine-learning algorithms pose?

Algorithmic decision-making has been described as the new "regulatory frontier" *inter alia* for non-discrimination law and equality protection.[14] Discrimination can infect machine-learning algorithms at several entry points. In order to operate, an algorithm needs two elements: a code that formalises a problem in mathematical terms; and data, that is a set of input variables that the machine can learn from. Several typologies have been proposed in the literature to describe how either the code or the data can be biased, thus injecting discrimination into algorithmic operations. Kleinberg et al. identify risks of discrimination in the choice of outcome that is entrusted to the algorithm, in the selection of input information used to train it (what they call "candidate predictors") and in the training procedure used, including the training data.[15]

Barocas and Selbst also propose a five-tier taxonomy of the sources of discrimination. Their first entry point corresponds to outcome definition, or what the authors call "problem definition" through the construction of "target variables", themselves broken down into "class labels".[16] The second entry point is the training data used for the learning process, which can be biased either because of the way it has been collected or because the data itself is biased and is then relied on to teach the algorithm.[17] A third entry point is the selection of relevant features

[11] See Jeffrey Dastin, 'Amazon scraps secret AI recruiting tool that showed bias against women'(<https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-airecruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G> accessed 23 February 2019. See also Pauline T. Kim, 'Data-Driven Discrimination at Work' (2017) 48 William & Mary Law Review 857 and Miriam Kullmann, 'Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law' (2018) 34 The international journal of comparative labour law and industrial relations 1.

[12] See e.g. Carole Pateman, *The sexual Contract* (Stanford University Press 1988) and Pierre Bourdieu, *La Domination Masculine* (Seuil 1998). See also Susanne Burri and Helga Aune, 'Sex Discrimination in Relation to Part-Time and Fixed-Term Work' (European network of legal experts in gender equality and non-discrimination & European Commission 2013) < https://www.equalitylaw.eu/downloads/2804-sex-discrimination-en> accessed 21 September 2019.

[13] See e.g. Boden, *Artificial intelligence: a very short introduction*, 39-43 for relevant distinctions between *inter alia* supervised, unsupervised, reinforcement and deep learning.

[14] See Hacker Philip and Petkova Bilyana, 'Reining in the Big Promise of Big Data: Transparency, Inequality, and New Regulatory Frontiers' Northwestern Journal of Technology and Intellectual Property.

[15] Jon Kleinberg and others, 'Discrimination in the Age of Algorithms' (2019) 25548 NBER Working Paper Series, 21-23.

[16] Solon Barocas and Andrew D. Selbst, 'Big Data's Disparate Impact' (2016) 104 California law review 671, 677-680.

[17] Ibid, 681-687.

for the model, that is, what attributes should be considered relevant in the algorithmic model. A fourth entry point is "proxy discrimination", where discrimination arises from so-called "redundant encodings", i.e. "cases in which membership in a protected class happens to be encoded in other data" which is considered correlated to the outcome by the algorithm.[18] A typical example would be the use of a non-protected category (e.g. residence or zip code) as a proxy for a protected category (e.g. race). A fifth and last entry point they identify is the masking of intentional discrimination, which programmers can hide in the architecture of the algorithm.

Hacker reduces these sources of discrimination to two main entry points. "Biased training data" covers situations where the training data is either incorrectly handled, for instance when the correct answer taught to the machine is biased, or skewed because of historical discrimination. [19] "Unequal ground truth" corresponds to situations where the "best available approximation of reality" (or empirically observable data) is not equal between groups — for instance, if risks of car crash are higher among men than among women. Since measuring risks is highly complex and costly for insurance policy providers, the groups themselves (men and women) can be used as proxies to compensate the lack of more granular information.[20] This explains why gender has been used as a proxy for a long time before the landmark *Test-Achats* decision of the Court of Justice of the European Union (CJEU) in 2011.[21] While this is now prohibited in the EU, there is a risk that machine learning uses other indirect proxies (such as tastes and behaviours) that are correlated with protected proxies to estimate risks. This would be the case for example if tastes for certain types of sports or cars that are held prevailingly by men are used to estimate given risks in the insurance sector.

I. 3. Bias through the lens of non-discrimination law: demystifying algorithmic discrimination

Despite the risks highlighted above, the discriminatory potential of machine-learning algorithms should not be mystified.[22] First, discriminating (in the broad sense) is part of continuous operation of algorithms. From a legal point of view, it is not a problem in itself: the problem only arises when algorithms discriminate based on legally protected categories. In the context of EU law, these protected categories are sex, race or ethnic origin, as well as disability, age, sexual orientation and religion or belief.[23] Nationality is also covered in particular cases, including in relation to the freedom of movement of persons, goods, services and capital.[24] Second, and perhaps most importantly from a non-discrimination law perspective, algorithms reinforce and propagate patterns of inequality that already exist in the social fabric. A well-

[18] Ibid, 691.

[19] Philipp Hacker, 'Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law' (2018) Common market law review , 1146-1148.

[20] Ibid, 1449.

[21] C-236/09 *Association Belge des Consommateurs Test-Achats and Others* ECLI:EU:C:2011:100.

[22] We intend to tackle question of positive action by algorithms in subsequent research.

[23] These grounds are covered by directives 2006/54/EC and 2004/113/EC for sex, Directive 2000/43/EC for race or ethnic origin and Directive 2000/78/EC for the grounds of disability, religion or belief, age and sexual orientation.

[24] Article 18 TFEU.

known adage among computer scientists – "garbage in, garbage out" – conveys the idea that any discriminatory algorithmic output comes from bias injected into algorithms by human beings.[25] In other words, and as reframed by Mayson, "bias in, bias out".[26]

From the point of view of discrimination theory, the sources of algorithmic discrimination can be categorised into two overarching types of inequality (re-)producing mechanisms.[27] On the one hand, stereotyping and prejudice affect the equal representation of groups in society. On the other, past discrimination that has been institutionalised and reified over the course of history is reflected in structural forms of inequality. Algorithmic discrimination has the potential to reinforce both distributive inequality by maintaining discriminatory access to social goods (e.g. labour, health services, social benefits, exercise of rights, etc.) and symbolic inequality by misrepresenting or rendering invisible certain groups of population (e.g. search engines returning sexualised images of black women).[28] Algorithmic discrimination therefore both arises from, and further entrenches, hierarchising status beliefs and stereotypes as well as structural institutionalised patterns of inequality. From the perspective of discrimination law, the mechanisms through which discrimination might invade algorithms can thus be classified along these two main axes.[29]

I.3.1 Stereotyping in framing, labelling and modelling

Difference and social divisions are associated with (negative or positive) stereotypes in human cognitive schemes, which contribute to solidifying and maintaining social hierarchies over time.[30] There are several ways in which bias can pervade algorithmic design during the developing process. It is possible, for instance, that a software reflects the biases and stereotypes that are held by its developers.[31] This can be done intentionally, but mostly prejudice will be implicit. Stereotyping can influence the framing of the problem posed, and of the output looked for. For instance, if an algorithm is developed to find out who is the best candidate, the output will likely vary according to the definition given to 'best'. If the definition of 'best' ascribes great weight to so-called leadership skills, the output is likely to discriminate against female candidates, leadership often being stereotypically conceived of as a male characteristic[32] and stressing abilities and qualities that are usually typified as 'male', such as being ambitious,

[25] See Sandra G. Mayson, 'Bias In, Bias Out' (2018) 128 Yale Law Journal .

[26] Ibid.

[27] See Charles Tilly, *Durable Inequality* (University of California Press 1998); Cecilia Ridgeway, 'Why Status Matters for Inequality' (2014) 79 American Sociological Review 1; Nancy Fraser, *Social Justice in the Age of Identity Politics: Redistribution, Recognition, Participation* (Wissenschaftszentrum Berlin für Sozialforschung, Berlin: 1998).

[28] See e.g. Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (St. Martin's Press 2018) and Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism.*

[29] Barocas and Selbst, 'Big Data's Disparate Impact'. 'Masking' will be discussed in section 1. 3.

[30] See Iyiola Solanke, 'Stigma. A limiting principle allowing multiple consciousness in anti-discrimination law?' in Dagmar Schiek and Victoria Chege (eds), *European Union Non-Discrimination Law Comparative Perspectives on Multidimensional Equality Law* (Routledge-Cavendish 2009)

[31] See e.g. Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism.*

[32] See e.g. Janine Bosak and Sabine Sczesny, 'Am I the Right Candidate? Self-Ascribed Fit of Women and Men to a Leadership Position' (2008) 58 Sex Roles 682.

forceful, self-sufficient and self-confident.[33] Stereotypes and prejudices can also impact the labelling process. For instance, if white men wearing a white overall are likely to be categorised as doctors, women wearing the same dress will more likely be identified as nurses.[34] Externalising data labelling to internet users through microwork platforms like Mechanical Turk or captcha-based services like reCaptcha might lead to prejudiced outcomes on yet a bigger scale.

