

Discrimination Revised. How AI Is Reshaping Anti-Discrimination Law*

Costanza Nardocci

TABLE OF CONTENTS: 1. Introducing AI-Based Discrimination. – 2. What's New, Why, and What Is Not. – 3. When EU (But Also, Common Law) Anti-Discrimination Law Falls Short. – 4. The Pitfalls of Understanding AI-Based as a Pure Human-Driven Discrimination: The Proxy, Proxy Discrimination, and the New Victim Paradigm. – 5. Did Someone Say, "Discriminatory AI"? Regulating and Sanctioning AI-based Discrimination. – 6. The Law Does not Say, but: Procedural Remedies at the Times of AI-Based Discrimination. – 7. Conclusions: A Call for Awareness.

1. Introducing AI-Based Discrimination

When discussing artificial intelligence and human rights, discriminatory AI is always quite present in the public and academic debate¹.

Despite the undoubted and socially accepted relevance of the discriminatory implications of AI², existing and prospective regulations

* This article was subjected to double-blind peer review.

¹ The literature that first began to examine extensively the discriminatory implications associated with artificial intelligence technologies is the Anglo-American one. In particular, it is worth highlighting at the outset the essay by D.K. Citron and F. Pasquale, *The Scored Society: Due Process for Automated Predictions*, in *Washington Law Review*, 2014, p. 1 ff. The work is highly interesting, offering an overview of the main critical issues characterising the functioning of artificial intelligence technologies—opacity, arbitrariness of the determinations made, disproportionate impact on certain groups—as well as illustrating several proposals to mitigate them. These may be summarised in the adoption of policies and regulations ensuring the transparency of AI systems, an assessment of their impact and recommended uses, and the protection of individuals' rights. See also, from the early years in which the topic began to attract attention, T. Zarsky, *Understanding Discrimination in the Scored Society*, in *Washington Law Review*, 2014, p. 1375 ff.

² On the discriminatory implications of AI, The issue gained significant public attention starting in 2014, following the publication of the White House report [Big Data: Seizing Opportunities, Preserving Values](https://obamawhitehouse.archives.gov/sites/default/files/2014/02/big_data_seizing_opportunities_preserving_values.pdf), available at obamawhitehouse.archives.gov, which in particular includes several examples of discriminatory effects produced through the use of artificial intelligence techniques. The most well-known example concerns the consequences

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

seldom question whether anti-discrimination law should be reconsidered in light of the torsion discrimination encompasses when coupled with AI.

The consequences of such a poor investigation into the here defined AI-based discrimination are the scarce consideration of how artificial intelligence technologies may eventually result in unreasonable treatments towards individuals and social groups, and the absence of tailored normative responses in the recently adopted, and also perspective, AI legislations worldwide³.

generated by the Street Bump App, used in the city of Boston and developed in collaboration with the Mayor's Office of New Urban Mechanics. The app sought to use data collected from residents' smartphones to assess which areas of the city required redevelopment, through an analysis of road and neighborhood conditions. The main issue with the app stemmed from the fact that economically poorer groups were less likely to own a smartphone on which the app could be installed. As a result, municipal interventions would have focused only on neighborhoods with higher concentrations of residents belonging to more affluent socioeconomic classes. The app therefore ended up discriminating against certain groups of residents on the basis of economic conditions. The literature began to explore more deeply the intersections between artificial intelligence and discriminatory phenomena particularly after the publication of the 2014 White House Report. An interesting study has shown that, among works published from 2014 to 2019 in indexed journals, only 14 studies addressed the issue from the perspective of its legal implications and consequences. This refers to the work of M. Favaretto, E. De Clercq and B. Simone Elger, *Big Data and Discrimination: Perils, Promises and Solutions. A Systematic Review*, in *Journal of Big Data*, vol. 6, no. 12, especially p. 5 ff. An even more recent study highlights the existence of a significant gap between research conducted in the United States on artificial intelligence and discrimination, and the current state of theoretical development in continental Europe. This refers to S. Wachter, B. Mittelstadt and C. Russell, *Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI*, in *Computer Law & Security Review*, 2020, p. 1 ff. Recently, in this field, see also the well-known works of K. Crawford, including *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, New Haven, 2021. In previous decades, the literature dealt only occasionally with the role of bias in the functioning of technologies. In particular, see B. Friedman and H. Nissenbaum, *Bias in Computer Systems*, in *ACM Transactions on Information Systems*, 1996, p. 330 ff.; M. Bruce and A. Adam, *Expert Systems and Women's Lives: A Technology Assessment*, in *Futures*, 1989, p. 480 ff.

³ Reference is chiefly to the European Union Regulation, the so-called "AI Act", Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence, but also to the other existing laws on AI, as in the case of the laws recently adopted by Japan and, before it, by North Korea. On the scarce relevance attributed to discriminatory AI in the regulatory framework see paragraph no. 5.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

Contrary to that approach is, instead, the understanding that discrimination possesses specific traits in the AI realm, which separates it from the already known human-driven discrimination⁴.

The juxtaposition between AI-based and human-driven discrimination essentially lies in the supplementary and sometimes even protagonist role of the machine *vis-à-vis* the humans, which ultimately leads to alternative, if not heterogeneous, manifestations of discriminatory treatments.

The article intends to illustrate the core elements featuring AI-based discrimination. It untangles the relationships AI-based discrimination entertains with EU anti-discrimination law, looking at how direct and indirect discrimination react towards discrimination generated by AI technologies. It then proceeds to argue that AI-based discrimination could not be misconceived as a pure manifestation of human deliberate or unintentional willingness to discriminate, but, rather, as a separate form of unreasonable treatment, where humans and automatic agencies intersect with one another.

Through the comparison between human-driven and AI-based discrimination, the paper eventually questions the adequacy of existing anti-discrimination laws to tackle discriminatory AI and advocates for the introduction of a new set of mechanisms to counter the prejudicial effects deriving from discriminatory AI.

2. *What's New, Why, and What Is Not*

At first glance, someone could be tempted to argue that nothing changes when discrimination meets AI, as there is always a human in the loop or, at least, behind AI⁵. That is a recurrent statement when considering the leading role played by humans in the preliminary phases of the building

⁴ On this, extensively, C. Nardocci, *Algoritmi, eguaglianza, discriminazione. Le sfide dell'intelligenza artificiale*, Turin, 2025.

⁵ This is true particularly with regard to the selection of the data fed to AI systems, that, instead of truthfully represent the external reality, are intrinsically human and, therefore, partial. On this, see K. Crawford, *The Hidden Biases in Big Data*, in *Harvard Business Review*, 2013, who emphasizes that: «[d]ata are assumed to accurately reflect the social world, but there are significant gaps, with little or no signal coming from particular communities. While massive datasets may feel very abstract, they are intricately linked to physical place and human culture». On the relationship between data and human culture, see, also, L. Gitelman, *"Raw Data" Is an Oxymoron*, Cambridge (MA), 2013.

Costanza Nardocci
Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

and functioning of AI technologies and, first, in the choices related to the dataset, which could suggest an overlap between human-driven and AI-based discrimination. In short, if AI is just a product of human behaviour, discrimination derived from AI technologies cannot be conceived as something different from the paradigm of discrimination caused by human conduct. In this sense, AI should be just something added to the traditional scheme, incapable of interrupting the causal link between the discriminatory conduct and its effects.

The denial of AI agency, thus, leads to the enhancement of human agency, neglecting the differences and, even sooner, the legitimacy of a debate on the existence of a new form of discrimination, which is derived or partially caused by AI technologies.

Contrary to that, though, AI possesses agency. Somehow, sometimes, it shows alternative features compared to human agency, but it does.

The recognition of AI's agency in the discrimination discourse contributes to explaining the divergent path of discrimination when the conduct stops being entirely human, resulting instead in a wide and extremely varied range of intersections between humans and AI technologies. While in human-driven discrimination, they are the only ones bearing the responsibility for the violation of the principles of equality and non-discrimination, in AI-based discrimination, humans are not alone in generating the discriminatory outcomes.

It is, thus, primarily the conduct that is subjected to profound changes when the unreasonable difference in treatment is also caused by AI.

The heterogeneity between human-driven and AI-based lies, therefore, and first of all, in the conduct: exclusively human in the case of human-driven discrimination; mixed or half human and half automatic in the paradigm type of AI-based discrimination.

The literature has thoroughly investigated the ways AI might discriminate by distinguishing a number of phases that pose discriminatory risks⁶. These are: the selection of the data fed into the machine; the training of the data, which could be autonomous run or, conversely, guided by the programmer; the identification and selection of the “target variables” and “class labels”, which are used to group into categories; the “feature selection”, meaning the choice of the features used by AI⁷; more

⁶ On this, extensively, S. Barocas and A.D. Selbst, *Big data disparate impact*, in *California Law Review*, vol. 104, 2016, p. 671 ff.; F.Z. Burgesius, *Discrimination, artificial intelligence, and algorithmic decision-making*, Council of Europe Publications, 2018.

⁷ The concept of features selection refers to the process by which the programmer and, subsequently, the AI system decides which attributes use in making its choices.

Costanza Nardocci
Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

fundamentally, the choice of the “proxy” as the element AI will refer to make distinctions which may eventually turn out to be of a discriminatory nature.

A few points might be made regarding the selection and training of the data.

The first, mentioned already, is the selection of the data fed into the machine. During this phase, bias might act as one of the key factors leading to discrimination. The heterogeneity of human biases widely affects the quality of the data and their representation of the outside reality in cases of over- or under-representation of certain groups or categories, which are usually those already suffering from structural discrimination⁸.

The second step or phase is the training of the data, which could be autonomous run or, conversely, guided by the programmer. Depending on the type of AI technologies, this phase may, therefore, show a more or less determinant influence on the programmer: the more AI functions as a machine learning system, the less the programmer will be able to control and supervise the outcome of the technology at stake. Put differently, the machine will take precedence over the human in the possible discriminatory outcome of the machine.

The training of the data is also relevant as it could witness two interesting phenomena: the poisoning of the data and the inaccurate or updated information provided to the machine in light of scientific and technological innovation. Data poisoning occurs when the human – and here the liability lies almost entirely with the programmer – deliberately feeds the machine with data that is poisoned, unhealthy, untruthful, and misleading. Instead, the latter case concerns AI technologies whose dataset has not been updated in a way consistent with the advancement and development of innovation, therefore impairing AI’s ability to respond to the tasks assigned.

