

LINCOLN MEMORIAL UNIVERSITY LAW REVIEW

VOLUME 9

SUMMER 2022

ISSUE 3

A WAY OUT FOR EUROPE: HOW CAN EUROPE COMBAT DISCRIMINATION BY AUTOMATED DECISION-MAKING SYSTEMS?

Elvin Evrim Dalkılıç¹

I. INTRODUCTION

Humans are allowing more and more decisions to be made by machines, but at what cost?² In other words, as

¹ Associate Prof. Dr. Elvin Evrim Dalkılıç, PhD, Bilkent University Faculty of Law, elvin@bilkent.edu.tr, ORCID: 0000-0003-2873-2476. After graduating from Ankara University Faculty of Law in 2000, she completed her PhD at Gazi University in 2009. Elvin Evrim Dalkılıç was awarded the title of associate professor in 2019. She started her academic career as a research assistant at Bilkent University Faculty of Law, Department of Administrative Law in 2003 and has been continuing her academic career as a full-time faculty member at the same faculty since 2010. Dr. Elvin Evrim Dalkılıç teaches administrative law, administrative judiciary, and international human rights law at Bilkent University Faculty of Law.

² See Symposium, *Rise of the Machines: Artificial Intelligence, Robotics, and the Reprogramming of Law: Power, Process, and Automated Decision-making*, 88 FORDHAM L. REV. 613, 613–14 (2019) (“Automated decision-making algorithms evaluate teachers, approve or reject loan applications, choose whom to search in an airport security line,

humans, do we overtrust technology? Could it be because humans are evolved as social animals where life originally did not lead by technology? Today, automated decision-making (“ADM”) systems based on complex algorithms are increasingly used to make decisions impacting people’s rights and interests.³ These systems undoubtedly have some advantages: they are more cost-effective, faster, and better at processing large datasets than their human counterparts.⁴ However, the widespread use of ADM systems also raises human rights concerns, namely that they may implement discriminatory practices behind an inscrutable veneer of technology.⁵ Humans may inadvertently introduce bias into ADM systems, allowing prejudicial decisions to become standardized. Moreover, ADM systems may reinforce and perpetuate existing biases because, unlike humans, ADM systems cannot consciously attempt to counteract learned biases.⁶

allocate police officers on the beat, and determine eligibility for government benefits, among a litany of other commercial and government decisions.”).

³ “The use of the term ‘automated decision-making’ has become common in scientific discussion and (legal) practice. The term addresses both the use of algorithms for decision-making support of human decision-makers and the automated execution of decisions, although these are not always clearly differentiated from each other. For both types, the terms ‘automated decision-making systems’ (ADM systems) or ‘automated decision systems’ are also used.” CARSTEN ORWAT, FED. ANTI-DISCRIMINATION AGENCY, RISKS OF DISCRIMINATION THROUGH THE USE OF ALGORITHMS 21 (2019), available at https://www.antidiskriminierungsstelle.de/EN/homepage/_documents/download_diskr_risiken_verwendung_von_algorithmen.pdf?__blob=publicationFile&v=1.

⁴ See, e.g., Ifeoma Ajunwa, *An Auditing Imperative for Automated Hiring Systems*, 34 HARV. J. L. & TECH. 621, 632 (2021).

⁵ See Symposium, *supra* note 2, at 614 (“[T]he faith we tend to put in the power of technology shields algorithmic systems from critical interrogation, in general.”).

⁶ COUNCIL OF EUR., STUDY ON THE HUMAN RIGHTS DIMENSIONS OF AUTOMATED DATA PROCESSING TECHNIQUES AND POSSIBLE REGULATORY IMPLICATIONS 27 (2018), available at

Given the gravity of these issues, it is important to review the way Europe—particularly the Council of Europe (“CoE”) and the European Court of Human Rights (“ECtHR”)—has responded to the proliferation of ADM systems. After an expert study on the human rights and regulatory implications of automated data processing techniques, the CoE’s Committee of Ministers adopted a declaration in 2019 “on the manipulative capabilities of algorithmic processes.”⁷ The declaration decries the “growing threat” to human decision-making posed by automated systems and reads, “The effects of the targeted use of constantly expanding volumes of aggregated data on the exercise of human rights in a broader sense, significantly beyond the current notions of personal data protection and privacy, remain understudied and require serious consideration.”⁸ In 2020, the CoE’s Committee of Ministers issued a recommendation “to member States on the human rights impacts of algorithmic systems.”⁹ The recommendation stresses “the need to ensure

<https://edoc.coe.int/en/internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html>.