Stereotyping also happens where the lack of information about the features of a certain group creates uncertainty. Some proxies might yield a high level of predictive accuracy but at the same time be discriminatory. This is because, despite its general statistical relevance, the use of this feature is unfair at the individual level. In particular, in the absence of perfect information or more granular data and in front of the cost of obtaining such data, stereotypes and generalisations regarding certain groups of population might be relied on as a way to approximate reality. As mentioned above, until the *Test-Achats* case, insurers for instance used gender as a proxy (an actuarial factor) to calculate risks and thus estimate premiums in the absence of granular information about individuals.[35] Group stereotyping reduces uncertainty but might also create "unsound generalizations [that] may deny members of these populations the opportunity to prove that they buck the apparent trend" or that the stereotype is unfounded.[36] Stereotyping is problematic because it can have performative effects, that is reinforce and enhance discrimination and inequality.[37] Research has for instance shown that gender stereotypes and unequal gender representation in online image searches influence users' perceptions of the real world and thus their representation of certain professions as typically male or female.[38]

I.3.2 Structural discrimination engrained in data

Structural discrimination, which is the product of past discrimination institutionalised over time and now reflected in many ways in the organisation of society, is mirrored in data. In particular, structural discrimination engrained in training data will prove problematic. First, if this data accurately represents reality, it could still reflect structural inequalities which machine learning algorithms would then reproduce. The danger is that discrimination manifested in statistical correlations becomes further reified by algorithms, entrenching exclusion and inequality. For instance, if an algorithm is trained using a dataset gathering all employees hired in the past in

[33] See Alice Eagly and Steven Karau, 'Role Congruity Theory of Prejudice Toward Female Leaders', Psychological Review 2006, vol. 109, No. 3, 573-598, at 574.

[34] See Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism.* On the consequences of such stereotypes, see Matthew Kay and others, 'Unequal Representation and Gender Stereotypes in Image Search Results for Occupations' (2015) Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.

[35] C-236/09 *Association Belge des Consommateurs Test-Achats ASBL and Others v Conseil des ministres* [2011] EU:C:2011:100.

[36] Barocas and Selbst, 'Big Data's Disparate Impact' , 687.

[37] On gender stereotyping as discrimination, see e.g. Rebecca J. Cook and Simone Cusack, *Gender Stereotyping. Transnational Legal Perspectives* (University of Pennsylvania Press 2010) and A. S. H. Timmer, *Gender Stereotyping in the Case Law of the EU Court of Justice* (2016). See also Frederick F. Schauer, *Profiles, probabilities, and stereotypes* (Belknap Press of Harvard University Press 2003).

[38] Matthew Kay, Cynthia Matuszek and Sean A. Munson, 'Unequal Representation and Gender Stereotypes in Image Search Results for Occupations', (Association for Computing Machinery).

order to predict which employees should be hired in the future based on their potential performance, it might reproduce discrimination against women and overwhelmingly correlate the male gender to expected job performance.[39] This is because past hiring decisions are infected with discriminatory gender stereotypes that will be reproduced by the algorithm. In a more diffuse way, language itself is a "naturally skewed data".[40] Because language reflects our representation of the world, which is deeply hierarchical and biased, it is also a vector of structurally engrained discrimination, the patterns of which will be reproduced by machine-learning algorithms if not addressed.[41] Research has for instance shown how natural language processing systems "exhibit gender bias mirroring stereotypical gender associations" such as "Man is to computer programmer as Woman is to homemaker", thereby "pick[ing[up on historical biases encoded in their training corpus".[42] This shows how gender inequality, and in particular stereotypical gender roles (female homemaker vs. male breadwinner) are deeply embedded in various types of data.

It could be, moreover, that data collection and selection itself is skewed and misrepresents reality. For instance, collecting data from all those who have an internet connection might exclude poor or rural populations, who will then not be represented at all in the sample used to train algorithms. In the same vein, the project 'The Coded Gaze', hosted at MIT, reveals how training data in which black women are under-represented leads to face recognition software being less performant on recognizing black women's faces while performing much better on black and white men's faces as well as white women's faces.[43] This example shows how machine-learning algorithms can put certain groups of population at a disadvantage through reproducing structural discrimination. The 'Gender Shades' study reveals how some of the datasets chosen to train face recognition devices are biased, as is the case of a dataset of celebrity faces containing 77,5% male faces and 83,5% white faces, leading to underperformance in relation to the underrepresented categories of population.[44] If a face recognition software was trained on randomly selected datasets containing the faces of public political figures, e.g. members of parliaments, the selection would also be biased as it would reflect the structurally unequal access to politics for men and women, and in particular women from racialised minorities, in many countries.[45] This lack of representation of certain groups of populations is problematic because it leads the algorithm to 'unsee' them, thus making them invisible. The learning being based on past discrimination, inequality is reproduced by

[39] See for instance Amazon's hiring algorithm: Dastin, 'Amazon scraps secret AI recruiting tool that showed bias against women'.

[40] See Aylin Caliskan, Joanna J. Bryson and Arvind Narayanan, 'Semantics derived automatically from language corpora contain human-like biases' (2017) 356 Science.

[41] Kaiji Lu and others, *Gender Bias in Neural Natural Language Processing* (2018).

[42] See Lu and others, *Gender Bias in Neural Natural Language Processing* 2-3 and T. Bolukbasi and others, 'Man is to computer programmer as woman is to homemaker? Debiasing word embeddings' (2016) Advances in Neural Information Processing Systems 4356.

[43] See The Coded Gaze: https://www.ajlunited.org/the-coded-gaze. See also Buolamwini and Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*.

[44] See , 3.

[45] This is why the researchers of the above-mentioned study carefully selected datasets from various countries to reflect a wide spectrum of skin tones, and chose Parliaments were the gender parity was highest in order to avoid reproducing structural discrimination through algorithmic bias. See Buolamwini and Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 5-6.

machines, in this case ignoring further the existence of black women in the public sphere by denying them visibility. Such erasure could lead to extremely grave discriminatory consequences if built in other technologies such as healthcare services or automated cars.

Finally, because a model always approximates reality, it uses proxies that are generalisable: "Indeed, the very point […] is to provide a rational basis upon which to distinguish between individuals and to reliably confer to the individual the qualities possessed by those who seem statistically similar".[46] Some of these proxies might be statistically accurate predictors but at the same time embed structural discrimination and thus reproduce inequality. On the one hand, banning protected categories from the data available to machine learning algorithms does not prevent discriminatory outcomes, as machines use non-protected data as proxies for these categories.[47] On the other hand, refining granularity so as to find a feature that performs as good or better but at the same time does not lead to unfair generalisations about a protected group might be costly. For example, proxies such as the reputation of the schools from which candidates graduated, despite being unfair to candidates from minorities and not adequately reflecting a person's actual job-related skills, are used because they allow to reduce uncertainty in front of otherwise difficultly measurable items (e.g. work performance potential) by substituting them with easily accessible and low cost information.[48] Choosing such proxies might however lead to so-called "redundant encoding" or "feedback effect", that is the reification and essentialisation of false correlations inherited from past discrimination into the present.[49] In other words, inequalities that have been socially constructed over time through accumulated past discrimination become considered 'natural properties' of certain groups of population. The effect is performative: taking inequality into account as a relevant group characteristic in present decisions about disadvantaged groups further contributes to systematically depriving these groups from equal opportunities. In the example above, the under-representation of minority candidates in elite graduate schools is institutionalised as a valid selection rule by the algorithm and reproduced, thus further entrenching inequality.

The consequences of machine-learning algorithms in relation to bias and data-driven inequality in terms of automation, reproduction and even performativity of discrimination are therefore grave and it is essential to evaluate whether non-discrimination laws in place are fit to prevent and redress them. This is our agenda for the next sections.

II The applicability of the current EU non-discrimination law regime to cases of algorithmic discrimination

This section explores the new challenges machine-learning algorithms pose in the specific context of the protection EU law offers against discrimination. It probes the robustness of the

[46] Barocas and Selbst, 'Big Data's Disparate Impact', 677.

[47] See Betsy Anne Williams, Catherine F. Brooks and Yotam Shmargad, 'How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications' (2018) 8 Journal of Information Policy .

[48] See Barocas and Selbst, 'Big Data's Disparate Impact'.

[49] Kim, 'Data-Driven Discrimination at Work'.

legal framework against situations of algorithmic discrimination that are likely to arise in the context of current machine learning applications.

The general principle of non-discrimination in EU law finds expression in a set of primary and secondary legal provisions protecting people from discrimination based on their sex, race or ethnic origin, disabilities, religion or belief, age and sexual orientation. Nationality is protected in certain areas of EU law too.[50] The personal scope of these provisions covers workers (Directives 2006/54/EC for the ground of sex – hereinafter the Gender Recast Directive, Directive 2000/43/EC for the ground of race or ethnic origin – hereinafter the Race Equality Directive and Directive 2000/78/EC for the grounds of disabilities, religion or belief, age and sexual orientation – hereinafter the Employment Equality Directive). In the case of sex (Directive 2004/113/EC – hereinafter the Gender Goods and Services Directives) and race or ethnic origin certain cases (Race Equality Directive 2000/43/EC) consumers and service providers are covered too. The material scope of the EU general principle of non-discrimination therefore spans the labour market, where the protection is at its widest, and the consumption market, where it is more limited.[51]

II. 1. Algorithmic discrimination in the labour market

The first and most extensive field in which EU non-discrimination legislation has been deployed is employment and labour relations.

II.1.1 Equal pay

Historically, the principle of equal pay, now enshrined in Art. 157 TFEU and Article 4 of Directive 2006/54/EC, foresees that women and men should be paid equally when performing equal work and work of equal value. While the implementation of this principle is still challenging in non-digital sectors, this is even truer in the digital economy, especially for on-demand services.[52] Empirical studies on platform work show that the hourly income of women averages to thirds of men's hourly wage.[53] In the case of driving services, for instance, drivers' pay depends on their evaluation by an algorithm that computes their score using features such as their availability rate and clients' ratings. The use of these features channels both stereotypes and structural inequalities into the scoring process. Citing Uber as an example, Küllman for instance exposes how stereotypes about the non-reliability of women could influence clients' ratings and feed into the overall evaluation of drivers and thus into the calculation of their pay, disadvantaging female workers.[54] By analogy, if drivers were considered to be employees, this situation would potentially amount to discrimination in light of the CJEU's jurisprudence in

[50] See Art. 18 TFEU.