Besides the specifics of these phases, whose analysis goes beyond the scope of the investigation, what matters the most is that each of these shows the multifaceted forms that the discriminatory conduct takes in the context of AI-based discrimination and demonstrates the non-exclusive, but rather concurrent, role of the human in the series of actions that eventually lead to the discriminatory effect.

But the analysis of the conduct is not enough to fully understand AI-based discrimination and to grasp its argued heterogeneity *vis-à-vis* human-

⁸ See, on this, J. Lerman, *Big Data and Its Exclusions*, in *Stanford Law Review Online*, vol. 66, 2013, p. 55 ff.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

driven discrimination. While the conduct exhibits new features, because of the said mediation of the machine in the chain of events leading to the unreasonable difference in treatment, the origins or causes of discrimination do not differ in the two types of discrimination.

In line with the argument that sees discrimination rooted in collective prejudices endorsed by the dominant group towards minority communities (the so-called out-groups) resulting from inter-ethnic conflicts⁹, both human-driven and AI-based discrimination originates from overlapping causes nurtured in human nature. Therefore, the origin of discrimination does not separate human-driven from AI-based, in that in both cases human prejudices represent the primary cause of the discriminatory conduct that follows through.

However, despite the human origin of discrimination, which may suggest that AI should be cut off from the chain linking the causes to the discriminatory effects, AI does impact the ways discrimination reveals itself and negatively impacts individuals and social groups. In other words, while the primary causes of discrimination remain a human domain, AI has opened the doors to a variety of obscured and often unpredictable discriminatory conduct that hardly reconcile with the traditional categories of EU anti-discrimination law.

The mismatch between how human discrimination has been captured and sanctioned under the law and the novelties AI-based ones have brought into the legal scenario justifies the need to reconsider whether and to what extent anti-discrimination law is adequate and should be applied to effectively tackle discriminatory conduct resulting from an interplay between humans and AI systems.

⁹ Reference is made to the so-called conflict power theories and to D.L. Horowitz, *Ethnic Groups in Conflict*, Oakland, 1985. On this, see also M.N. Marger, *Race and Ethnic Relations. American and Global Perspectives*, Boston, 2009; A.D. Smith, *The Ethnic Revival*, Cambridge, 1981, and Id., *The Ethnic Origins of Nations*, Chichester, 1988.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

3. *When EU (But Also, Common Law) Anti-Discrimination Law Falls Short*

Framing AI in the context of EU anti-discrimination law is crucial to test the heterogeneity of AI-based discrimination from the legal and traditional conceptualization of discrimination.

From this angle, the theories of direct and, to some extent, even indirect discrimination prove to be inadequate to intercept and describe the core features of discrimination generated by AI¹⁰.

Beginning with the former, under EU law, direct discrimination implies the recurrence of the following elements: a difference in treatment among two or more comparable situations; the lack of an objective and reasonable justification underneath the distinction between the two or more comparable situations; the explicit reliance on one or more legally suspect grounds, the so-called factors of discrimination, that act as the basis of the difference in treatment.

Moreover, direct discrimination also requires that all the elements described above should also be supported by the proof of the recurrence of a direct and causal link between the conduct and the discriminatory effects, which means that there would not be direct discrimination should the effects not be caused by a human action.

Given the elements featuring the concept of direct discrimination under EU law, AI-based discrimination deviates from it for several reasons¹¹.

¹⁰ The more recent literature appears increasingly inclined to argue that traditional anti-discrimination law is inadequate to address disparities in treatment stemming from the functioning of AI technologies. This holds true for both EU anti-discrimination law and U.S. anti-discrimination law. See, in this regard, T.B. Gillis and J.T. Spiess, *Big Data and Discrimination*, in *The University of Chicago Law Review*, vol. 86, 2019, p. 458 ff.; S. Barocas and A.D. Selbst, *Big Data's Disparate Impact*, cit.; C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York, 2017; and, in greater detail, R. Xenidis, *Tuning EU Equality Law to Algorithmic Discrimination: Three Pathways to Resilience*, in *Maastricht Journal of European and Comparative Law*, vol. 27(6), 2020, p. 736 ff.; M. Mann and T. Matzner, *Challenging Algorithmic Profiling: The Limits of Data Protection and Anti-Discrimination in Responding to Emergent Discrimination*, in *Big Data & Society*, 2019, p. 1 ff.; M. Lees, *The New Profiling: Algorithms, Black Boxes, and the Failure of Anti-Discrimination Law in the European Union*, in *Security Dialogue*, vol. 45(5), 2014, p. 494 ff.

¹¹ On this, see P.N. Geslevich and Y. Le Aretz, *Learning Algorithms and Discrimination*, in W. Barfield and U. Pagallo (eds), *Research Handbook of Artificial Intelligence and Law*, Cheltenham, 2018, p. 88 ff.; J. Adams-Prassl, R. Binns and A. Kelly-Lyth, *Directly Discriminatory Algorithms*, in *Modern Law Review*, vol. 86, 2023, p. 144 ff. Also sharing the view

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

The first mismatch arises from the conduct that originates the unlawful distinction. While in human-driven discrimination, the difference in treatment follows a human action, in the context of AI-based discrimination, vice versa, the conduct represents, as said, the product of an agency that is mixed, not centred on the human only, but rather, dependent on the human and machine simultaneous interaction. That also means that, when AI is involved, the causal link between the conduct and its effects is complicated by the mediation of the machine, which plays a role altogether with the human in perpetrating discrimination.

Moreover, given that AI intersects with humans in extremely diverse ways depending on the specifics of the AI technology, in the case of AI-based discrimination, AI's contribution to the discriminatory outcome will likely feature various levels of intensity. Therefore, as a result, AI may: influence an already discriminatory human conduct, without adding much to an already enacted unlawful treatment; contribute with more or less pervasive degrees to discriminate together with the human; become the main responsible for the discriminatory effect, as in all cases when human agency is minimized.

Regardless of the specifics of each case, the identification of the causal link between the conduct, infiltrated by the AI system, and the discriminatory effect will often be not seldom prevented by the often-obscure functioning of the AI technology.

Besides the conduct, a second deviation from the structural elements of direct discrimination covers the element that the unlawful treatment chooses to rely on in unreasonably differentiating the said comparable situations¹². Whereas in direct discrimination, the distinction should always and explicitly be based on one or more factors of discrimination, AI-based discrimination seldom grounds its distinctions on legally suspect features, opting instead for a different element, the proxy, which acts like a suspect

that so-called “algorithmic discrimination” cannot be framed within the doctrinal structures of direct discrimination are C.S. Yang and W. Dobbie, *Equal Protection Under Algorithms: A New Statistical and Legal Framework*, in *Michigan Law Review*, vol. 119(2), 2020, p. 91 ff.; D. Hellman, *Measuring Algorithmic Fairness*, in *Virginia Law Review*, vol. 106(4), 2020, p. 811 ff.; J.R. Bent, *Is Algorithmic Affirmative Action Legal?*, in *The Georgetown Law Journal*, vol. 108, 2020, p. 803 ff.; R. Xenidis, *Tuning EU Equality Law to Algorithmic Discrimination*, cit., p. 747, who observes that «it [is] difficult for algorithmic proxy discrimination to be considered as direct discrimination because its definition in EU law involves a causality link between a given treatment and a protected ground, while inferential analytics rely on correlations».

¹² On this, see, extensively, A. Datta et al., [Proxy Non-Discrimination in Data-Driven Systems](#), 2017, available at *arXiv*.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

ground without formally being one¹³. Here, being the proxy an element that is "other" compared to the traditional factors of discrimination, unlawful treatments grounded on proxies, although predictive of individual affiliation to protected features, won't be sanctioned under the law¹⁴.

In light of the above, AI-based discrimination sits uneasily with the legal understanding of direct discrimination, whose structural aspects are either missing or altered¹⁵.

¹³ There is no universal definition of what a proxy is. Nonetheless, the literature offers some coordinates. Among others, G. Karger offers a reconstruction of several definitional proposals in *The Proxy Problem in Disparate Treatment*, 2024, available at ssrn.com, where it also refers to the definition proposed by L. Alexander in *What Makes Wrongful Discrimination Wrong? Biases, Preferences, Stereotypes, and Proxies*, in *University of Pennsylvania Law Review*, vol. 141, 1992, p. 149 ff. According to Alexander, p. 167, proxies are «traits [...] which, though immaterial in themselves, we believe to be highly correlated with those traits in which we are primarily interested». Other authors describe the proxy as an attribute, without offering a definition that goes beyond stating its heterogeneity vis-à-vis traditional grounds of discrimination. In this perspective, M. Lees, *The New Profiling*, cit.; R. Gellert, K. de Vries, P. de Hert and S. Gutwirth, *A Comparative Analysis of Anti-Discrimination and Data Protection Legislation*, in B. Custers, T. Calders, B. Schermer and T. Zarsky (eds), *Discrimination and Privacy in the Information Society*, Berlin, 2013, p. 61 ff., who, with regard to the functioning of AI technologies, note that: «discrimination will not concern any of the protected grounds, but rather attributes such as income, postal code, browsing behaviour, type of car, etc., or complex algorithmic combinations of several attributes».

¹⁴ The topic evokes that of the opportunity to potentially enlarge the list of the legally suspect grounds provided under national Constitutions and international human rights law treaties. Reference is made to the The reference is to the soundness – understood as appropriateness in light of the objectives pursued by the legislator (thus, reasonableness) – or lack thereof of the option that favours selecting a more or less limited set of individual qualities that cannot serve as grounds for distinguishing between two comparable situations. This is a topic addressed in the case law of constitutional and national courts which, however – at least within Europe and as a product of the civil law tradition – have rarely departed from the provisions of positive law, relying instead on extensive interpretations, whether of constitutional or statutory rank when viewed from the standpoint of national legal systems, or of primary or secondary law when the relevant framework is that of the European Union. This is also the reason why, as will be discussed, the notion of a proxy, which is central to understanding AI-based discrimination, was already well established in the U.S. context – more inclined to expand suspect grounds through judicial interpretation – whereas in Europe, by contrast, it was only the advent of new technologies that thrust the notion of proxies into the spotlight, making both its legal definition and the clarification of its role in the functioning of artificial intelligence systems urgent.