⁷ COUNCIL OF EUR., DECLARATION BY THE COMMITTEE OF MINISTERS ON THE MANIPULATIVE CAPABILITIES OF ALGORITHMIC PROCESSES (2019), *available*

at https://search.coe.int/cm/pages/result_details.aspx?ObjectId=090000168092dd4b.

⁸ *Id.* at ¶¶ 7, 9.

⁹ COUNCIL OF EUR., RECOMMENDATION CM/REC (2020) 1 OF THE COMMITTEE OF MINISTERS TO MEMBER STATES ON THE HUMAN RIGHTS IMPACTS OF ALGORITHMIC SYSTEMS (2020), *available at* https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154.

that racial, gender and other societal and labor force imbalances that have not yet been eliminated from our societies are not deliberately or accidentally perpetuated through algorithmic systems, as well as the desirability of addressing these imbalances through using appropriate technologies.”¹⁰ The European Convention has imposed similar guidelines on Human Rights (“ECHR”).¹¹

This paper consists of three parts. Part one assesses the challenges of using ADM systems and the rise of human rights concerns, particularly how they pertain to the prohibition of discrimination in Europe. Part two discusses the European response to these challenges through regulations of the CoE and judgments of the ECtHR. Finally, part three aims to raise awareness by sharing the views and recommendations of experts and scholars in this field.

II. DISCRIMINATION BY ADM SYSTEMS

Critics of ADM systems often point to their opacity and unpredictability.¹² These concerns also increase awareness that ADM systems may significantly curtail human rights. One of the fundamental human rights cited in this context is the prohibition of discrimination. In a 2018 study published by the CoE, Prof. Frederik Zuiderveen Borgesius wrote that, although ADM systems may produce discriminatory effects, “they do not necessarily perform worse than humans.”¹³ Indeed, in some

¹⁰ *Id.*

¹¹ *Id.*

¹² See, e.g., Symposium, *supra* note 2, at 619 (“[T]he opacity of decision-making algorithms prevents those harmed by automated systems from determining either how a decision came about or the logic and reasoning behind it.”).

¹³ FREDERIK ZUIDERVEEN BORGESIUS, COUNCIL OF EUR., DISCRIMINATION, ARTIFICIAL INTELLIGENCE, AND ALGORITHMIC DECISION-MAKING 31 (2018), available at <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>.

cases, ADM systems discriminate only “because they were trained on data that reflect discrimination by humans.”¹⁴ Absent this flawed input, ADM systems can be used “to discover existing inequality that might have remained hidden otherwise.”¹⁵

The legal distinction between direct and indirect discrimination may help ascertain whether ADM systems promote or prevent discrimination. This distinction has been described as follows:

Direct discrimination occurs where a decision-maker bases her decision directly on criteria or factors which are regarded as unlawful (such as race, ethnicity, religion, gender, sexual orientation, age, or disability). . . . Indirect discrimination occurs where a certain characteristic or factor occurs more frequently in the population groups against whom it is unlawful to discriminate (such as a person with a certain racial or ethnic background living in a certain geographical area; women having fewer pensionable years because of career breaks). Since algorithmic decision-making systems may be based on the correlation between data sets and efficiency considerations, there is a danger that such systems perpetuate or exacerbate indirect discrimination through stereotyping. Indirect discrimination is only present where differential treatment cannot be justified.¹⁶

Another study that did not rely on the legal distinction between direct and indirect discrimination uses the term “digital discrimination.” The study’s authors define the term as

¹⁴ *Id.*

¹⁵ *Id.*

¹⁶ COUNCIL OF EUR., *supra* note 6, at 27–28.

unfair treatment “caused by automated decisions, usually taken by intelligent agents or other AI-based systems.”¹⁷ Digital discrimination is widespread in many fields and is present in systems assessing risks for policing, banking loans, and insurance payouts.¹⁸

ADM systems are not intuitive when compared to human decision-making. These systems can be complex to the point that they defy human understanding. Therefore, discrimination by ADM systems can be difficult to detect.