[51] This also covers social protection in the case of gender and race.

[52] November 6th is the date that symbolises the average gender pay gap in the EU each year, by marking the day where women stop being paid until December 31st each year.

[53] See Arianne Renan Barzilay and Anat Ben-David, 'Platform Inequality: Gender in the Gig-Economy' (2017) 47 Seton Hall law review 393.

[54] See Kullmann, 'Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law'.

Bougnaoui.[55] In this case, the Court explained that clients' preferences regarding dealing with an employee who did not wear a hijab could not be considered a genuine and occupational requirement and thus could not justify discrimination on the basis of religion. Following the same reasoning, prejudiced preferences expressed by clients in ratings should not lawfully inform workers' pay and influence their working conditions in discriminatory ways.

Another feature that might influence the algorithm that calculates drivers' pay is their availability rate, which might disadvantage women who generally carry most of the burden of care duties.[56] Here again, the gendered organisation of society might weight negatively in algorithmic scoring, infringing on the principle of equal pay and thus indirectly discriminating against women. By analogy, the CJEU case law on part-time work, notably in *Bilka-Kaufhaus*, has made clear that excluding part-time workers from certain segments of pay or benefits could amount to indirect sex discrimination if "a far greater number of women than men" are impacted by the exclusion.[57] This casts doubts as regards the lawfulness of algorithmic scoring to calculate pay when it affects certain protected categories such as gender in a disproportionately disadvantageous way.

In both cases, gender is not a feature inputted in the algorithmic decision-making procedure, but still works to the structural disadvantage of female workers. EU non-discrimination law covers these situations and thus seems fit to address such types of algorithmic discrimination. However, the application of the equal pay principle in this regard highly depends on whether the relationship between gig economy platforms and companies on the one hand and workers on the other can be called an employment relationship from the point of view of EU law.[58] The scope of the principle of equal pay in fact seems to exclude service providers who are genuine self-employed workers.[59]

II.1.2 Employment: hiring, promotion and working conditions

Another configuration of the general principle of non-discrimination is its translation into a prohibition of discrimination on grounds of gender (Gender Recast Directive 2006/54/EC), race and ethnic origin (Race Equality Directive 2000/43/EC), disability, sexual orientation, religion or belief and age (Employment Equality Directive 2000/78/EC) in recruitment procedures, promotions, working conditions and professional training. Hiring and promotion procedures perhaps illustrate best the risk algorithmic decision-making poses with regard to the perpetuation of inequalities. Some digital human resources services have specialized in offering

[55] C-188/15 *Asma Bougnaoui and Association de défense des droits de l'homme (ADDH) v Micropole SA* EU:C:2017:204.

[56] See Kullmann, 'Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law'.

[57] C-170/84 *Bilka - Kaufhaus GmbH v Karin Weber von Hartz* EU:C:1986:204.

[58] This is a question which national courts have been recently confronted with: see the UK decision *Uber B.V. v Aslam and others* [2018] EWCA Civ 2748 and the French decision *CA Paris*, 6-2, 10 January 2019 confirming that Uber drivers are workers. However, 'bogus' or economically dependent self-employed workers, despite being classified as self-workers under national law, might be regarded as workers under EU law as ruled by the CJEU in *Allonby*, thus extending the application of the equal pay principle to platform workers. See C-256/01 *Debra Allonby v Accrington & Rossendale College, Education Lecturing Services, trading as Protocol Professional and Secretary of State for Education and Employment* EU:C:2004:18.

[59] Kullmann, 'Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law', 16-17.

companies scoring methods evaluating which employees perform best and should thus progress towards higher-ranked positions. Amazon for instance had to renounce to an algorithmic model that systematically discriminated against female and racialized employees.[60] The 'smart' algorithm that would screen and rank employees had been trained on datasets that featured the profiles of employees promoted in the past. Although gender and race were not part of the set of variables inputted in the model, the machine-learned algorithm considered that these features correlated with candidates' performance. Inevitably, the training data used in machine learning reflects past discrimination and is the product of human bias in relation to protected categories in past decisions about which employees to promote.

In the same vein, machine-learning algorithms are found to discriminate against members of protected categories when used in the context of hiring procedures. Of course, it is unlawful to input protected criteria as relevant variables in the model. This could be captured as direct discrimination under EU non-discrimination law. As mentioned earlier, even when algorithms are blinded to sensitive information about membership in protected categories, existing correlations in training data might still produce discriminatory outcomes in unintended ways.[61] Kim for instance explains the case of a machine learning algorithm that used the distance between workers' home and workplace "as a predictor of employee job tenure", which first appeared benign, but in fact proved discriminatory "because […] housing patterns are correlated with race".[62] Identifying a particular geographic area as structurally disadvantaged, with lower levels of education and economic activity on average, the algorithm gave a disadvantageous score to its residents, who in fact mostly belonged to a racialized group.

Following Article 2(1)(b) of Directive 2006/54/EC and Article 2(2)(b) of Directive 2000/43/EC and 2000/78/EC, both cases are likely to amount to indirect discrimination under EU law because they put bearers of protected characteristics at a disproportionate disadvantage even though this was not intended. In addition, the concept of discrimination by association which the CJEU developed in *Coleman* and *CHEZ* would strengthen the level of protection afforded by EU law.[63] This concept could in fact cover cases of algorithmic misclassification: because the level of granularity of an algorithm might not extend to the individual level, some individuals who are not members of protected groups could be associated with these groups and misclassified. Following the CJEU's jurisprudence, they would be afforded the same level of legal protection as members of protected groups themselves. This scenario would be similar to *CHEZ* (2014) in the area of goods and services, where the resident of a 'racialised' area mostly populated by Roma people was excluded from accessing certain electricity services based on the racist stereotypes which the company held against the Roma population of this area.[64] The company, explaining that it feared electricity theft in the area, installed electricity meters so high that clients could not reach them to control their electricity consumption. The

[60] See Dastin, 'Amazon scraps secret AI recruiting tool that showed bias against women'.

[61] See Williams, Brooks and Shmargad, 'How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications'.

[62] Kim, 'Data-Driven Discrimination at Work', 873.

[63] C-303/06 *S. Coleman v Attridge Law and Steve Law* EU:C:2008:415 and C-83/14 *CHEZ Razpredelenie Bulgaria AD contre Komisia za zashtita ot diskriminatsia* EU:C:2015:480.

[64] C-83/14 *"CHEZ Razpredelenie Bulgaria" AD v Komisia za zashtita ot diskriminatsia* EU:C:2015:480.

applicant in *CHEZ* was not of Roma origin herself but the CJEU held that she was the victim of the racist stereotypes and ascriptions that targeted her neighbourhood. More generally, the analysis the CJEU performed in *CHEZ* could also be a source of inspiration to tackle proxy-based algorithmic discrimination. This should be taken in combination with *Coleman*, an employment case in which the mother of a disabled child claimed to have been harassed and otherwise discriminated against as her employer refused her time off work and flexible working arrangements to care for her child. In *Coleman*, the CJEU recognised that is "the prohibition of direct discrimination […] is not limited only to people who are themselves disabled" but extend to people who are associated with disabled persons, e.g. their caregivers. Therefore, with regard to such employment-related cases of algorithmic discrimination the current EU law regime seems to provide some useful yardsticks.

II.1.3 Algorithmic stereotyping in job ads and opportunities

Algorithms are used to advertise job openings, in particular through online platforms for ad delivery such as Facebook. With the collection of more and more personal data and the personalisation of online ads based on collected data, the issue of discrimination prompted by algorithmic profiling in advertisement becomes pressing. Recent studies have shown that substantial risks exist that targeted advertising of job positions end up discriminating against users as well as reinforcing stereotypical labour divisions in society. Most telling are the results ensuing from Ali et al.'s experiment, revealing that employment ads set to target the same audience and delivered by Facebook ended up reaching an 85% female audience for cashier positions in supermarkets, while taxi driver positions reached a 75% black audience and lumberjack positions an audience that was male at 90% and white at 72%.[65] Algorithmic profiling, by multiplying the possibilities for targeted advertising, therefore pose the question of discrimination and the protection EU law can offer.

Because job advertising is part of the recruitment process, it falls within the scope of employment and is thus covered by the Gender Recast Directive, the Employment Equality Directive and the Race Equality Directive.[66] These directives in fact apply to "conditions for access to employment […] including selection criteria and recruitment conditions".[67] The protection EU non-discrimination law affords on grounds of gender, race, sexual orientation, age, religion or belief and disability should thus be understood to cover cases in which job-related ads are delivered in discriminatory ways, perpetuating stereotypes and reinforcing existing patterns of structural discrimination. Algorithmic discrimination in job-related advertising could be compared to the situations which the CJEU dealt with in *ACCEPT* and *Feryn*.[68] In these cases, the Court deemed discriminatory statements made in public by employers, respectively on grounds of sexual orientation and ethnic origin, to fall within the

[65] Muhammad Ali and others, 'Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes' (2019) available at https://arxivorg/abs/190402095 .

[66] See Paul Post and Rikki Holtmaat, 'A False Start: Discrimination in Job Advertisements' (2014) 2 European Gender Equality Law Review 12, 13.

[67] Art. 14(1)(a) of Directive 2006/54/EC; Art. 3(1)(a) of Directive 200/43/EC and Directive 2000/78/EC.