¹⁵ While in direct discrimination there is no additional element to take into account to evaluate the existing difference with the ways AI discriminates, in the context of the disparate treatment theory, there is an additional element to consider that is the lack

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

What is missing, in short, is the expressed reference to a suspect ground protected as such under the law. Altered, *vice versa*, is the conduct that turns out to be the product of an intertwined relationship between the human and the machine instead of being entirely human.

Lastly, while in direct discrimination there is no doubt that a human will eventually be held responsible for the unlawful treatment, in AI-based discrimination, the establishment of the liability will require separating the human involved from AI and identifying what should be attributed to the former to ground his/her responsibility before the law. A process that will eventually be impaired by the often-obscure functioning of the AI system involved¹⁶.

Besides direct discrimination, indirect discrimination also proves to be problematic when coupled with AI-based discrimination¹⁷.

Although ideally more equipped to intercept AI-based discrimination, in that in both cases the unlawfulness insists on the impact (the effect) rather than on the treatment (the conduct), which could more effectively respond to the shortcomings of applying the theory underneath direct discrimination, indirect discrimination proves, nonetheless, to be seemingly inadequate to describe and, thus, functions as a background framework for sanctioning discriminatory AI¹⁸.

The main points of departure from the traditional scheme of indirect discrimination revolve around three main arguments.

The first lies in the fact that discrimination arising from the use of AI originates in differences based on factors that operate as predictive indicators of an individual's membership in a protected class (so-called proxies). This feature prevents the satisfaction of the first constitutive element of the concept of indirect discrimination (and, by extension, disparate impact), which instead presumes that the unequal treatment results from a facially neutral rule, drafted in such a way as to preclude, *prima*

of intentionality behind the conduct leading to the AI-based discriminatory effect. Intentionality cannot, in fact, be said to feature a conduct that sees a participation of an automated system clearly incapable of wanting to generate a discriminatory effect. In the literature, on the concept of intentionality associated to AI-based discrimination, see A.Z. Huq, *What is discriminatory intent?*, in *Cornell Law Review*, vol. 103, 2018, p. 1211 ff.; S. Fredman, *Discrimination law*, Oxford, 2011.

¹⁶ See, again, on this and among many others, M. Lees, *The new profiling*, cit.

¹⁷ In support of this thesis, see, extensively, M. Mann and T. Matzner, *Challenging Algorithmic Profiling*, cit.

¹⁸ See, for the same thesis, V. Calderon, *Unintentional algorithmic discrimination: how artificial intelligence undermines disparate impact jurisprudence*, in *Duke Law & Technology Review*, vol. 24, 2024, p. 29 ff.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

facie at least, any disparate treatment. To be sure, even in traditional cases of indirect discrimination, the factor underpinning a facially neutral distinction may itself be somehow predictive of protected-class membership, much like a proxy in algorithmic decision-making.

Yet the structure of AI-based discrimination diverges in crucial ways.

In algorithmic contexts, the relevant correlation need not arise between a conventional ground of discrimination and an apparently neutral factor predictive of it. Rather, the correlation may emerge among multiple proxies themselves, rather than between a single proxy and a traditional protected ground, as in cases of traditional, meaning purely human, indirect discrimination. In this respect, it is precisely the interlocking of proxies – and their attenuated, mediated relationship with traditional protected characteristics – that sets the discriminatory operation of algorithmic systems apart from the conventional model of indirect discrimination¹⁹.

That being said, some connection with a traditional discriminatory ground must ultimately exist, for the absence of such a link would negate the discriminatory character of the conduct altogether. Even so, proving the existence of this correlation – whether between the predictive factor employed by the algorithm and the protected class, or among proxies that collectively reproduce the effect of that class – poses formidable evidentiary challenges. These challenges stem from the defining features of artificial intelligence systems: opaque data, opaque correlations, and opaque mechanisms of operation.

Ultimately, there must be some connection with the traditional discriminatory factor, since its absence would, from the outset, exclude the very possibility of discrimination. Nevertheless, proving the existence of a correlation between the predictive element employed by the machine and the discriminatory factor – or between the proxies and the latter – in judicial proceedings is a task that will eventually be complicated by the well-known features of AI systems: the opacity of data, the obscurity of the existing correlations, the lack of transparency concerning how the machine processes and utilizes them.

¹⁹ It should also be noted, as the literature highlights, that the elements underpinning the degree of overlap between the discriminated group and the protected group are themselves unclear and discretionary. In this sense, R. Xenidis, *Tuning EU Equality Law to Algorithmic Discrimination*, cit., especially p. 747. These critical issues are bound to increase when the discriminated group displays characteristics, that are heterogeneous in relation to the classical representation of the social group, understood here as a category.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

A second aspect concerns the construction of the model more specifically. As noted, indirect discrimination is defined on the basis of the disproportionately adverse impact experienced by one category compared to another, which can be assessed by examining the effects of the conduct.

The problem that arises in the context of AI is that the dataset often does not adequately mirror the outside reality, as it fails to faithfully reproduce the relationships, or more precisely, the proportions between groups and categories in their actual sizes. This results in so-called data sets containing groups or categories that are either underrepresented or overrepresented according to proportions that do not necessarily reflect reality.

In both cases, the lack of neutrality inherent in the data upon which AI operates makes it difficult to demonstrate the occurrence of indirect discrimination in terms of a comparative assessment of effects, because the reference framework, the dataset, is incapable of reproducing the external phenotypic heterogeneity. Demonstrating the occurrence of a disproportionately greater discriminatory effect is, therefore, particularly challenging²⁰.

Another criticism concerns the *tertium comparationis*²¹. The opaque functioning of AI systems, combined with the impossibility of accessing the dataset²², may, in fact, prevent the identification of the comparator, thereby hindering the comparative assessment required to prove the disparate effect in a given case. Moreover, the predictive capacity of the elements employed by AI to make distinctions – the proxies – likewise complicate the

²⁰ Equally problematic is the very notion of particular disadvantage or of proportionately greater prejudice which, as noted by some, remains insufficiently defined. This lack of clarity exacerbates cases of AI-based discrimination that could in principle be sanctioned — because they fall under the framework of indirect discrimination — but that should instead escape censure due to the difficulty of qualifying the discriminatory effect as “disparate”. In this regard, see R. Xenidis, *Tuning EU Equality Law to Algorithmic Discrimination*, cit., p. 747, who rightly observes that: «the definition of the “particular disadvantage” to be experienced by a protected group is unclear [...] under EU law [and] has not received a consistent interpretation by the Court of Justice»; similarly, S. Wachter, *Affinity Profiling and Discrimination by Association in Online Behavioural Advertising*, in *Berkeley Technology Law Journal*, vol. 35, 2020, p. 367 ff., at 402-403.

²¹ On this, see E. Lundberg, [*Automated decision-making vs indirect discrimination. Solution or aggravation?*](#), 2019.

²² This is an aspect on which regulatory interventions are making significant progress, with a strong emphasis on ensuring transparency requirements. Such transparency should properly extend both to the construction of the dataset and to all the stages governing the operation of the artificial intelligence system, insofar as this is possible given the often-noted opacity of the more complex AI technologies.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

identification of the targeted group disproportionately affected by AI. As it will be further discussed, proxies *favor* the development of new alternative, unstable, and dynamic groups²³, which are difficult to map onto the categories that are identified by legally established traits and that are traditionally qualified as factors of discrimination, therefore, preventing or limiting the comparative assessment.

Lastly, as in the case of direct discrimination, causation is problematic even in the context of indirect discrimination, where the obscure mechanisms that guide the functioning of AI may greatly hinder the proof of the recurrence of indirect discrimination.

The limits of EU anti-discrimination law to describe the features of AI-based discrimination, or, put differently, the inadequacy of the theories of direct and indirect discrimination to apply to discriminatory AI, appear even more evident by looking at the element that more than any other represents AI-based discrimination prominently, the proxy.

4. *The Pitfalls of Understanding AI-Based as a Pure Human-Driven Discrimination: The Proxy, Proxy Discrimination, and the New Victim Paradigm.*

«[W]e shouldn't use race because essentially it creates this negative feedback loop, then you say, OK, well, OK, let's not use race, but should we use zip code, which of course is a proxy for race in our segregated society?

And so once they acknowledge that zip code is just as good as race, then you're like, OK, so how do we choose our attributes? Because there are so many proxies to race. And it's really actually very tricky.

It's tricky. And I'm not trying to claim that it's easy.”²⁴

A second area where AI-based discrimination diverges from human-driven discrimination is the reason behind the unlawful differential treatment or impact.

While anti-discrimination law connects the unlawful distinction to a factor of discrimination qualified as such by the law according to a closed list of human traits, when discrimination arises from AI, it is very unlikely that it will hinge, *prima facie* at least, on legally recognized factors of discrimination.

²³ On this, see paragraph no. 4.b.

²⁴ C. O'Neil, [When Not to Trust the Algorithm](#), in *Harvard Business Review*, 2016.

Costanza Nardocci
Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

Should AI rely on legally protected grounds, its use would be, of course, prohibited, but, more often, AI relies, instead, on wide and heterogeneous sets of elements called proxies²⁵.

When framed in the context of AI, thus, the proxy performs as a factor of discrimination in purely human-driven discrimination even without being one, and ultimately, supports the distinction upheld by the machine.

Although akin in their discriminatory capacity, in that both have the potential to influence an unlawful conduct and to cause a seemingly discriminatory impact, the factor of discrimination and the proxy differ from one another on several levels.

The first difference is predominantly ontological in nature. Whereas the factors of discrimination identify human features, be they immutable (as for race or ethnicity) or temporarily mutable, meaning subject to modifications over time depending on facts and circumstances that involve the holder, the proxy seldom recalls traits that define human beings in their most inherent aspects. Conversely, the proxy may indicate preferences of all sort (food, movies, songs) or it may be very circumstantial as when AI makes its distinctions based on the possession of an object (a car, a smartphone), or on the presence, even if just random, in a specific geographical space such as a neighbourhood, a city, a State, a Country.