There is no turning back from the era of digitalization. When using ADM systems, it is important to seek out and design systems to prevent differential treatment that is unjustified and unlawful. A decision that relies on racially biased data is indirectly discriminatory. As Dr. Carsten Orwat states, “[h]umans are the ones to set the decision-making rules or, in the case of machine-learning methods, algorithms generate parts of the decision-making rules based on the analysis of data.”¹⁹ When ADM systems are based on human decisions, “it is likely that the same biases which potentially undermine the human decision-making are replicated and multiplied in the algorithmic decision-making systems, only that they are then more difficult to identify and correct.”²⁰

It is difficult to determine whether humans are better at preventing discrimination than ADM systems. With ADM systems, however, it is harder to predict where discrimination begins and ends under the terms of legal liability.

¹⁷ XAVIER FERRER ET AL., BIAS AND DISCRIMINATION IN AI: A CROSS-DISCIPLINARY PERSPECTIVE 1 (2020), available at <https://arxiv.org/pdf/2008.07309.pdf>.

¹⁸ *Id.*

¹⁹ ORWAT, *supra* note 3, at 21.

²⁰ COUNCIL OF EUR., *supra* note 6, at 28.

III. EUROPEAN SAFEGUARDS AGAINST THE DISCRIMINATORY OUTCOMES OF ADM SYSTEMS

Safeguards are necessary to protect individuals from the discriminatory outcomes of ADM systems. These safeguards include legal instruments and judicial review of decisions made by ADM systems. While important, these safeguards have largely been left behind by the pace of technological advancement. There is “a critical incompatibility between European notions of discrimination and existing work on algorithmic and automated fairness.”²¹ On the other hand, non-discrimination and data protection laws are the primary legal regimes protecting people against AI-driven discrimination. In this context, the work of the CoE and the judgments of the ECtHR may serve as guides in the fight against discrimination.

Many treaties and constitutions prohibit discrimination, including the ECHR. Article 14 of the ECHR states, “The enjoyment of the rights and freedoms set forth in this Convention shall be secured without discrimination on any ground such as sex, race, color, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status.”²² The ECHR prohibits both direct and indirect discrimination. According to the ECtHR, direct discrimination arises where there is “a difference in the treatment of persons in analogous, or relevantly similar, situations,” and said differential treatment is

²¹ SANDRA WACHTER ET AL., WHY FAIRNESS CANNOT BE AUTOMATED: BRIDGING THE GAP BETWEEN EU NON-DISCRIMINATION LAW AND AI 1 (2020), *available at* <https://arxiv.org/ftp/arxiv/papers/2005/2005.05906.pdf>.

²² Convention for the Protection of Human Rights and Fundamental Freedoms, art. 14, Nov. 4, 1950, 213 U.N.T.S. 222.

based “on an identifiable characteristic.”²³ With regard to indirect discrimination, the ECtHR notes that “a difference in treatment may take the form of disproportionately prejudicial effects of a general policy or measure which, though couched in neutral terms, discriminates against a group. Such a situation may amount to ‘indirect discrimination,’ which does not necessarily require discriminatory intent.”²⁴ Thus, ADM systems may cause indirect discrimination under this definition.

Non-discrimination law has several weaknesses when it comes to ADM systems.²⁵ For instance, “The prohibition of indirect discrimination does not provide a clear and easily applicable rule.”²⁶ The ECtHR accepts that “a suspicion of indirect discrimination can be rebutted if the alleged discriminator can invoke an objective justification,” but whether the alleged discriminator can invoke an objective justification depends on all the circumstances of the case.²⁷ Because this test must be applied case-by-case, it is often unclear from the outset whether a decision or practice breaches the prohibition of indirect discrimination.²⁸ Additionally, indirect discrimination can remain hidden within ADM systems. For example, a customer who applies for a loan on a bank’s website may be denied by an ADM system due to the customer’s race; however, the customer has no way of knowing that the ADM system discriminated against him, making the decision difficult to challenge.²⁹ Another weakness of non-discrimination law is that statutes and regulations tend to focus on discrimination against protected classes.³⁰ However, ADM

²³ BORGESIU, *supra* note 13, at 32–33.