[68] See C-54/07 *Centrum voor gelijkheid van kansen en voor racismebestrijding v Firma Feryn NV* EU:C:2008:397 and C-81/12 *Asociaţia Accept v Consiliul Naţional pentru Combaterea Discriminării* EU:C:2013:275.

scope of the prohibition of direct discrimination even in the absence of identified victims.[69] In *ACCEPT*, an important shareholder of a major football team in Romania had declared in public that he would not want any gay football player in his team. The CJEU decided that these statements could amount to discrimination because "it may be inferred [from them] that [the club] has a discriminatory recruitment policy".[70] In *Feryn*, the director of a major company selling and installing doors had publicly stated that he would not employ immigrants because the company's clients would be reluctant to give them access to their houses for installation purposes. The Court found that the public statement at stake could amount to discrimination because they had the effect of deterring potential job applicants from ethnic minority background from applying.[71] It is notable that these cases extend the protection against discrimination to situations of dissuasion that exclude certain groups from given job opportunities even prior to the selection process, even without identifiable complainants.[72] Since recital 8 of the Race Equality and the Employment Equality Directives states that their aim is "to foster conditions for a socially inclusive labour market" and "to foster a labour market favourable to social integration", it could be argued by analogy that stereotyping based on protected grounds in the advertising of jobs, which de facto results in excluding certain groups from certain job opportunities, could be considered discriminatory.[73] In fact, such algorithmic stereotyping not only influences online users' representations in terms of career opportunities through reinforcing prejudice through the phenomenon of "echo chambers"; it also affects the real-world distribution of job opportunities, thereby maintaining segregation and inequality in the labour market.[74]

Furthermore, the use of algorithms in the distribution of labour also risk directly reproducing and reifying biased preferences and harmful stereotypes by systematically offering less work opportunities to protected minorities on the basis of users' ratings and evaluations. For instance, a study by Hannák et al. shows that "perceived gender and race are significantly correlated with worker evaluations" on online platforms, "which could harm the employment opportunities afforded to the workers".[75] Algorithms in fact use "rating and review data to power recommendation and search systems", therefore allowing customers' prejudices to affect the automated distribution of work.[76] Under the guise of maximising efficiency, machine learning

[69] See ibid.

[70] *ACCEPT* [49].

[71] *Feryn* [25].

[72] In this sense, see *ACCEPT* [45] and [52] : "It is irrelevant in that regard that […] the system of recruitment […] is not based on a public tender or direct negotiation following a selection procedure requiring the submission of applications" and "the fact that a professional football club might not have started any negotiations with a view to recruiting a player presented as being homosexual does not preclude the possibility of establishing facts from which it may be inferred that that club has been guilty of discrimination". See also *Feryn* [23].

[73] This reasoning also applies to the Gender Recast Directive, of which recital 11 states that "marked gender segregation on the labour market" should be addressed.

[74] See Kay, Matuszek and Munson, 'Unequal Representation and Gender Stereotypes in Image Search Results for Occupations'. See also e.g. Nicholas DiFonzo, 'The Echo Chamber Effect', *The New York Times* (22 April 2011) <https://www.nytimes.com/roomfordebate/2011/04/21/barack-obama-and-the-psychology-of-the-birther-myth/the-echo-chamber-effect> accessed 21 September 2019.

[75] Anikó Hannák and others, *Bias in Online Freelance Marketplaces: Evidence from TaskRabbit and Fiverr* (2017).

[76] Ibid, 1915.

algorithms might end up aggravating structural inequalities through “misrecognition” and “maldistribution” by relying on biased input data.[77] As stated in the previous section, the fact that EU non-discrimination law does not accept customers’ prejudices as a valid justification for discrimination could be used as a basis to challenge such types of discrimination, provided that these situations fall within the scope of employment.[78]

These are but some examples of how algorithmic discrimination may collide with and be captured by EU non-discrimination law. Similar problems could arise in various other employment-related fields which the non-discrimination directives cover, such as the access to social security, vocational training, redundancy schemes, etc. EU non-discrimination law offers quite some potential to address algorithmic discrimination occurring in the labour market, both when it comes to protected grounds and groups as well as protected fields (access to employment, employment conditions and pay). This is further supported by the horizontal direct effects recognized by the Court of Justice to the principle of non-discrimination in the jurisprudential series spanning *Mangold*, *Kücükdeveci*, *Dansk Industri* and *Egenberger*.[79] However, this protective potential would need to be confirmed by the Court of Justice, a task which will require some interpretive creativity to adapt existing regulations to digital forms of discrimination.

II. 2. The provision and supply of goods and services

In the case of the consumption and supply of goods and services, EU non-discrimination law offers a much more limited protection. In the absence of a so-called ‘Horizontal Directive’[80] covering all discrimination grounds, its personal scope is restricted to the grounds of sex and race or ethnic origin as laid down in Directive 2004/113/EC and 2000/43/EC. In addition, the material scope of this protection might be constrained by a number of exceptions in national transposing laws.[81] While the absence of horizontal protection is a problem per se in this area, it becomes exacerbated in light of the risks linked to algorithmic discrimination. Quite paradoxically, the area of goods and services is where algorithmic discrimination could potentially have its greatest and gravest impact. Discrimination in this field can take a number of forms.

II.2.1 Structural misrepresentation in the access to goods and services

[77] See Axel Honneth and Nancy Fraser, *Redistribution or Recognition? A Political-Philosophical Exchange* (Verso 2004).

[78] See *Bougnaoui* [2017].

[79] C-144/04 *Werner Mangold v Rüdiger Helm* EU:C:2005:709; C-555/07 *Seda Kücükdeveci v Swedex GmbH & Co. KG* EU:C:2010:21; C-441/14 *Dansk Industri (DI) contre Succession Karsten Eigil Rasmussen* EU:C:2016:278; C-414/16 *Vera Egenberger v Evangelisches Werk für Diakonie und Entwicklung e.V.* EU:C:2018:257. See further section IV and also Raphaële Xenidis, 'Transforming EU Equality Law? On Disruptive Narratives and False Dichotomies' (2019) Yearbook of European Law .

[80] 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 COM(2008) 426 final.

[81] See Eugenia Caracciolo di Torella and Bridgette McLellan, *Gender equality and the collaborative economy* (European Commission, 2018).

The problem of algorithmic discrimination already starts at the stage of data collection. Data collected to train algorithms might exclude some portions of the population. This can result in lesser performance of ensuing algorithms in relation to the groups excluded. Unavailability or underperformance of goods and services, based on these algorithms, fitting the needs of women, racialised, religious and sexual minorities, the elderly or people living with disabilities can prove extremely discriminatory. For instance while AI applications are more and more used in the healthcare sector, the issue of under-representation of women and racialised minorities in data used to train algorithms becomes important. Health services that do not take into consideration the specific needs of women and minorities, who might suffer from different issues or exhibit different symptoms compared to the white male population which research primarily focuses on, might discriminate in their availability or performance.[82] Many more fields than healthcare services are concerned by this issue, e.g. housing, driving, etc.[83] Such discrimination could fall under the scope of Directive 2004/113/EC for gender and 2000/43/EC for race or ethnic origin. In fact, recital 17 of the Gender Goods and Services Directive states that "[t]he principle of equal treatment in the access to goods and services does not require that facilities should always be provided to men and women on a shared basis, as long as they are not provided more favourably to members of one sex".

II.2.2 Algorithmic stereotyping in advertising

Algorithmic discrimination is also likely to take place in advertising services, for instance on platforms that use algorithms to match providers and buyers, in particular in ad targeting and delivery on the consumption market. Optimisation and personalisation of ads and ad delivery rely on the gathering of personal data by online platforms and third parties and the profiling of users based on their personal characteristics. As Ali et al. show, "significant skew in delivery [take place] along gender and racial lines for 'real' ads for [*inter alia* credit and] housing opportunities despite neutral targeting parameters".[84] Gendered and racial skews are for example observed in relation to housing ads on Facebook.[85] This might be because advertisers themselves target the ads at certain audiences, or because ad delivery services match ads and audiences based on their profile, interests and the success rates of past ads. The second scenario is particularly worrisome because it entails that "some users [are] less likely than others to see particular ads based on their demographic characteristics" even when "advertisers set their targeting parameters to be highly inclusive".[86] For instance, it has been shown that ads which target the same audience might be distributed to highly differentiated users based on ad delivery services' stereotypical assumptions about their content and images, for instance cultural stereotypes leading to racial profiling or gender stereotypes leading to sex-based targeting.[87]

[82] See e.g. Oras A Alabas and others, 'Sex Differences in Treatments, Relative Survival, and Excess Mortality Following Acute Myocardial Infarction: National Cohort Study Using the SWEDEHEART Registry' (2017) 6 Journal of the American Heart Association .

[83] See e.g. Caroline Criado Perez, *Invisible Women: Data Bias in a World Designed for Men* (Abrams Press, New York 2019).

[84] Ali and others, 'Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes', 1.

[85] Ibid.

[86] Ibid.

[87] Ibid, 2.