In not being descriptive of a human quality defined as suspect under the law, the proxy is not an indicator of a risk of discrimination per se. It could become one or perform like a factor of discrimination when it establishes a relationship of such a kind with a legally suspect ground that relying on the proxy would be as if the difference was based on a traditional ground of discrimination. These cases draw attention to the correlations the proxy develops with the traditional factors of discrimination, which brings us to the second major difference between the proxy and a legally suspect ground. In fact, if it becomes suspect just in case it operates in such a way to call into question a traditional factor of discrimination, contrary to a suspect ground, it means that the proxy possesses no inherent discriminatory capacity.

²⁵ The concept of proxy is not new to anti-discrimination law and, before that, to constitutional law. On this, see, extensively, L. Alexander and K. Cole, *Discrimination by Proxy*, in *University of Minnesota Law School*, vol. 14, 1997, p. 453 ff., and G. Karger, *The Proxy Problem in Disparate Treatment*, cit.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

Yet, the proxy may be suspect, as the literature argues, discussing the so-called “proxy discriminatory”²⁶ rightly to emphasize the crucial role of the proxy in discriminatory AI. What is, thus, dangerous about the proxy is the ability to predict individual affiliations to protected groups. Some examples could clarify the meaning and the implications of the proxy’s correlations with protected grounds.

When AI uses the Zip code as a classifier, there is nothing suspicious at first glance, because the Zip code does not qualify as a factor of discrimination, which makes its use perfectly legal. However, if its use disproportionately impacted individuals of low-income status, the proxy would become suspect, because it would predict individuals’ affiliation to a protected group, as a result of the correlation established with social classes, which, as is known, constitute a factor of discrimination.

Moreover, in the context of health insurance systems, especially in the U.S., scores are frequently utilized to determine eligibility for insurance coverage. This scoring mechanism may inadvertently favour individuals with better insurance or higher socioeconomic status. Consequently, an AI system for hospital resource allocation could perpetuate unequal access to healthcare, with socioeconomic conditions and insurance status serving as proxies. Another instance, then, involves diagnostic disparities, where proxies such as race, ethnicity, or gender play a role. For example, using symptoms that mainly manifest in males as a basis for diagnosis may lead to discrimination against women, as their pathology might not be accurately associated. Additionally, AI algorithms in diagnostic tools may exhibit variations in accuracy across different racial or ethnic groups, resulting in disparities in the identification and treatment of certain medical conditions.

In short, the proxy seldom, directly or indirectly, recalls human qualities, but rather it frequently establishes connections and more or less tight correlations with traits that are used to divide individuals and social groups along gender, racial, or ethnic lines. That is another way of saying that, by relying on the proxy, AI directly or indirectly discriminates depending on elements widely differentiated among one another and other

²⁶ On the concept of proxy discrimination, see, extensively, H. Weerts, A. Kelly-Lyth, R. Binns and J. Adams-Prassl, *Unlawful Proxy Discrimination: A Framework for Challenging Inherently Discriminatory Algorithms*, in *Association for Computing Machinery, Inc.*, 2024, p. 1850 ff.; M.C. Tschantz, *What is Proxy Discrimination?*, in *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*, Seoul, Republic of Korea, ACM, New York, 2022; A.E.R. Prince and D. Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, in *Iowa Law Review*, vol. 105, 2020, p. 1257 ff.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

than the traditional factors of discrimination laws and Courts target and scrutinize.

In light of the otherness of the proxy compared to the traditional grounds of discrimination, the literature has coined the expression “proxy discrimination”²⁷ to capture AI-based discrimination and to emphasize the role of the proxy in shaping the traits of the new form of discrimination.

Additionally, as much as AI may discriminate based on an element (the proxy) that correlates with one or more protected grounds, either directly or indirectly, proxy discrimination in AI could, thus, resemble a direct or an indirect form of discrimination depending on the role of the proxy in the process leading to a discriminatory outcome. As a consequence, when AI makes discriminatory choices based explicitly and directly on elements that are predictive of individual affiliation to a protected group, AI-based discrimination has been defined as “direct proxy discrimination”. On the other hand, when the rule behind the decision adopted by AI is coached in neutral terms, lacking any apparent correlations with protected grounds, but its effects happen to discriminate against a targeted group, the difference in treatment may be regarded as a form of “indirect proxy discrimination”.

Therefore, the link between the suspect grounds will define proxy discrimination as either direct or indirect, depending on the predictive capacity of the proxy, which could lead to one or the other type in light of its direct or indirect ability to establish a correlation with a legally protected feature.

Leaving the proxy aside, many challenges arise when confronting proxy discrimination.

The first one rests on the difficulty of capturing *ex-ante* which elements – human or of other nature – may act as proxies in the functioning of AI systems and, as a result, which are the potential or actual correlations are between the proxy and legally protected grounds of discrimination. In other words, with proxy discrimination, the programmer, the users, and, eventually, the “victim” might not be fully aware of the established correlations between the element AI uses to make distinctions and one or

²⁷ On this, see, extensively, A.E.R. Prince and D. Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, cit., p. 1277 ff.; A. Datta et al., *Proxy Discrimination in Data-Driven Systems*, cit.; J. Grimmelmann and D. Westreich, *Incomprehensible Discrimination*, in *California Law Review*, vol. 7, 2017, p. 164 ff.; L. Alexander and K. Cole, *Discrimination by Proxy*, cit., and, also, C. Nardocci, [*Proxy Discrimination in Artificial Intelligence: What We Know and What We Should Be Concerned About*](https://chairesante.ca/), available at *chairesante.ca*, 9 February 2024.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

more factors of discrimination and, therefore, of the risks of discrimination associated with the AI system at issue.

The unawareness of the existing correlations goes even further. In fact, that lack or partial knowledge is expected to have further impacts on the risk assessment of AI systems, as it could be extremely difficult to unveil when and how the proxy may or may not be predictive of individual affiliations to a protected group, that is to say, when AI discriminates against someone.

Another reason why proxy discrimination should be carefully handled has to do with the difficulties in predicting individuals' affiliation to a social group, which matters for the purpose of the identification of the victim.

Not only are proxies, in fact, hard to detect *per se*, but the same reasoning applies to individual affiliations, which, in the context of proxy discrimination, no longer ground exclusively on human features, but, rather, on the interplay among proxies and protected grounds.

That also means, that the more is difficult to establish when and how the proxy predicts individual affiliations to protected groups, the more unlikely will be to have a clear knowledge of whether the element AI uses to make its decisions (the proxy) will discriminate individuals belonging to legally protected groups²⁸ or to “other” groups, resulting instead from the said intersections between the proxy and a suspect ground of discrimination.

In other words, proxy discrimination may obscure the identification of the victim, who will eventually face barriers in accessing justice and, even before that, in knowing that to be the target of discriminatory AI.

Proxy discrimination has, in fact, brought to light other challenges for anti-discrimination law, which deals with the victim of the unlawful treatment. The peculiarity of the victim and his/her status in the context of proxy discrimination leads to a discussion of the new victim paradigm of discriminatory AI.

First, as long as the proxy could be unknown or difficult to identify in advance, the same is expected to happen to the individual and/or the social group exposed to the unlawful treatment, who could remain unknown or might not be entirely aware of having been targeted by discriminatory AI.

There is more, though.

²⁸ For an in-depth examination of the ways in which AI technologies discriminate against traditionally disadvantaged groups, see the contributions published in A. Quintavalla and J. Temperman (eds), *Artificial Intelligence and Human Rights*, Oxford, 2023.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

Besides the likelihood that the victim(s) do not realize having been discriminated against, proxy discrimination also contributes to the creation of new collective victims and, thus, of new mechanisms of individual affiliations, as a result of the indefinite possibility of existing correlations among proxies and protected features²⁹. Proxy discrimination, in fact, creates “new” or, at least, “other” factors of discrimination, since the proxy acts as a factor of discrimination in light of its unlawful effects *vis-à-vis* the victim(s); favours the emergence of “new” victims³⁰, as the proxy relies on a much broader range of human features or human-related qualities, that regroup individuals around other shared traits; lastly, establishes new mechanisms of individual affiliations and, therefore, “new” minority groups that result from the said correlations among the proxies and the legally protected grounds.

Beyond the discriminatory capacity, there is, thus, the ability of the proxy to create new social formations and, particularly, new minority groups³¹.

Beyond the discriminatory capacity, there is, thus, the ability of the proxy to create new social formations and, particularly, new minority groups.

In light of the definition of minority, according to which a minority group qualifies as such because of its subordinated position *vis-à-vis* a dominant group and in light of common features that are suspect under the

²⁹ In this regard, see K. De Vries, *Identity, profiling algorithms and a world of ambient intelligence*, in *Ethics and Information Technology*, vol. 12, 2010, p. 71 ff., at 76, who focuses in particular on a reconstruction of the concept of identity which – framed in the AI context – assumes decisive importance, allowing to distinguish «a device used to decide who is in and who is out; who is us and who is them; who is likely to be a good customer and who is not; who is allowed to pass the border and who is not». On the indefinite types of correlations among proxies and protected grounds see D. Boyd, K. Levy and A. Marwick, *The Networked Nature of Algorithmic Discrimination*, in *Open Technology Institute, New America, Data & Discrimination*, 2014, p. 53 ff., who argue that: «[t]he notion of a protected class remains a fundamental legal concept, but as individuals increasingly face technologically mediated discrimination based on their positions within networks, it may be incomplete. In the most visible examples of networked discrimination, it is easy to see inequities along the lines of race and class because these are often proxies for networked position. As a result, we see outcomes that disproportionately affect already marginalized people».

³⁰ Part of the literature converges on this as a consequence of the argued heterogeneity of AI-derived discrimination, including B. Mittelstadt, *From Individual to Group Privacy in Big Data Analytics*, in *Philosophy and Technology*, vol. 30, 2017, p. 475 ff.; and S. Wachter, *Affinity Profiling and Discrimination by Association in Online Behavioural Advertising*, cit.

³¹ For this thesis, see, also, C. Nardocci, *Minority Rights in the Era of Artificial Intelligence*, in *European Yearbook on Minority Issues*, vol. 22, 2025, p. 1 ff.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

law, the “new” social formations created by proxies may be called algorithmic or AI-targeted minorities. Algorithmic or AI-targeted minorities are, therefore, communities that share with the traditional minorities³² a subordinated status within the society, but that differ from them a first glance because of the common feature that is not protected, but a proxy correlated with one or more factors (s) of discrimination³³.

The proxy is the main element of distinction between AI-targeted minorities and traditional minorities, leading to further elements of differentiation that echo what the proxy looks like. As the proxy is temporarily mutable, so the social formations aggregated around it are dynamic in contrast to the predominant stability over time of traditional and historical minorities. AI-targeted minorities are, in fact, dynamic over time, because their existence is ephemeral, dependent on one element that is supposed to evolve, disappear, as in the case of preferences, possessions, and locations. In other words, AI-targeted minorities are random communities that happen to share something apparently neutral (the proxy), which instead turns out to be suspect in light of its correlations with a suspect ground of discrimination.

Now, the instability and the shared element, the proxy, are not enough to describe AI-targeted minorities and their fragile condition under the law.

While speaking about the proxy tackles the objective side, the traditional definition of minority groups adds something more. That is the subjective feeling, again shared by the members, of being part of a group that is sidelined by the dominant group and subjected to unlawful differential treatments. Conversely to traditional minorities, the feeling of solidarity among the members is lacking in the context of AI-targeted minorities. In being randomly aggregated and mutable over time in their

³² There is no universal definition of minority. Nonetheless, the most widely accepted definition was suggested by Francesco Capotorti, in *Study on the persons belonging to ethnic, religious and linguistic minorities*, in UN Subcommission on Prevention of Discrimination and Protection of Minorities. *Special Rapporteur to carry out a Study on the Rights of Persons belonging to Ethnic, Religious and Linguistic Minorities*, 1979, according to which a minority is «[a] group numerically inferior to the rest of the population of a State, in a non-dominant position, whose members – being nationals of the State – possess ethnic, religious or linguistic characteristics differing from those of the rest of the population and show, if only implicitly, a sense of solidarity, directed towards preserving their culture, traditions, religion or language».

³³ In favor of the identification of new social formations, that differ to the traditional minority groups, L. Taylor, L. Floridi and B. Van Der Sloot, *Group Privacy New Challenges of Data Technologies*, Cham, 2017.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

associations to one another, the members of AI-target minorities very rarely feel that they are unlawfully treated and share a sense of solidarity towards other members who are largely unknown. In short, the members of AI-targeted minorities are essentially unaware of their condition and of their affiliation to the group.

If being associated with the proxy to other individuals alike and becoming part of a minority group would not curtail the applicability in their respect of the individual rights recognized to those who belong to a minority group, the risks could be less serious and worrisome than they really are. Yet it is quite the opposite.

Reference is made to individual rights that counterbalance the collective rights entitled to traditional minority groups and, specifically, the right to self-identify as a member of the minority community³⁴, the right to dissent with the rules governing the internal affairs of the community, and ultimately, the right to exit the group³⁵. Rights that constitute a corollary of the constitutional principle of self-determination, which presupposes that the person has knowledge of his/her status in society and of his/her affiliation to a minority group.

While the individual is usually and fully aware of his/her affiliation to a minority group, the same conditions hardly apply to AI-targeted minorities, whose members rarely know about their association with a minority created by AI technologies. Evidently, the less the members will be aware of their status, the less they will be placed in a condition to exercise the rights associated with being part of a minority. No right to self-identification, since there is no knowledge of the affiliation to the community; no right to exit as there could hardly be a conflict between the member and group, since the individual does not know about his/her

³⁴ The reference is to Art. 3(1) of the Council of Europe Framework Convention for the Protection of National Minorities, according to which: «Every person belonging to a national minority shall have the right freely to choose to be treated or not to be treated as such and no disadvantage shall result from this choice or from the exercise of the rights which are connected to that choice».

³⁵ On the right to exit, see, extensively, L. Green, *Rights of exit*, in *Legal Theory*, vol. 4(2), 1998, p. 165 ff.; C. Kukathas, *Exit, freedom, and gender*, in D. Borchers and A. Vitikainen (eds), *On exit: Interdisciplinary perspectives on the right of exit in liberal multicultural societies*, Berlin, 2012, p. 34 ff.; W. Kymlicka, *The Rights of Minority Cultures*, Oxford, 1995, and Id., *Multicultural Citizenship: A Liberal Theory of Minority Rights*, Oxford, 1995; K.C. Murat Mertel, *Toward a Substantive Right of Exit*, Kingston, 2007; K. Henrard, *Devising an Adequate System of Minority Protection. Individual human rights, minority rights, and the right to self-determination*, Leiden, 2000; J. Packer, *On the Content of Minority Rights*, in J. Raikka (ed), *Do we need minority rights? Conceptual Issues*, Leiden, 1997, p. 121 ff.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

association to the minority community; eventually, no right to dissent with the internal rules of the minority, that, besides, it is very unlikely will have some, given its random and ephemeral existence, that should prevent the possibility for the newly formed community to benefit from a governing structure.

In light of the above, there are, thus, numerous reasons why the proxy should be placed at the centre of the investigation into discriminatory AI. Be it the element that grounds the unlawful distinction generated by AI, or that favours the establishment of unprecedented mechanisms of collective affiliation, creating “new” victims, the proxy should be adequately considered by legislators and domestic and supranational Courts in their attempts to regulate and tackle discriminatory AI.

5. Did Someone Say, “Discriminatory AI”? Regulating and Sanctioning AI-Based Discrimination

Amid the increasing significance of the discriminatory implications associated with AI technologies and the deviations from the traditional schemes of anti-discrimination law, regulators worldwide have until

Now, seldom adopted legislative responses to tackle AI-based discrimination.

It is, in fact, primarily discrimination as a phenomenon per se that seems not to attract the interest of domestic legislators and supranational organizations. Before arguing in favour of or against the said heterogeneity separating human-driven from AI-based one, references to discrimination rarely go beyond the invitation to ensure the compliance of AI technologies with the constitutional principle of equality and non-discrimination. Moreover, despite notable exceptions, anti-discrimination law is not called into question, neither in terms of its invoked applicability to tackle discrimination deriving from AI nor to suggest its partial revision to reconcile with the novelties AI-based discrimination brings about.

Let’s begin with Europe, where the first comprehensive regulation of AI technologies, the AI Act, very well documents this trend.

The 2021 first draft of the Regulation named, in fact, the word discrimination just twice in the entire text³⁶, and likewise, it seems not to be

³⁶ [The text is available at *digital-strategy.ec.europa.eu*](https://digital-strategy.ec.europa.eu/en). Among many others, discuss the first regulatory attempt C. Casonato and B. Marchetti, *Prime osservazioni sulla proposta di*

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

much inspired by a human-rights approach, endorsing instead a hierarchical vision of fundamental rights prioritizing privacy and data protection.

Contrary to the first draft, the 2024 adopted version, which came officially into force on August 1st, amended some of the loopholes of the previous proposal by introducing an explicit mention of the European Union Charter of fundamental rights³⁷ and expanding the obligations to safeguard all principles protected therein. While the choice speaks of a willingness to secure a more robust compliance of AI technologies with fundamental rights, the AI Act does not feature references to EU anti-discrimination law and to the 2000 Directives nos. 43³⁸ and 78³⁹.

In line with the never-abandoned preference to link AI regulation to data protection law, the only form of discrimination contemplated and sanctioned under the AI Act is profiling following the approach of the GDPR. Thus, Article 5(1)(d) prohibits the use of AI systems «for making risk assessments of natural persons to assess or predict the risk of a natural person committing a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics».

Regolamento dell'Unione europea in materia di Intelligenza Artificiale, in *BioLaw Journal*, vol. 3, 2021, p. 415 ff.; M. Kop, *EU Artificial Intelligence Act: The European Approach to AI*, in *Transatlantic Antitrust and IPR Developments*, 2021; W.G. Voss, *AI Act: The European Union's Proposed Framework Regulation for Artificial Intelligence Governance*, in *Journal of Internet Law*, vol. 25(4), 2021, p. 7 ff.; M. Veale and F.Z. Borgesius, *Demystifying the Draft EU Artificial Intelligence Act*, in *Computer Law Review International*, vol. 4, 2021, p. 97 ff.; L. Floridi, [The European Legislation on AI: A Brief Analysis of its Philosophical Approach](#), 2021, available at [ssrn.com](#).

³⁷ See, on this, especially, Recital 1 about the purpose of the regulation, that is described as «to improve the functioning of the internal market by laying down a uniform legal framework in particular for the development, the placing on the market, the putting into service and the use of artificial intelligence systems (AI systems) in the Union, in accordance with Union values, to promote the uptake of human centric and trustworthy artificial intelligence (AI) while ensuring a high level of protection of health, safety, fundamental rights as enshrined in the Charter of Fundamental Rights of the European Union (the 'Charter')», or, even, Recital 2, that explicitly requires the use of AI systems to be in compliance with «the values of the Union enshrined as in the Charter, facilitating the protection of natural persons, undertakings, democracy, the rule of law and environmental protection, while boosting innovation and employment and making the Union a leader in the uptake of trustworthy AI». Equally relevant is Recital 7, which bounds the regulation of high-risk AI systems to the respect of the fundamental principles safeguarded under the EU Charter of Fundamental Rights, and, lastly, Art. 1 that is the only provision of the AI Act, featuring an explicit reference to the Charter.

³⁸ Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin.

³⁹ Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

Moreover, Article 6(4), which sets the rules about high-risk AI systems, limits the derogations set forth under paragraph no. 3 in all cases of AI systems entailing profiling of natural persons⁴⁰.

The AI Act, thus, replicates the approach of the GDPR, which explicitly prohibits profiling under Article 22. However, it fails to acknowledge the difference in scope between the two regulations, which would have suggested embracing a wider interpretation of the notion of discrimination and sanctioning not only profiling, but also AI-based discrimination in the first place, regardless of its interpretation as a separate or overlapping form of human-driven discrimination.

Whereas, thus, profiling is recognized with a proper autonomous configuration, the AI Act mentions the right to non-discrimination a number of times in the Preamble, and only in two cases, both regarding the governance of high-risk AI technologies. First, under Article 10, which states that the training, validation, and testing data sets should be examined to exclude the recurrence of «possible biases that are likely to affect the health and safety of persons, harm fundamental rights or lead to discrimination prohibited under Union law [...]». And then, under Article 77, requiring national public authorities to ensure high-risk systems referred to under Annex III to comply with fundamental rights, «including the right to non-discrimination».