Algorithmic stereotyping of images and content leading to gendered and racial differentiation in ad exposure might lead to discrimination through offering less opportunities to certain groups, but also through reinforcing harmful stereotyping. Beside optimisation in delivery, economic considerations alone will lead low budget ads to be shown to less valuable audiences (for instance men) while high budget ads will be shown to audiences that are more demanded because considered more attractive (for instance women), which might lead to discrimination when these audiences correlate with protected categories such as gender.[88]

The problem is that EU non-discrimination law covers advertising only in a restricted way, given the existence of a so-called hierarchy of grounds and the lack of agreement on the pending proposal for a Horizontal Directive evening the level of protection for all grounds.[89] Advertising is excluded from the scope of the Gender Goods and Services Directive, therewith allowing for gender-targeting ads.[90] The Race Equality directive however remains silent on this point, thus potentially outlawing race-based advertising for goods and services in the EU. Other grounds are not covered in the ambit of the consumption market. The protection EU law offers in this field is therefore insufficient and risks allowing algorithmic stereotyping reinforcing existing symbolic and structural inequalities. In addition, other obstacles arise linked to the application of EU non-discrimination law. In cases where commercial offers are highly personalised, establishing discrimination might be an issue. Hacker has for instance expressed doubts regarding whether such personalised products or services would fall under the definition of "goods and services which are available to the public" mentioned in Art. 3(1)(h) of the Race Equality Directive and Art. 3(1) of the Gender Goods and Services Directive.[91]

II.2.3 Algorithmic price discrimination

Another configuration of algorithmic discrimination on the goods and services markets translates in price discrimination. Big data and the availability of information based on individual characteristics such as "location, age, gender, employment status" or behavioural observations (e.g. consumer behavior, search histories, etc.) offer companies new possibilities to "mine consumers' digital footprints, using machine learning algorithms to enable digital retailers to predict the price that individual consumers ('final end users') are willing to pay for particular items, and thus offer them different prices".[92] Targeting can thus lead to differentiated prices based on protected categories such as gender and ethnic background. A major concern is

[88] Ali and others, 'Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes'.

[89] On the Horizontal Directive, see Proposal for a Council Directive on implementing the principle of equal treatment between persons irrespective of religion or belief, disability, age or sexual orientation COM(2008)0426 final. On the issue of a hierarchy of grounds, see e.g. Erica Howard, 'The Case for a Considered Hierarchy of Discrimination Grounds in EU Law' (2006) 13 Maastricht Journal of European and Comparative Law 445. The term refers to the unequal scope of protection afforded to grounds of discrimination, with the widest scope for race or ethnic origin (employment, goods and services, social protection and education), followed by gender (employment, social security, goods and services) and then age, sexual orientation, disability and religion or belief (employment).

[90] See Art 3(3) of Directive 2004/113/EC : "This Directive shall not apply to the content of media and advertising nor to education".

[91] Philipp Hacker, 'Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law' (2018) 55 Common Market Law Review.

[92] Christopher Townley, Eric Morrison and Karen Yeung, 'Big Data and Personalized Price Discrimination in EU Competition Law' (2017) 36 Yearbook of European Law 683, 684 and 688.

"that ACPD [algorithmic consumer price discrimination] may result in a form of social sorting, with some groups routinely discriminated against, particularly on the basis of gender, race, or geographic location".[93] Algorithmic price discrimination could thus enhance existing patterns of discrimination in pricing, for instance gender-based pricing, and entail higher costs or lacking supply for some protected categories.

EU non-discrimination law prohibits pricing on the basis of gender and race or ethnic origin through Directives 2004/113/EC and 2000/43/EC. This has been confirmed by the Court of Justice in the already mentioned *Test-Achats* case outlawing the use of gender as an actuarial factor in insurance policy pricing.[94] However, unclarities remain regarding the application of existing legislation to other types of goods and services. In addition, other protected grounds are not covered in this field, so that price discrimination based on sexual orientation, disability, religion or belief and age would not be unlawful under EU law. Moreover, other types of price discrimination, for instance based on socio-economic resources which is not a protected ground in the EU equality framework, might also occur, facilitated by machine-learning algorithms. Algorithms therefore more than ever pose the question of the boundaries between prohibited discrimination and discrimination not (yet) covered by the law, in particular on the consumption market. The legal framework seems to fall short in effectively tackling algorithmic discrimination on the consumption market because of its limitation both in terms of protected grounds and areas that are excluded from its scope.

This section has provided some examples of algorithmic discrimination, which illustrate how algorithms pose an increased risk of discrimination in some configurations. It has highlighted the risks linked to the hierarchy of equality protection offered by EU law, with race equality enjoying the broadest scope of protection, followed by gender equality, while the protection on other grounds of discrimination – disability, religion or belief, sexual orientation and age – does not apply to goods and services at all.[95]

III. Conceptual challenges, uncertainties and limits of EU non-discrimination law in the digital era

The examples above illustrate how the problem of algorithmic discrimination and its many forms raise questions regarding the scope and application of EU non-discrimination law. This section zooms in on the two main conceptual tools contained in EU non-discrimination law: direct discrimination and indirect discrimination. A preliminary challenge is that, depending on the context and on whether the focus is on the algorithmic operation itself or the inclusion of its output in the decision-making process of a human operator, discrimination can be conceived

[93] Ibid, 719.

[94] C-236/09 *Association Belge des Consommateurs Test-Achats ASBL and Others v Conseil des ministres* EU:C:2011:100.

[95] See e.g. Lisa Waddington and Mark Bell, 'More Equal than Others: Distinguishing European Union Equality Directives' (2001) 38 Common Market Law Review 587 and Erica Howard, 'The Case for a Considered Hierarchy of Grounds in EU Law' (2006) 13 Maastricht Journal of European and Comparative Law 445.

as either direct or indirect. The co-involvement of humans and machines in decision-making processes makes it difficult to classify discrimination strictly as differential treatment or disproportionate disadvantage given the complexity of decision-making chains. Questions arise as to which steps a discrimination analysis should focus on: the scoring process by an algorithm or as the case may be the incorporation of such scoring in final decisions by a human operator? The boundaries between the concepts of direct and indirect discrimination are therefore blurred in the context of algorithmic discrimination. Hence, as Kim puts forward, a "mechanical application of [the] existing doctrine will fail to address the real sources of bias when discrimination is data-driven".[96]

III.1 Direct discrimination: a sound conceptual basis with limited applicability?

Direct discrimination is defined in EU law as a situation in which "one person is treated less favourably than another is, has been or would be treated in a comparable situation".[97] In the context of algorithms, direct discrimination captures situations where models are not neutral in relation to a protected ground. If any element of an algorithmic rule or code is not neutral towards a protected ground, the result will fall under the concept of direct discrimination. This can be the case, for instance, when a protected ground is directly inputted in an algorithmic model as a relevant variable and treated as a negative factor.

One of the strengths of EU non-discrimination law in this context is the irrelevance of intent. This feature of EU non-discrimination law separates the debate from US legal analyses of discrimination, where the notions of 'motive' and 'intent' are central to a finding of so-called 'disparate treatment'.[98] Whether a protected ground was treated differently as a result of intention or not does not matter in EU law, which potentially allows the concept of direct discrimination to capture a broad range of situations where protected grounds would be used as relevant variables by an algorithmic model even though it was not the programmers' intention to discriminate.

However, since developers strive for accuracy, cases of direct discrimination will be rather rare because directly inputting discrimination in an algorithmic model is likely to reduce its predictive value, which constitutes an important disincentive.[99] In addition, awareness of legal obligations is generally well established with regard to the ban on directly treating protected groups differently. Despite the rareness, one risk which the concept of direct discrimination would cover, however, is the concealing of discrimination under the veneer of neutrality.[100] In sum, if the concept of direct discrimination is fit to capture certain situations of algorithmic discrimination, its relevance is likely to be less important than that of indirect discrimination.

[96] See Kim, 'Data-Driven Discrimination at Work', 866, 869, 908.

[97] E.g. Art. 2(2)(a) Directive 2000/43/EC.

[98] *McDonnell Douglas Corp. v. Green*, 411 U.S. 792 (1973).

[99] See Hacker, 'Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law', 1152.

[100] This is what Barocas and Selbst call "masking". See Barocas and Selbst, 'Big Data's Disparate Impact'.

Issues of proof of differential treatment of these groups are however likely to complicate victims' task of proving direct discrimination. Machine learning processes are often described as 'black boxes' in light of the difficulties in understanding whether the parameters of an algorithmic model are neutral towards protected categories.[101] In particular, the opacity of such models to laypersons, but also the proprietary nature of certain commercial algorithms and the ensuing lack of disclosure, make it difficult for lawyers and judges alike to understand what is in the box and whether it is discriminatory. In addition, the CJEU established in *Meister* that a right to recruitment information does not exist vis-a-vis an employer, even though the company's refusal to provide such information could count as an element towards the establishment of a *prima facie* case of discrimination.[102] By analogy, *Meister* does not place applicants in an easy situation when it comes to accessing information about a suspect algorithm. Finally, performing the comparison-based test inherent in non-discrimination law might be a challenge for judges in light of this lack of transparency.[103] To establish direct discrimination, judges in general select real or hypothetical comparators and examine whether the protected group at stake has been treated in a differential way by the algorithmic model. Establishing algorithmic direct discrimination might thus be complicated because of the above-mentioned hurdles.

III.2 Indirect discrimination: a broader reach but more justifications

Indirect discrimination is likely to capture many situations of algorithmic discrimination. It refers to situations 'where an apparently neutral provision, criterion or practice would put [members of a protected category] at a particular disadvantage compared with other persons, unless that provision, criterion or practice is objectively justified by a legitimate aim and the means of achieving that aim are appropriate and necessary'.[104] In the words of McCrudden and Prechal, "[i]ndirect discrimination prohibits practices that formally apply to all from having the effect of disadvantaging individuals of particular protected groups, unless those practices can be shown to be objectively justified by a legitimate aim and the means of achieving that aim are appropriate and necessary".[105] Two elements of the concept of indirect discrimination make it particularly relevant to algorithmic discrimination. First, on the surface the treatment of protected groups is neutral. This allows capturing a wide array of situations in which algorithms do not operate on the basis of protected groups directly, and even situations where algorithms were made explicitly blind to these groups so that they are not picked as relevant variables. Indirect discrimination seems fit to capture a large spectrum of apparently neutral but indeed discriminatory algorithmic outputs, for instance situations in which training data is biased towards certain groups (under- or over-inclusion), the phenomenon of 'redundant encoding' through which structural discrimination is reproduced by algorithmic models, as well as proxy

[101] See Pasquale, *The black box society : the secret algorithms that control money and information*.