Besides the European Union, in 2024, Europe welcomed a second, rather important attempt to introduce some core principles regarding the relationship of artificial intelligence with human rights.

On May 17th, the Council of Europe adopted, in fact, the first global treaty on AI, the Framework Convention on artificial intelligence, human rights, democracy, and the rule of law⁴¹, open to signatures since September 2024.

⁴⁰ On this, it should be highlighted that the notion of profiling adopted by the Regulation fully overlaps with and follows that of the GDPR pursuant to Art. 4(4), according to which profiling means: «any form of automated processing of personal data consisting of the use of such personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements».

⁴¹ For some commentaries on the text, see, among others, the contributions by A. Hars, *Conceptual Difficulties in the Transformation of Human Rights to the Realm of Artificial Intelligence*, in *Acta Humana*, 2024, p. 123 ff.; F.P. Levantino and F. Paolucci, *Advancing The Protection Of Fundamental Rights Through AI Regulation: How The EU And The Council Of Europe Are Shaping The Future*, in P. Czech, L. Heschl, K. Lukas, M. Nowak and G. Oberleitner

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

Contrary to the AI Act, the Framework Convention is inspired by a more explicit and concrete human-rights-based approach to AI technologies.

Although it does not speak of anti-discrimination law, as applied or potentially applicable to AI-based discrimination, which goes beyond the competences of the Council of Europe, the Framework Convention dedicates two provisions to the principle of non-discrimination, and, before that, the Preamble also repeatedly affirms the urge to effectively tackle discrimination deriving from AI technologies.

Interestingly, the Framework Convention attributes two significances to the principle of equality and non-discrimination. Under Article 10, the Framework Convention includes the principle of equality and non-discrimination among the principles that should be safeguarded in all activities during the lifecycle of AI systems. According to Article 10, the Framework Convention states that all the activities undertaken by the State parties must comply with the principle of equality and non-discrimination «as provided under applicable international and domestic law». In making an explicit reference to international and domestic anti-discrimination law, Article 10 marks a profound difference compared to the EU Regulation. Therefore, while the AI Act chose not to include anti-discrimination law under the spectrum of relevant EU law for the application and implementation of the Regulation, the Framework Convention is rooted in the idea that the right not to be discriminated should be enforced and not merely formally invoked. Should the EU have adopted a similar approach, this would have entailed the explicit reference to the already mentioned 2000 EU anti-discrimination Directives, which, conversely, do not appear in the text.

Such a diversity in the interpretation and role of the right to non-discrimination between the AI Act and the Framework Convention turns out to be even clearer when looking at the second legal provision that the treaty of the Council of Europe introduced to make the obligation to respect the right at stake even more binding.

(eds), *European Yearbook on Human Rights 2024*, Leiden, 2025, p. 3 ff. For a commentary on the versions preceding the one ultimately adopted, see C. Nardocci, *La (seconda) svolta del 2024. Anche il Consiglio d'Europa decide di regolamentare l'intelligenza artificiale*, in *BioLaw Journal*, vol. 1s, 2024, p. 73 ff. Furthermore, on the process that led to the adoption of the Framework Convention, see E.H. Morawska, *Council of Europe Standards and Activities Related to AI: Towards a Framework Convention on AI and Human Rights?*, in M. Balcerzak and J. Kapelańska-Pręgowska (eds), *Artificial Intelligence and International Human Rights Law*, Cheltenham, 2024, p. 25 ff.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

Thus, Article 17 strongly reaffirms that the implementation of the Framework Convention should be secured without discrimination on any grounds, connecting the principle to the effective realization of the Treaty by the State parties.

What is missing in the Framework Convention, but it came through with no surprise, is the absence of any allusion to the peculiarities of AI-based discrimination. The Framework Convention seems to be moving, instead, on a different level, one that is more concerned with the definition of sets of core principles that reflect the values of the Council of Europe and that the Council of Europe aims to see fully safeguarded before the increasing resort to AI technologies.

Eventually, the AI Act's (mis)understanding of the relationships between discrimination and AI is not reflected in the context of the Framework Convention that, in short, "did what it could with what it is, and what it had", a supranational organization with no competence over State parties.

Beyond Europe, a more satisfactory proposal to respond to the challenges brought up by AI-based discrimination comes from the law adopted by the State of Colorado, the first of its kind regulation worldwide entirely dedicated to "algorithmic discrimination", so-called CAIA⁴².

The ground-breaking law, which is expected to come into force on February 1st, 2026, intends as its primary goal to introduce safeguards to protect individuals and social groups from algorithmic discrimination.

There are many reasons why the CAIA should be considered as a promising example of a legislative response to the challenges and risks associated with AI technologies.

First, the law introduces a definition of algorithmic discrimination under Part 17, dedicated to artificial intelligence.

According to the CAIA, «algorithmic discrimination means any condition in which the use of artificial intelligence system results in an unlawful differential treatment or impact that disfavors an individual or group of individuals based on their actual or perceived age, color, disability, ethnicity, genetic information, limited proficiency in the English language, national origin, race, religion, reproductive health, sex, veteran status, or other classifications protected under the laws of this State or federal law».

Without arguing in favour or against its heterogeneity vis-à-vis human-driven, the CAIA's definition of algorithmic discrimination is worth

⁴² Reference is to the [*Colorado Act Concerning Consumer Protections in Interactions with Artificial Intelligence Systems*](#), SB 24/205. The full text is available at content.leg.colorado.gov.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

considering first, because it emphasizes that it recurs every time an AI system is involved in the discriminatory conduct that eventually leads to the unlawful differential treatment or impact. While the statement could appear to be obvious in its content in that AI systems cannot but be directly or indirectly involved, it, nonetheless, focuses on one of the elements that separates AI-based from human-driven ones, meaning the said mixed agency between the human and the machine.

A second aspect to highlight lies in the reference to the individual and collective dimensions of the discriminatory effects. Without being a specific trait of AI-based discrimination, it, nonetheless, has the merit to recall that discrimination is a collective phenomenon impacting on individual and collective basis, contrary to the preference given by domestic laws to interpret discrimination as a purely individual, unreasonable, and, therefore, unlawful difference in treatment.

Lastly, the wide list of suspect grounds, made of a variety of factors of discrimination together with the relevance attributed to perceived and not only actual protected features, contributes to increasing the chances of intercepting and sanctioning AI-based discriminations.

Alongside the definition, the CAIA identifies a list of conduct that falls outside the scope of the legislation and that cannot be considered as manifestations of algorithmic discrimination.

Besides the significance of the first legislative definition of algorithmic discrimination, the second reason why the CAIA appears to be particularly meritorious is the precise and detailed enumeration of the obligations of the developer and the deployer to avoid algorithmic discrimination. The choice is highly commendable, as it supports a clear establishment of the responsibilities of the two main actors in the phases that precede putting AI systems on the market of AI systems. Moreover, it represents the first legislative attempt to identify strategies to cope with the specifics of algorithmic discrimination, which, in so doing, gains the uniqueness that is instead predominantly missing on a regulatory, but also theoretical, basis.

The CAIA does not mention the proxy – that the above analysis qualifies as a key trait in the context of discrimination generated by AI systems –, nor does it emphasize its role in the description of algorithmic discrimination. Yet, the preference for an open list of suspect grounds and the equalization between actual and perceived protected features seems to be moving in a coherent direction, with the inclusion or overlaps between proxy discrimination and CAIA's definition of algorithmic discrimination.

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

Yet, the preference for an open list of suspect grounds and the equation between actual and perceived protected features seems to be coherent with the notion of proxy adopted in the context of the so-called discrimination. In fact, proxies may very well identify with some of the listed characteristics that do not overlap with the traditional factors of discrimination despite their potential discriminatory capacity.

The comparative analysis does not offer anything else comparable to the CAIA, with the only exceptions of other US States that followed Colorado's pioneering lead⁴³, starting with the newly approved California law⁴⁴, which prohibits the use of AI that discriminates against applicants and employees on protected features under the Fair Employment and Housing Act (FEHA).

The comparative analysis does not offer anything else comparable to the CAIA, with the only exceptions of other US States that chose to follow Colorado's pioneering lead.

Similarly, besides the Council of Europe, the other international organizations have preferred soft law mechanisms to address the challenges brought up by artificial intelligence, including some references to discriminatory AI.

Thus, the 2024 UN Resolution “Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development”⁴⁵, encourages the member States to promote safe, secure and trustworthy AI systems to «help protecting individuals from all forms of discrimination, bias, misuse or other harm, and avoid reinforcing or perpetrating discriminatory or biased application and outcome through the life cycle of artificial intelligence systems». The UN Resolution also has the merit of recalling some of the strategies to contrast discriminatory outcomes, such as the analysis and mitigation of bias encoded in datasets, and to adopt the notion of “algorithmic discrimination”, suggesting –

⁴³ On this, see: the State of Illinois with Bill HB 3773, approved in June 2024, that tackles algorithmic discrimination in the workplace and Bill SB 2203, “Preventing Algorithmic Discrimination Act”, introduced in February 2025; the State of Connecticut with Bill SB2, introduced on 2 April 2024 and finally approved in January 2025, titled “An Act Concerning Artificial Intelligence”; the State of Virginia with Bill H.R. n. 2094 of 2025; the State of Texas with the proposed Bill HB n. 1709, introduced on December the 23rd, 2024; the State of New Jersey with its “Guidance on Algorithmic Discrimination” and the New Jersey Law Against Discrimination of January 2025, that extended the State’s anti-discrimination law to cover discrimination generated by AI technologies.

⁴⁴ Reference is to Bill AB no. 1018 of 2025.

⁴⁵ [The news and the contents of the Resolution can be consulted at *digitallibrary.un.org*.](http://digitallibrary.un.org)

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

perhaps just implicitly – that it represents something slightly different from the traditional forms of discrimination acted by humans only.

However, the United Nations did not adopt a binding treaty⁴⁶, like the Council of Europe, and other regional international organizations, such as the Organization of the American States (OAS)⁴⁷, the ASEAN, the Association of Southeast Asian States⁴⁸, and the African Union⁴⁹, shared a similar approach, although recent times saw an increase of initiatives that could lead, potentially, even to more concrete responses in the context of discriminatory AI.