[102] C-415/10 *Galina Meister v Speech Design Carrier Systems GmbH* EU:C:2012:217.

[103] Exceptions to the comparability requirement exist, e.g. in the context of pregnancy discrimination: C-177/88 *Elisabeth Johanna Pacifica Dekker v Stichting Vormingscentrum voor Jong Volwassenen (VJV-Centrum) Plus* EU:C:1990:383.

[104] E.g. Art. 2(2)(b) Directive 2000/43/EC.

[105] Christopher McCrudden and Sacha Prechal, The Concepts of Equality and Non-Discrimination in Europe: A practical approach (2009), 35.

discrimination in which variables which correlate with a protected ground are used as relevant features or labels in an algorithm.

Furthermore, the concept of indirect discrimination allows shifting the focus of the analysis onto the effects of algorithms, instead of their rules, parameters and content. By the same token, it shifts the analysis from the perpetrator of discrimination to the victims.[106] There is no need to know exactly how an algorithm operates, rather, what is relevant to the case is only its discriminatory impact. Not having to open the 'black box' is likely to be an advantage for victims of discrimination. In that perspective, the focus of the concept of indirect discrimination on the group affected by the discriminatory measure rather than on the individual applicant might also be a better conceptual fit with the way algorithms operate.

The concept of indirect discrimination could also provide a safety net in case no direct discrimination can be found in light of difficulties of proof regarding the non-neutrality of algorithmic rules. Establishing a *prima facie* case of indirect discrimination might, in some cases, be easier for applicants. In fact, testing and measuring algorithmic output could prove easier than measuring the effect of a neutral rule in a human setting, so that statistics (even if not required to establish indirect discrimination in EU law) could represent an easily available means of evidence. Current discussions focus on how algorithmic auditability, that is the possibility that algorithms are checked, reviewed and monitored by third parties, could facilitate such measures.[107] That said, access to this information in the context of proprietary algorithms could pose further problems.[108]

Once a *prima facie* case of indirect discrimination is established, the burden of proof shifts onto the defendant. This provision, applied to algorithmic discrimination, places the burden on the providers of algorithm-based services (e.g. employers, service providers, etc.) to demonstrate that their algorithm is not discriminatory. EU rules on the burden of proof strengthen applicants' position. However, it has been argued that the strength of the concept of indirect discrimination is compromised by the specific challenges posed by the possibility of objective justification in the context of algorithms.[109] Indeed, a third dimension of the concept of indirect discrimination is the possibility for it to be objectively justified within the ambit of a proportionality test.[110] First, the discriminatory output of the algorithm at stake must serve a legitimate aim. This is an easy step for defendants as the use of algorithmic models in itself will serve legitimate business purposes (e.g. ranking or scoring algorithms to find out which employees are most performant,

[106] See Alan David Freeman, 'Legitimizing Racial Discrimination Through Antidiscrimination Law: A Critical Review of Supreme Court Doctrine' (1978) 62 Minnesota Law Review 1049.

[107] See Rumman Chowdhury and Narendra Mulani, 'Auditing Algorithms for Bias' *Harvard Business Review* (<https://hbr.org/2018/10/auditing-algorithms-for-bias?referral=03758&cm_vc=rr_item_page.top_right>.

[108] See discussion in the next section (IV).

[109] See Hacker, 'Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law'.

[110] See C-170/84 *Bilka-Kaufhaus GmbH v Karin Weber von Hartz* EU:C:1986:204; C-96/80 *J.P. Jenkins v Kingsgate (Clothing Productions) Ltd.* EU:C:1981:80; C-127/92 *Dr. Pamela Mary Enderby v Frenchay Health Authority and Secretary of State for Health* EU:C:1993:859. To be justified, a measure or practice shall be "objectively justified by a legitimate aim and the means of achieving that aim [must be] appropriate and necessary".

estimating a default risk, etc.).[111] If the aim is found to be legitimate, the measure at stake also needs to be deemed appropriate and necessary, that is effective and proportionate. In cases of algorithms, it is likely that courts accept that they are effective means to the aim of making accurate predictions and the like.[112] Algorithms are in fact developed precisely to ensure a level of precision and granularity that human minds are not able to reproduce. Hence, the requirements of a legitimate aim and the appropriateness of an algorithm meeting that aim are likely to be satisfied.

However, following the last prong of the proportionality test conducted by the CJEU — the necessity requirement — there must be “no other means of achieving [the same] aim that imposes less of an interference with the right to non-discrimination”.[113] By contrast to other academic views, we contend that the Court of Justice is less likely to accept broad justifications for this last part of the proportionality test.[114] The recent *Achbita* case also points towards a rejection of blanket measures by the Court and an acceptation of more narrowly tailored policies or practices.[115] In this case, the Court examined whether employees’ prohibition to wear a religious sign in the form of a hijab was a blanket ban or limited to those workers who sustained interactions with the clients. [116] Only the second situation could fulfil the necessity requirements. By analogy, in case of algorithmic discrimination, the question is likely to become one about the trade-off between business efficiency and non-discrimination. This will call for a balancing act between accuracy and equality. Hence indirect discrimination seems conceptually fit to tackle a large spectrum of issues, but its application appears challenging, thus calling into question whether it will provide a systematic redress against algorithmically induced discrimination.

In addition to these frictions, situations of algorithmic discrimination could be conceived of as either direct or indirect discrimination depending on whether the algorithmic operation itself is considered or its application in decision-making (depending on whether the decision-making process involves a human supervisor). Distinguishing between the two concepts might not always be a clear-cut case given the complexity of the human-machine relationship and the fragmentation of algorithmic decision-making systems. Frontiers between direct and indirect discrimination might become blurred in certain configurations. Responding to the critique that

[111] See e.g. Hacker, 'Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law', 1161.

[112] See ibid.

[113] European Union Agency for Fundamental Rights and Council of Europe, *Handbook on European non-discrimination law* (European Union and Council of Europe, 2018), 45.

[114] See e.g. Hacker, 'Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law'.

[115] See C-157/15 *Samira Achbita and Centrum voor gelijkheid van kansen en voor racismebestrijding v G4S Secure Solutions NV* EU:C:2017:203, [42]: ‘As regards, in the third place, the question whether the prohibition at issue in the main proceedings was necessary, it must be determined whether the prohibition is limited to what is strictly necessary. In the present case, what must be ascertained is whether the prohibition on the visible wearing of any sign or clothing capable of being associated with a religious faith or a political or philosophical belief covers only G4S workers who interact with customers. If that is the case, the prohibition must be considered strictly necessary for the purpose of achieving the aim pursued.’

[116] Despite the contestable nature of the reasoning and outcome in Achbita, the CJEU shows a refusal to accept blanket bans as necessary.

"the rules governing indirect discrimination may be noticeably more flexible than those relating to direct discrimination", AG Sharpston indicated in *Bougnaoui* that "[i]t might be objected that the application of the rules laid down by EU law to the latter category is unnecessarily rigid and that some 'blending' of the two categories would be appropriate".[117] In addition to this, another issue arises in case algorithms are used to disguise direct discrimination. If used to conceal intentional discrimination, algorithms could engender so-called covert direct discrimination, but such a configuration would be even harder to prove and would probably be treated as indirect discrimination by the CJEU.[118] Finally, the enforcement of EU non-discrimination law might also prove challenging, a problem which the next section tackles in detail.

IV. Enforcement challenges: a need to rethink the application of EU non-discrimination law?

Ensuring compliance with and securing enforcement of the EU non-discrimination principle in relation to algorithmic discrimination might be problematic for various reasons. First of all, conceiving the fundamental right to equality as the other side of the coin of the non-discrimination principle, one can distinguish two main ways of enforcement: the individual rights-based approach, relying upon private individuals bringing claims to court, and the monitoring, supervisory approach, relying upon a variety of public institutions, tools and mechanisms to take the necessary action to implement and enforce equality. Both ways reveal hurdles and weaknesses, which cast doubt on whether current enforcement mechanisms of EU non-discrimination law are fit to tackle algorithmic discrimination. We identify the most pressing enforcement issues below and unpack the challenges they pose for the future.

IV.1 The individual rights-based approach

For a long time, the prevalent mechanism for enforcing rights deriving from EU law has been the initiation of national court proceedings. The Court of Justice's recognition of the supremacy of EU [then EEC] law and of its direct effect has been key to this individual approach to enforcement.[119] Citizens have played a very important role in tackling discrimination and in bringing about norms regarding the effective judicial protection of their right to equality. National courts have by now turned to the CJEU in hundreds of cases, with a duty to leave aside any national rule of law conflicting with EU law. A striking, early example of this process is the case of Mrs Defrenne, who made a significant contribution to the development and enforcement of the equal pay principle by bringing various cases to the national court in Belgium, leading in turn to three preliminary references to the CJEU.[120] We take the Defrenne

[117] Opinion of AG Sharpston, *Asma Bougnaoui and Association de défense des droits de l'homme (ADDH) v Micropole SA*, 13 July 2016 EU:C:2016:553, [65].

[118] On the concept of covert direct discrimination, see Oddny Mjöll Arnadóttir, *Equality and Non-Discrimination under the European Convention on Human Rights*, vol 74 (Kluwer Law International 2003).

[119] C-6/64 *Flaminio Costa v E.N.E.L.* EU:C:1964:66 and C-26/62 *NV Algemene Transport- en Expeditie Onderneming van Gend & Loos v Netherlands Inland Revenue Administration* EU:C:1963:1.

[120] Mrs Defrenne suffered in particular from the impact her earlier compulsory retirement age than for male colleagues had on her retirement pension. See C-80/70 *Gabrielle Defrenne v Belgian State* EU:C:1971:55; C-43/75 *Gabrielle Defrenne v Société anonyme belge de navigation aérienne Sabena* EU:C:1976:56 and C-149/77 *Gabrielle Defrenne v Société anonyme belge de navigation aérienne Sabena* EU:C:1978:130.

cases as our lead example to illustrate the problems that emerge with the individual rights based approach in relation to algorithmic discrimination.

As stated above, a first major issue concerns the actual identification of algorithmic discrimination. How to overcome the lack of transparency in the nature and effects of algorithms? How can victims and lawyers establish evidence of discrimination in light of algorithms' technological and technical complexity? While the CJEU indicated in *Meister* that a private employer is not required to disclose information regarding the recruitment process, it also established in a consistent line of case law that transparency is a precondition for the effective enforcement of the non-discrimination principle.[121] Tackling the problem of differential treatment in retirement age and pension calculations, the Court required in *Danfoss* that job classification systems be transparent and that employees be entitled to access information as to how pay structures are set up and tasks and functions evaluated.[122] Without transparency, female employees may in fact not even be aware that they are being discriminated against. As a consequence, even though the CJEU has sent contradictory signals, one could argue that basic transparency, openness and disclosure requirements need to apply to algorithmic models to enable identification of discriminatory practices and effects and to allow applicants to make *prima facie* cases of discrimination. This is all the more relevant in the context of current research on auditable algorithms, algorithmic accountability and explainable AI, all geared towards enabling a better understanding of how AI operates, enhancing the interpretability and increasing the transparency of AI-assisted decision-making.[123]

In case a *prima facie* case of discrimination can be established, a second major issue nevertheless arises, which concerns how to ensure access to justice. In the digital sphere, a variety of problems exist in this regard, which differ from more classic cases such as the *Defrenne* case. To Mrs Defrenne, it was clear that her employer, the airline Sabena, was to be brought to court because it was responsible for the infringement of the equal pay principle. In relation to algorithmic discrimination, however, the question of liability may be more delicate. Is the designer and developer of discriminatory algorithms or the provider who bases its services on such algorithms to be held liable?[124] A further question concerns the relevant jurisdiction: what law governs the conflict that arises between victims of discrimination and algorithmic service providers and what court should be turned to? Upon deciding these questions, a more fundamental interrogation might arise: is it a realistic and viable strategy for citizens to bring a case to court in light of the technical expertise required and the related costs incurred?

[121] *Meister* [2012].

[122] C-109/88 *Handels- og Kontorfunktionærernes Forbund I Danmark v Dansk Arbejdsgiverforening, acting on behalf of Danfoss* EU:C:1989:383.

[123] See e.g. Henrik Palmer Olsen, Jacob Livingston Slosser, Thomas Hildebrandt, Cornelius Wiesener, 'What's in the Box? The Legal Requirement of Explainability in Computationally Aided Decision-Making in Public Administration (2019) *iCourts Working Paper Series* No. 162 and Céline Castets-Renard, 'Régulation des algorithmes et gouvernance du machine learning : vers une transparence et "explicabilité" des décisions algorithmiques ?' (November 2018) Revue Droit&Affaires .

[124] See section III.2. on the complexity of the human-machine relationship.

A third important problem is that of the restricted invocability of EU equality law according to whether the defendant party is a public or private entity. In *Defrenne II*, the Court decided that Article 157 TFEU (ex-Article 119 EEC) can also be invoked vis-à-vis a private employer, accepting not only the vertical, but also the horizontal direct effect of this provision. Yet, as seen above, many other equality rights in the area of gender, age, race and other grounds have only been substantiated in EU directives which in and of themselves lack horizontal direct effect and will only apply to private relations after their transposition into national law. While the *Mangold-Kücükdeveci-Dansk Industri-Egenberger* series of case law has established the existence of horizontal direct effect in relation to the general principle of non-discrimination and Article 21 of the EU Charter of Fundamental Rights, an important question remains regarding the extent to which consumers and users of algorithmic services can enforce their EU right to equality in light of existing restrictions and the most likely private nature of the legal conflicts occurring.[125]

While not seeking to provide a comprehensive overview and analysis of the drawbacks of the individual rights-based approach in non-discrimination law, the above sketch reveals the lack of clarity and ensuing insecurity for claimants, which considerably affects the potential success of a claim of algorithmic discrimination before courts. Besides, serious question marks must be put to the desirability of placing too much emphasis on, and faith in, individuals going to court to claim their right to equality. Comparative research on the enforcement of gender equality law in the Member States of the EU reveals a multitude of other, persisting problems that deter people from initiating legal action to protect their right to equality.[126] These range from institutional problems (e.g. length of proceedings, lack of expertise and assistance, lack of trust in the judiciary, lack of sufficient compensation), financial problems (e.g. cost of proceedings, lack of legal aid) to uncertainty about the outcome and fear of victimisation, by the employer, family and society.[127] The *Defrenne* series of cases, which lasted over a decade and asked the CJEU to rule on different aspects of the equal pay principle on three occasions, is a token of the courage and stamina that may be required from citizens who decide to enforce their rights and the lawyers who support them.[128] *Defrenne* also shows the limits of an adversarial system in relation to the effective protection of equality rights. How realistic is it to expect people to

[125] See C-144/04 *Werner Mangold v Rüdiger Helm* EU:C:2005:709; C-555/07 *Seda Kücükdeveci v Swedex GmbH & Co. KG* EU:C:2010:21; C-441/14 *Dansk Industri (DI), acting on behalf of Ajos A/S v Estate of Karsten Eigil Rasmussen* EU:C:2016:278 and C-414/16 *Vera Egenberger v Evangelisches Werk für Diakonie und Entwicklung e.V.* EU:C:2018:257.

[126] See Isabelle Chopin, Carmine Conte and Edith Chambrier, *A comparative analysis of non-discrimination law in Europe 2018: The 28 EU Member States, the former Yugoslav Republic of Macedonia, Iceland, Liechtenstein, Montenegro, Norway, Serbia and Turkey compared* (European Commission, Luxembourg: 2019) and Alexandra Timmer and Linda Senden, *A comparative analysis of gender equality law in Europe 2018: A comparative analysis of the implementation of EU gender equality law in the EU Member States, the former Yugoslav Republic of Macedonia, Iceland, Liechtenstein, Montenegro, Norway, Serbia and Turkey* (European Commission, Luxembourg: 2019).

[127] See ibid.

[128] The questions concerned the horizontal effect of ex-Art. 119 EEC, whether a retirement pension is to be considered 'pay' for the purposes of this Article, whether a different age limit for male and female crew workers is admissible, whether air hostesses and stewards are doing equal work, whether there is an entitlement to compensation for being paid less while doing the same work, whether the scope of ex-article 119 EEC extends beyond pay, etc.

go to court to obtain answers to such a multitude of questions before seeing justice done to their case? Even more so, cases like *Defrenne* have been put to the Court under the preliminary ruling procedure of Article 267 TFEU which provides a solution to specific cases only. Similar cases, existing and future ones, will only be resolved and ruled out once the legislator makes the required amendments to the law.

Algorithmic discrimination thus exacerbates the weaknesses of the individual justice approach and forces us to explore other avenues that may reduce the need for going to court in the first place. Below we consider the elements of public supervisory approach already in place and what this approach can bring for dealing with the case of algorithmic discrimination.

IV.2 The public supervisory approach

What are currently relevant institutions, procedures and tools available at the public, supervisory level that may be put to practice with a view to combatting algorithmic discrimination? What are their limits? What other private monitoring and enforcement mechanisms might be helpful in this regard?

The 'classic' EU law mechanism for supervising and enforcing compliance is the infringement procedure under Article 258 TFEU, which the European Commission can initiate against non-compliant Member States. However, an important limitation for acting against algorithmic discrimination is that only Member States can be held liable for infringement of non-discrimination law when EU directives impose obligations that need to be transposed into national law. Thus, infringement proceedings cannot be initiated against private actors. Yet, while algorithmic discrimination can be a concern in relation to public services and entities[129], complaints are likely to mostly target private actors. As a consequence, it will be the duty of national public bodies to make sure that private actors relying on machine-learning algorithms comply with EU non-discrimination law, and these duties would need to be made clear on the basis of EU law itself too. Only such national bodies could be held liable by the European Commission for not doing so, as opposed to the private actors responsible for algorithmic discrimination. Clearly, this mechanism is thus ill-equipped to deal with the issue of algorithmic discrimination.

As such, one could say that the enforcement system as it stands today relies heavily not only on national courts, but also potentially on national compliance bodies and agencies, especially because EU equality law is a domain that so far lacks strong supervisory EU agencies as we witness them in other domains of EU law such as financial services, migration, air control, chemicals, fisheries etc.[130] In these domains, the EU has recognised the limits of the individual

[129] See e.g. Henrik Palmer Olsen, Jacob Livingston Slosser, Thomas Hildebrandt, Cornelius Wiesener, 'What's in the Box? The Legal Requirement of Explainability in Computationally Aided Decision-Making in Public Administration.