⁴⁶ Suggests the opportunity that even the United Nations should adopt a treaty on AI M. Falstein, [Algorithmic discrimination and the existing international law of equality and non-discrimination](#), available at [austlii.edu.au](#).

⁴⁷ The Organization of the American States (OAS) recently adopted some noteworthy soft law documents that do not insist, however, on the discriminatory implications of AI. Reference is to the declaration “Towards The Safe, Secure, and Trustworthy Development and Deployment of Artificial Intelligence in the Americas: the Importance of Governance, Regulatory, and Institutional Frameworks” and to the following Action Plan, adopted as a result of the meeting held on December 12th and 13th, 2024 between the Ministers and other authorities responsible for science and technology of the OAS member States.

⁴⁸ Reference is to the regional guide [AI governance and ethics](#), whose full text can be consulted at [asean.org](#). The text does not focus on the specifics of AI-based discrimination, although it emphasizes the importance to adopt strategies to cope with discrimination arising as a result of the massive resort to AI technologies. One of the suggested safeguards should be to «include human interventions and checks on the algorithms and its outputs» and to ensure that «the design, development, and deployment of AI systems align with fairness and equity principles», alongside measures to «mitigate potential biases during data collection and pre-processing, training, and inference». Interestingly, the focus is on the same phases the literature argues are at risk to generate discriminatory and AI-related effects.

⁴⁹ The contribution of the African Union is expressed by the [AI Continental Strategy](#), which constitutes a very detailed document whose overall purpose is to advance the development of AI technologies in the African continent. The full text can be read at [au.int](#). Like for the ASEAN, the *AI Continental Strategy* does not endorse the view that AI discriminates in its own ways, so to speak. Nonetheless, discrimination associated to AI technologies is central in the structure and *rationale* of the document. It is first described as both a system-level and structural risk associated to AI and, then, it is linked to the lack of diversity featuring datasets. Diversity and inclusion represent a key goal of the *AI Continental Strategy*, in that separating the text from the other regional international human rights soft law documents adopted by the other regional organizations. See, in particular, p. 28 of the *AI Continental Strategy*, cit.

Costanza Nardocci
Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

6. *The Law Does not Say, But: Procedural Remedies at the Times of AI-Based Discrimination*

The acknowledgment of the specifics of AI-based discrimination does not end with the description of its lines of departure from the traditional understanding of discrimination as a purely human-driven phenomenon.

The argued inadequacy of anti-discrimination law brings the discussion forward to the identification, if any, of alternative or, to the very least, complementary mechanisms to ensure the enforcement of the principles of equality and non-discrimination.

The reasons are evident and numerous.

Among these, reference is chiefly made to the liability discourse, where issues largely arise from the difficulty of allocating responsibilities, particularly in cases involving highly sophisticated and increasingly autonomous AI systems.

The opacity and insufficient transparency of many AI systems frequently obstruct the detailed and comprehensive reconstruction of the chain of actions leading to a potential human rights violation, thereby preventing effective *ex post* evaluations and the identification of the person(s) accountable.

Furthermore, as each action becomes increasingly opaque or blurred in its content and boundaries as a consequence of the (frequently) unknown and autonomous functioning of the machine, attribution to human agency progressively diminishes if not vanishes in the worst-case scenario.

With this regard, the existing regulatory framework doesn't help as it does not respond to the challenges surrounding the attribution of liability in the context of AI-based discrimination as of other human rights violations caused by AI. Until now, especially in the European scenario, preference has been given to the unfold of impact assessment strategies rather than to develop mechanisms to fill the existing gaps in the field of the allocation of responsibilities among the multiple actors interacting with the machine⁵⁰.

⁵⁰ Critically, on this, B. Botero Arcila, *AI liability in Europe: How does it complement risk regulation and deal with the problem of human oversight?*, in *Computer Law & Security Review*, vol. 54, 2024, 106012, p. 1 ff.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

A possible way out to circumvent the lack of normative instructions could be found first in the resort to some of the “new” rights, that legislators have been progressively associated to AI⁵¹.

First and foremost, the right to an explanation⁵², a corollary of the more well-known principle of transparency, could very well serve the purpose.

When it comes to discriminatory AI (but not only), a criterion to establish liability could rest on the knowledge, or lack of, the functioning of the machine. That is to say that the deployer or the end user or both could be held accountable of an AI-based discrimination every time they show no knowledge of how the machine works. Put differently, it is the non-compliance with the right to know, that could play out as a violation of the principles of equality and non-discrimination.

The reliance on the right to an explanation could, therefore, avoid the necessity to undergo an unlikely successful reverse engineering process for the identification of those responsible for an AI-based discrimination⁵³.

Ultimately, the so-called right to an explanation may mitigate certain obstacles to the detection of bias and, above all, to the reconstruction of the causal link between the relevant conduct and the resulting discriminatory effect. In doing so, it may contribute to addressing the current lacuna in the identification and verification of the existence of liability – both individual

⁵¹ Reference is chiefly made to the EU AI Act and to the CoE’s Framework Convention on artificial intelligence, human rights, democracy and the rule of law.

⁵² On the right to an explanation see S. Wachter, B. Mittelstadt and L. Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, in *International Data Privacy Law*, vol. 7(2), 2017, p. 1 ff., who argue that: «[t]wo kinds of explanations may be in question, depending on whether one refers to: system functionality, that is, the logic, significance, envisaged consequences and general functionality of an automated decision-making system, e.g. the system’s requirements specification, decision trees, pre-defined models, criteria, and classification structures; or to specific decisions, that is, the rationale, reasons, and individual circumstances of a specific automated decision, e.g. the weighting of features, machine-defined case-specific decision rules, information about reference or profile groups. Furthermore, one can also distinguish between explanations in terms of their timing in relation to the decision-making process: an ex ante explanation occurs prior to an automated decision-making taking place. Note that an ex ante explanation can logically address only system functionality, as the rationale of a specific decision cannot be known before the decision is made; an ex post explanation occurs after an automated decision has taken place. Note that an ex post explanation can address both system functionality and the rationale of a specific decision».

⁵³ For an endorsement of such instrumental use of the right to an explanation, see UK Court of Appeal, *Ed Bridges v. The Chief Constable of South Wales Police and others*, 11 August 2020.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

and collective – stemming from the deployment of AI systems, by ensuring the intelligibility, transparency, and explainability of their respective decision-making mechanisms.

Another example of a possibly fruitful applicability of AI-associated rights to tackle AI-based discrimination deals with the widely invoked principle of human oversight.

The underlying *rationale* largely mirrors the arguments put forward regarding the right to an explanation. Therefore, the proof of the absence of a human supervision over the functioning of the machine could be invoked to ground the accountability of those who were subjected to the positive obligation to control the AI system at stake. That would also mean, that no machine could replace a human being and, likewise, no machine could be held accountable for a human rights violation⁵⁴.

In both cases, the right to an explanation and to human oversight may serve to partially obviate to, or attempt to, the difficulties facing the identification of the liabilities resulting from the lack of human control over the AI systems.

Besides the applicability of AI-related rights, a second question arising in the context of discriminatory AI grapples with the role of anti-discrimination law. The sustained inadequacy of the existing normative provisions does not, in fact, exclude entirely their applicability to discriminatory AI.

While the distinction between direct and indirect discrimination, particularly, cannot be transposed when discussing discriminatory AI, the procedural remedies provided under anti-discrimination law reveal a different and potentially transferable dimension, that could increase the victim's right to access to justice.

On this, the rule regarding the reversed burden of the proof on the side of the respondent could certainly be preserved. Given the imbalance in terms of knowledge and access to data and information between the victim and all the actors involved in the design, functioning, deployment and use of the AI system, the rule that place on the respondent the obligation to prove his/her lack of culpability maintain its centrality even to tackle discriminatory AI.

A second provision likely to produce positive effects is the recognition of the standing of NGOs and other public entities to initiate

⁵⁴ On the principle of human oversight in its relationships with the liability discourse see the groundbreaking judgement of the Constitutional Court of Colombia, No. n. T-323 of 2024.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

judicial proceedings on behalf of the victim introduced in the European Union with the so-called 2000 anti-discrimination Directives. The provision may be usefully interpreted in conjunction with Article 77 of the *AI Act*, which delineates the powers of national public authorities and bodies tasked with the protection of fundamental rights, including the right to non-discrimination, in relation to high-risk AI systems.

Although narrow in scope – applying only to high-risk systems – the provision is noteworthy in that it grants national authorities access to relevant information on the functioning and datasets of such systems, potentially providing additional evidentiary support to the victim’s claims before the Courts.

Eventually, given the complexity of AI systems, the effective realization of the principles of equality and non-discrimination will require an integrated approach and a supportive role of computer scientists to unravel and deconstruct the phases of the functioning of the machine intended to identify each action and to attribute each of them to a human agent. With this regard, the right to an explanation and to human oversight will perform a supplementary role every time the identification of the person “behind” the machine, and potentially accountable, will be prevented by the opacity of the AI system. At this stage, though, is re-centering AI discourse on liability and, with it, human agency (and human accountability).

7. Conclusions: A Call for Awareness

Many are the reasons to discuss discriminatory AI.

Some rest on the data, which show the discriminatory potentials of AI technologies and call for legislative interventions to tackle them, relying on existing norms. Others stem from the argued heterogeneity of AI-based discrimination compared to the so-called human-driven one and advocate for alternative readings of anti-discrimination law to meet the novelties brought up by discrimination caused by AI.

One way or the other, discourses about the discriminatory implications of AI technologies have not yet conquered the scene. Nor does it in Europe, where the EU AI Act rests on a silent preference for a revisited privacy-based approach that continues to prioritize the protection of data over other fundamental rights, chiefly and among others, the principle of non-discrimination. Nor does it elsewhere, where the adopted or *in itinere* regulations seldom address the specifics of the interplay between AI and

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

discrimination, despite notable exceptions, as in the cases of the laws adopted by Colorado, Illinois, and California in the United States.

The scarcity of legislative attention towards discriminatory AI goes in parallel with the poor case-law on the domestic and supranational level.