[130] One example is the European Securities and Markets Authority (ESMA). See M. Scholten, Mind the Trend! Direct enforcement of EU law and policies is moving to 'Brussels', <http://eulawenforcement.com/?p=30> and M. Scholten and M. Luchtman, Law Enforcement by EU authorities. Implications for Political and Judicial Accountability, Edward Elgar, 2017.

rights based approach and instead of merely relying on domestic institutions for the protection of EU law rights and obligations, moved towards a more centralised enforcement. In the area of equality and non-discrimination, the only agencies in place are the European Institute of Gender Equality (EIGE)[131] and the Fundamental Rights Agency (FRA),[132] whose mandate only covers data collection, information, expertise and research. Two questions then impose themselves in this regard: which role could these European bodies possibly play in the future and which domestic bodies or agencies in particular could be called upon to tackle the problem of algorithmic discrimination? To begin with, in our view, EIGE, the FRA and national equality bodies could initiate awareness-raising campaigns and conduct studies on the topic of algorithmic discrimination. As there is currently little knowledge about algorithmic discrimination both among citizens and members of the bar and the judiciary, these campaigns and studies would help expose the diverse manifestations of algorithmic discrimination and could provide legal practitioners with guidance on building and treating cases of algorithmic discrimination. Depending on their mandate, national equality bodies could also play an important role in supporting individual claims, initiating class actions and bringing the issue to the attention of the legislator. The FRA started to study the impact of artificial intelligence and algorithms on fundamental rights in 2018, producing a focus paper entitled '#BigData: Discrimination in data-supported decision-making',[133] while the EIGE does not yet seem to have done so.[134] Equinet, the umbrella organisation at EU level for national equality bodies, has also recently called for proposals for a study on the equality implications of artificial intelligence and the role of equality bodies in tackling algorithmic discrimination.[135] The fact that the Finnish Presidency of the Council of Europe has organised a conference on this topic in February 2019 also shows increasing political awareness.[136] These are positive signals regarding public involvement in the near future.

Other legal building blocks might be of crucial importance to ensure the enforcement of non-discrimination law in the age of machine-learning and AI. Awareness has for instance been triggered within the framework of the General Data Protection Regulation (GDPR). This is the first piece of European legislation that recognises the phenomenon of algorithmic discrimination and sets rules and procedures to combat it, therewith providing a somewhat more specific monitoring and enforcement toolkit. In recital 71 of its preamble, it lays down a requirement to "implement technical and organizational measures" that "prevent, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or

[131] See EIGE's website: https://eige.europa.eu/about.
[132] See FRA's website: https://fra.europa.eu/en.
[133] Fundamental Rights Agency, '#BigData; Discrimination in data-supported decision-making', FRA Focus (May 2018) https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-focus-big-data_en.pdf accessed 3 May 2019.
[134] Its website not given any search results when using the keyword of 'algorithm[s]' or 'algorithmic'.
[135] See Equinet, 'Call for proposals : Equality, artificial intelligence and algorithmic discrimination' (24 April 2019) <http://equineteurope.org/2019/04/24/call-for-proposals-equality-artificial-intelligence-and-algorithmic-discrimination/> accessed 21 September 2019.
[136] See https://www.coe.int/en/web/portal/-/artificial-intelligence-helsinki-conference-conclusions.

that result in measures having such an effect".[137] The Regulation contains two key principles to address algorithmic discrimination: data sanitization, concerning the removal of special categories from datasets used in automated decision making, and algorithmic transparency, entailing a right to explanation.[138] The latter entitles data subjects to "meaningful information about the logic involved, as well as the significance and the envisaged consequences" when automated decision making or profiling takes place. Such information must be provided "in a concise, transparent, intelligible and easily accessible form, using clear and plain language".[139] The GDPR further refers to three specific tools by which these principles can be enforced: data impact assessments, codes of conduct and certification.[140] Therewith the GDPR does not only refer to the importance of data protection authorities – and potentially also of the European Data Protection Supervisor (EDPS) itself – in carrying out such impact assessment, but it also alludes to third-party auditing as a controlling mechanism with a view to certification.[141]

The case of data protection thus demonstrates how domestic public institutions in a specific area could be imposed such a monitoring and controlling duty, but also how private bodies and actors may possibly be engaged in such an exercise. The tools and procedures to be used for such purpose are not new as such, but already familiar to, and practised in, several EU law and policy domains. Impact assessment is a tool that has been progressively developed for over two decades now, and certification is an established instrument in relation to product safety, health and environmental inspections. These could be further developed as detection and enforcement tools, beyond the area of data protection in the field of equality and non-discrimination. Interestingly, we already witness such developments in the Member States, such as the Artificial Intelligence Impact Assessment, developed and proposed by the Dutch platform for the information society in December 2018.[142] Such initiatives can alleviate the burden imposed on service and goods providers to check whether the algorithmic models they buy from IT engineers are non-discriminatory. Development of third parties such as certification agencies, firms or branches in IT companies responsible for ethical questions, which will test algorithms for discrimination and certify them, is thus already a viable option within the framework of the Union's current toolbox. Besides the use of impact assessments, the FRA has also emphasised the need to involve different experts in oversight, stating that "reviews need to involve

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

[138] B.W. Goodman, A Step towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection, http://www.mlandthelaw.org/papers/goodman1.pdf See arts. 9 and 22 of the GDPR regarding the data sanitization principle.

[139] Art. 13(2)(f); art. 14(2)(g) and art. 12.

[140] Arts. 24 [impact assessments], 40 [codes of conduct], 42 [certification].

[141] See Goodman, o.c. In addition, Art 9 GDPR prohibits the "[p]rocessing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation" and Art. 22 GDPR protects data subjects from being subjected to "decision[s] based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her".

[142] See <https://ictinstitute.nl/the-artificial-intelligence-impact-assessment/>.

statisticians, lawyers, social scientists, computer scientists, mathematicians and experts in the subject at issue”.[143]

As such, quite some avenues could be further explored and developed within the framework of an ‘equality by design’ approach, that is an approach to algorithmic engineering that would take into account the concerns linked to equality and non-discrimination in each step of the algorithmic design and building process, so as to establish a more capable regulatory and enforcement framework to deal with algorithmic discrimination.[144]

V. Conclusions

In this contribution, we have mapped in what ways the use of machine-learning algorithms may lead to discrimination and to what extent this may collide with, and be covered by, the current EU legal framework. We found that while the body of EU equality law contains useful yardsticks to deal with algorithmic discrimination in the field of the labour market and access to goods and services, there are also important limitations to it. This is so, amongst others, because of discrepancies in the personal and material scope of the EU equality law directives. The lack of a coherent and comprehensive equality law framework regarding the various grounds of discrimination, including race, gender, religion, disability, sexual orientation and age, thus also appears problematic with regard to digital forms of discrimination. Furthermore, the complexity of the human-machine relationship might lead to a blurring of the conceptual difference between direct and indirect discrimination, as it ensues from the Court’s case law and has been consolidated in the EU equality directives. At the level of enforcement, it has become clear that existing problems connected to the still strong reliance on the individual rights-based approach are exacerbated in relation to algorithmic discrimination. This concerns inter alia problems of transparency, access to information, liability, burden of proof and effective judicial protection. There is thus a clear need to reflect, in further research, on how to effectively address these problems, as well as on which public supervisory mechanisms might be put into place to alleviate the burden imposed on individuals affected by such discrimination. All in all, the limitations and hurdles contained in the current EU equality law regime are in need of a more in-depth analysis, as well as the opportunities it may provide for better tackling this new form of digital discrimination.

At the same time, however, it must also be noted that while machine-learning algorithms sharpen the risk to see certain types of discrimination flourish, they also reduce the likeliness that other kinds of discrimination take place. Hence these new technologies represent both a

[143] See <https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-in-brief-big-data-algorithms-discrimination_en.pdf>.

[144] On ‘equality by design’ approaches in the platform economy, see e.g. Barzilay and Ben-David, 'Platform Inequality: Gender in the Gig-Economy', 430. The authors define equality by design as “the structuring of platforms in a manner that is sensitive to prevailing forms of gender discrimination, in ways that extend beyond merely omitting gender as a formal element of platforms’ template for profiles or not portraying women in a biased manner”.

risk and a chance for equality. The question of discrimination should first of all be understood not as a fatality inherent in algorithms but rather in the context of a trade-off. As the statistician George Box pointed out as early as 1979, "[a]ll models are wrong but some are useful".[145] The question is how to conceive of this usefulness and what values to prioritise. In particular, the trade off at stake is often one between ease of access, affordability and processability of data, accuracy of algorithmic output (e.g. predictions) and ethical considerations such as fairness and non-discrimination. The concept of equality by design thereby provides an opportunity for decision-makers and legislators to provide clear criteria for weighing in equality and non-discrimination in the context of this trade off. The challenge of algorithmic discrimination is therefore at the same time a chance for expressing a clearer and more concrete commitment to equality, and for establishing an operational framework for balancing equality with other concerns.

If algorithms enhance discrimination issues in some cases, they also provide an opportunity for reduced arbitrariness through increased rationality and explainability in decision-making procedures. While human decisions might also be called a "black box" because of their opaque and non-replicable nature, machine-learning algorithms offer a chance for more accountable decision-making, provided that certain transparency requirements are met.[146] Where human decisions cannot be reproduced changing one factor to test where discrimination comes from, algorithmic decisions might well be replicable in such a way, offering more control on how discrimination spreads and reproduces provided that a sufficient level of awareness exists. Therefore, certain principles such as transparency, explainability and accountability are fundamental to developing artificial intelligence applications if the aim is to turn existing risks of discrimination into an opportunity for increased equality. Devising and ensuring that these principles are respected along the entire algorithmic design chain will require a holistic multidisciplinary approach in which computer scientists, lawyers and social scientists, psychologists, sociologists, philosophers, etc. will have to join forces.

[145] George Box, 'Robustness in the strategy of scientific model building' in Robert Launer and Graham Wilkinson (eds), *Robustness in Statistics* (Academic Press 1979) , 202.

[146] See Frank Pasquale, *The Black Box Society: The Hidden Algorithms Behind Money and Information* (Harvard University Press 2015).