Until now, Courts have rarely sanctioned the discriminatory attitudes of AI technologies and, when they did, they have predominantly relied on the traditional categorizations of anti-discrimination law, stretched to adapt to the specifics and not always overlapping features of AI-based discrimination. The well-known case against the food delivery company *Deliveroo*, held by the Tribunal of Bologna (Italy)⁵⁵, is emblematic of this trend with its resort to the concept of indirect discrimination to decide a case of instead direct proxy discrimination. A similar approach has also been shared by the Australian Federal Court in *Tickle v. Giggle*⁵⁶, about the discrimination of a transgender woman by a women-only social media app, which, relying on a biometric system of identification, has not recognized the applicant. Here, the Court opted for an integral deference to existing norms of anti-discrimination law, qualifying the conduct as indirect discrimination, although the choices made by the AI system were not neutral at all, conversely hinging on sex. Rather, the issue was that the AI system was not built to recognize the third gender, adopting instead a binary reading of gender differences. Despite turning a blind eye to the specifics of the AI-based discrimination at stake, the judgment is, nevertheless, worth considering as the Australian Court concluded for the violation of the principle of equality and non-discrimination in a rare case of discriminatory AI successfully brought before the judiciary. Lastly, in a case involving the use of AI to establish the probabilities of recidivism of an indigenous inmate, the Canadian Supreme Court⁵⁷ refrained from engaging with the specific challenges posed by AI-based discrimination and instead confined its analysis to established anti-discrimination law.

Elsewhere, and vice versa, Courts have sometimes circumvented the discriminatory dimension of the applicants' submissions, insisting on other rights associated with AI. One example is *Ed Bridges v. The Chief Constable of*

⁵⁵ Court of Bologna, *Deliveroo Italy srl v. Nidil Cgil Bologna, Filg Cgil Bologna, Filcams Cgil Bologna*, January 12th, 2023. The full text in Italian of the decision is available at lpo.it.

⁵⁶ Federal Court of Australia, *Tickle v. Giggle*, August 23rd, 2024. The full text of the judgment is available at judgments.fedcourt.gov.au. A comment of the judgment is offered by A. Kerr, *Tickle v. Giggle – No Laughing Matter*, in *Denning Law Journal*, 2024, p. 331 ff.

⁵⁷ Supreme Court of Canada, *Ewert v. Canada*, no. 37233, 13 June 2018.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

South Wales Police, decided by the UK Court of Appeal of Wales⁵⁸, where preference has been given to the right to know, elevated as a criterion to ground the liability of the individual, over the investigation of the discriminatory capacity of the AI system of biometric identification used by the police. While the lack of knowledge of the functioning of the AI system may very well play out as a criterion to ground the liability of the agent, it is unfortunate that the Court did not investigate whether the applicant suffered from a violation of the principle of equality and non-discrimination as a result of the wrongful identification⁵⁹.

Moving further, if no judge has explicitly spoken of proxy discrimination in the context of AI yet, two judgements took an alternative route, passing through the alternative between the interpretation of AI-based discrimination as a mere human-driven discrimination, either direct or, more often, indirect, and the denial of the existence of discriminatory implications in the given cases

The first one is a Finnish case decided by the *National Non-discrimination and Equality Tribunal*⁶⁰, which sanctioned an AI system used by a credit institution company to assess creditworthiness, which proved to be discriminatory based on multiple variables echoing traditional factors of discrimination⁶¹. The judgment is noteworthy, especially in that the *National*

⁵⁸ *R (Bridges) – v – CC South Wales & others*, August 11th, 2020. The full text of the judgment is available at judiciary.uk.

⁵⁹ It aligns with *Ed Bridges v. The Chief Constable of South Wales Police* a case decided by the Hague District Court, case no. C/09/550982 /HAZA 18-388, February 5th, 2020, about “SyRI” (*Risk Indication System*), an AI technology used by the Dutch Government to identify various forms of fraud, and considered by the Court contrary to Art. 8 ECHR

⁶⁰ National Non-discrimination and Equality Tribunal of Finland, Finlandia, no. 216/2017, March 21st, 2018. Like in *Ed Bridges*, the Court did not ascertain the discriminatory implications of the AI system at stake, but it did, however, relied on transparency, and of the lack of, as a criteria to ground liability for the discriminatory effects caused by AI. See the following extract, where the Court held that: «[t]he importance of transparency, in the interest of verifiability, is [...] compelling, because using the risk model and the analysis that is carried out in that context carries the risk that discriminatory effects – unintentional or otherwise – occur. The Advisory Division stated in its opinion [...] that analysing large data sets, with or without deep learning/self-learning systems is undeniably useful, but may also yield undesirable results, including unjustified exclusion or discrimination», § 6.91.

⁶¹ Interestingly, the Tribunal focused on the mechanism behind the discrimination suffered by the victim. The extract that follows is particularly relevant, in that the Tribunal recognized that: «if an individual score is made up of statistical variables, the score in question is not an individual assessment based on the income and financial status of the

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

Non-discrimination and Equality Tribunal emphasized the differences existing between discrimination caused by humans entirely and that generated also by AI technologies.

The second judgement that even implicitly highlighted the peculiarities of discriminatory AI is *Mobley v. Workday Inc.*⁶². The case concerned an AI hiring tool (Workday) accused of discriminating based on race and ethnicity. The District Court of Northern California ascertained the recurrence of discrimination caused by AI, for the first time, rightly interpreting the relationship between the proxy and a traditional factor of discrimination. While it did not mention the proxy, the District Court found, however, that the discrimination caused by Workday was nevertheless based on other features directly correlated with race and ethnicity in that their use was predictive of the belonging of applicants to the said protected groups. Thus, in the Court's view, although Workday did not know the race and ethnicity of the applicants, it did nonetheless have access to other information from which it could have easily inferred the racial and ethnic background of the applicants, as well as their age, discriminating against the applicants based on protected grounds precisely as in a typical case of direct proxy discrimination.

Mobley v. Workday remains, for now, pretty much isolated⁶³, but significant, since, for the first time, a Court sanctioned as discriminatory a difference in treatment based on an element that, without being a factor of discrimination, was correlated with it and considered suspect like any other protected feature.

Despite the welcomed reading of discriminatory AI, on the side of the liability of the agent(s), the judgment revealed a tendency to emphasize the role of AI systems in causing the discriminatory effects, diminishing human responsibility. That interpretation of the criteria for determining whether an individual or an entity should be held liable for violations of the principles of equality and non-discrimination reflects another important

person in question, but a case of statistical profiling mainly based on reasons related to grounds of discrimination. The customer is not treated as an individual, but as a representative of statistical profiling based mainly on variables related to grounds of discrimination, which the creditor applies to all persons who fit the profile, such as men living in a certain residential area, having a certain first language and being of a certain age».

⁶² *Derek Mobley, Plaintiff, v. Workday, Inc.*, no. 23-cv-00770-RFL, July 12th, 2024.

⁶³ Additional noteworthy disparate impact claims about discriminatory AI in the United States context include: Fair Housing Act ("FHA"). *Huskey, et al. v. State Farm Fire & Casualty Co.*, 2023 WL 5848164 (N.D. Ill. Sept. 11, 2023); *Louis, et al. v. SafeRent Solutions, LLC and Metropolitan Management Group LLC*, 685 F.S upp. 3d 19 (D. Mass. 2023); *Open Communities v. Harbor Group Management Co., LLC, et al.*, Case No. 23-CV-14070 (N.D. Ill.).

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

trend that merits closer examination, to prevent discussions on liability for AI-based discrimination from being marginalized or left at the periphery of legislative initiatives. On this, the Court clearly stated that: «Workday's role in the hiring process is no less significant because it allegedly happens through artificial intelligence rather than a live human being who is sitting in an office going through resumes manually to decide which to reject» and, even more significantly, that «[d]rawing an artificial distinction between software decisionmakers and human decisionmakers would potentially gut anti-discrimination laws in the modern era».

Besides the regulatory (non)responses and the non-resolving case-law, the investigation has attempted to demonstrate the departure of AI-based discrimination from human-driven one, relying on the categories of EU anti-discrimination law. It has shown the distance from the concept of direct discrimination, the limited enforceability of indirect discrimination when AI is the primary cause of discrimination, hinging, instead, on a different construction of the relationships between AI and inequality that rests on the proxy with its indefinite and unpredictable discriminatory capacity. One way to address this could be through the concept of discrimination by association, as developed in the case law of the Court of Justice of the European Union⁶⁴, which – more than any other recognised form of discrimination – closely mirrors some of the core features of proxy discrimination. Discrimination by association could, therefore, bridge anti-discrimination law and the facets of discriminatory AI discussed here, supporting the development of complementary normative strategies.

Insisting on the necessity to re-interpret anti-discrimination law in light of the specifics of AI-based discrimination, the hope is that equality and non-discrimination will regain their centrality in the context of AI, and that legislators will eventually promote the much-awaited revision of the existing provisions of anti-discrimination law to ensure that discriminatory AI will eventually be tackled for what it is.

⁶⁴ Among the first judgements, see EU Court of Justice, [C-83/14], *CHEZ Razpredelenie Bulgaria AD v. Komisia za zashtita ot diskriminatsia*, 16 July 2015; EU Court of Justice, [C-303/06], *Coleman v. Attridge Law and Steve Law* EU, 17 July 2008.

Costanza Nardocci

Discrimination Revised.

How AI Is Reshaping Anti-Discrimination Law

ABSTRACT: The Article intends to illustrate the core elements featuring AI-based discrimination. It untangles the relationships AI-based discrimination entertains with EU anti-discrimination law, looking at how direct and indirect discrimination react towards discrimination generated by AI technologies. It then proceeds to argue that AI-based discrimination could not be misconceived as a pure manifestation of human deliberate or unintentional willingness to discriminate, but, rather, as a separate form of unreasonable treatment, where humans and automatic agencies intersect with one another.

Through the comparison between human-driven and AI-based discrimination, the Article eventually questions the adequacy of existing anti-discrimination laws to tackle discriminatory AI and advocates for the introduction of new sets of mechanisms to counter the prejudicial effects deriving from discriminatory AI.

KEYWORDS: artificial intelligence – discrimination – proxy – victims (new ones) – anti-discrimination law.

Costanza Nardocci – Associate Professor of Constitutional Law,
State University of Milan - “La Statale”, Milan, Italy
(costanza.nardocci@unimi.it)