²⁴ *Id.* at 33.

²⁵ *Id.* at 34.

²⁶ *Id.*

²⁷ *Id.* at 34–35.

²⁸ *Id.*

²⁹ *Id.* at 36.

³⁰ *Id.*

systems discrimination may fall outside these laws' scope.³¹ In light of these weaknesses, the CoE's actions take on more importance. The CoE's Data Protection Convention 108 and the Charter of Fundamental Rights of the European Union require each of the CoE's member states to have an independent Data Protection Authority with the powers of investigation.³² A Data Protection Authority helps to mitigate the risks of illegal discrimination by conducting data protection audits and ordering data controllers to give access to their data processing systems.³³

Furthermore, Convention 108 gives individuals a right "to obtain, on request, knowledge of the reasoning underlying data processing where the results of such processing are applied to him or her."³⁴ In other words, this provision purports to allow individuals to peek behind the curtain of ADM systems. Still, it is difficult to determine the extent to which these provisions have helped curb discrimination by ADM systems.³⁵

In addition to Convention 108, the CoE has issued declarations and recommendations regarding the protection of individuals against the disadvantageous outcomes of ADM systems. "Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes" emphasizes that "particular attention should be paid to the significant power that technological advancement confers to those - be they public entities or private actors - who may use such algorithmic tools without adequate democratic oversight or control."³⁶ The declaration then draws attention "to the growing threat to the right of human beings to form opinions

³¹ *Id.*

³² *Id.* at 39.

³³ *Id.* at 39-40.

³⁴ *Id.* at 44.

³⁵ *Id.* at 46.

³⁶ COUNCIL OF EUR, *supra* note 7, at ¶ 8.

and [m]ake decisions independently of automated systems, which emanates from advanced digital technologies.”³⁷ The declaration also encourages member states to assume their responsibility to address this threat through cooperation and regulatory frameworks.³⁸

Recommendation CM/Rec (2020)1 of the Committee of Ministers to Member States on the Human Rights Impacts of Algorithmic Systems stresses “the need to ensure that racial, gender and other societal and labor force imbalances that have not yet been eliminated from our societies are not deliberately or accidentally perpetuated through algorithmic systems, as well as the desirability of addressing these imbalances through using appropriate technologies.”³⁹ In this context, the recommendation recalls “the obligation of Member States under the Convention to refrain from human rights violations, including through algorithmic systems . . . to establish effective and predictable legislative, regulatory and supervisory frameworks that prevent, detect, prohibit and remedy human rights violations.”⁴⁰

In accordance with the prohibition of discrimination, the recommendation further states that member states and private actors should, among other things: (1) “carefully assess what human rights and non-discrimination rules may be affected as a result of the quality of data that are being put into and extracted from an algorithmic system, as these often contain bias and may stand in as a proxy for . . . gender, race, religion, political opinion,” etc.; (2) “ensure that all relevant staff members involved in the procurement, development,

³⁷ *Id.* at ¶ 9.

³⁸ *Id.*

³⁹ RECOMMENDATION CM/REC (2020) 1 OF THE COMMITTEE OF MINISTERS TO MEMBER STATES ON THE HUMAN RIGHTS IMPACTS OF ALGORITHMIC SYSTEMS, *available at* https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154.

⁴⁰ *Id.*

implementation, assessment and review of algorithmic systems with significant human rights impacts are adequately trained with respect to human rights and non-discrimination rules”; (3) “follow a standard framework for human rights due diligence to avoid fostering or entrenching discrimination throughout all life-cycles of their systems”; (4) “seek to ensure that the design, development and ongoing deployment of their algorithmic systems do not have direct or indirect discriminatory effects on individuals or groups that are affected by these systems, including on those who have special needs or disabilities or who may face structural inequalities”; (5) “should be cognizant of risks relating to the quality, nature and origin of the data they are using for training their algorithmic systems, with a view to ensuring that errors, bias and potential discrimination in datasets and models are adequately responded to within the specific context.”⁴¹ The most recent work of the CoE is the “Declaration by the Committee of Ministers on the risks of computer-assisted or artificial-intelligence-enabled decision-making in the field of the social safety net.”⁴² The declaration states, in part:

These systems can, if not developed and used in accordance with principles of transparency and legal certainty, amplify bias and increase risks. . . . Under such circumstances, they can replicate entrenched discrimination patterns, including as regards women, and can affect people in low-skilled and poorly paid jobs. . . . Biased and/or erroneous automated decisions can bring about immediate destitution, extreme poverty or even

⁴¹ *Id.*

⁴² COUNCIL OF EUR., DECLARATION BY THE COMMITTEE OF MINISTERS ON THE RISKS OF COMPUTER-ASSISTED OR ARTIFICIAL-INTELLIGENCE-ENABLED DECISION MAKING IN THE FIELD OF THE SOCIAL SAFETY NET (2021), *available at* https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=0900001680a1cb98.

homelessness and cause serious or irreparable harm to those concerned.⁴³

The declaration further draws the attention of member states to “the need to ensure that computer-assisted or AI-enabled decision-making systems are developed and implemented in accordance with the principles of legal certainty, legality, data quality, non-discrimination, and transparency.”⁴⁴

The ECtHR has handed down judgments in cases where the use of ADM systems and personal data allegedly violated rights enshrined in the ECHR. In *S. and Marper v. U.K.*, the ECtHR observed that States should “strike a fair balance” between protecting fundamental rights and developing new technologies.⁴⁵ There, the ECtHR held that a database that stored “fingerprints, biological samples and DNA profiles from anyone suspected but not convicted of criminal offenses, whatever their age, the nature, and seriousness of the offenses, without a time-limit or any independent review of the justification of the retention of data” violated the right to privacy under Article 8 of the ECHR.⁴⁶ “The blanket and indiscriminate nature of such a system failed to reflect a fair balance between the competing public and private interests.”⁴⁷ The ECtHR noted that “retention of unconvicted persons’ data may be especially harmful in the case of minors . . . given their special situation and the importance of their development and integration in society.”⁴⁸

⁴³ *Id.*

⁴⁴ *Id.*

⁴⁵ *S. and Marper v. U.K.*, 1581 Eur. Ct. H.R. 33, 35 (2008).

⁴⁶ COUNCIL OF EUR. & EUR. CT. H.R., GUIDE TO THE CASE-LAW OF THE OF THE EUROPEAN COURT OF HUMAN RIGHTS, DATA PROTECTION 45 (2022), *available at* https://www.echr.coe.int/Documents/Guide_Data_protection_EN_G.pdf.

⁴⁷ *Id.* at 45–46.

⁴⁸ *S. and Marper v. U.K.*, 1581 Eur. Ct. H.R. at 34 (2008).

In *Weber and Saravia v. Germany*, the ECtHR clarified that laws allowing for surveillance and the transmission of intercepted data should specify a definition of the categories of people who might be subject to surveillance; a limit on the duration of the measure; the procedure to be followed; the precautions to be taken when communicating the data to other parties; and the circumstances in which recordings are to be destroyed.⁴⁹ According to the ECtHR, the law must be clear to the point that citizens have an idea of when authorities can interfere with their right to respect for private life and correspondence.⁵⁰ The ECtHR has also specified that any interference with an individual's right to privacy should be subject to an independent, impartial, and proper oversight system.⁵¹

The ECtHR has yet to decide cases relating to ADM systems and the prohibition of discrimination. While the ECtHR has examined issues closely linked to personal data protection under Article 8 (right to respect for private and family life) and Article 9 (freedom of thought, conscience, and religion), it has not found separate issues under Article 14 (prohibition of discrimination).

IV. RECOMMENDATIONS

Concerns and challenges regarding the use of ADM systems should be taken seriously, especially when they implicate human rights issues. Artificial intelligence is becoming more prevalent daily, permeating nearly every aspect of our lives. So how can we respond to these challenges? Although legal scholars have begun conducting more studies on ADM systems, it seems they will always be two steps behind the pace of technological advancement. Still, the

⁴⁹ *Weber and Saravia v. Ger.*, App. No. 54934/00, Eur. Ct. H.R., ¶ 95 (2006).

⁵⁰ *Malone v. U.K.*, 82 Eur. Ct. H.R. (ser. A), ¶ 67 (1984).

⁵¹ *Klass v. Ger.*, 28 Eur. Ct. H.R. (ser. A), ¶ 55 (1978).

recommendations of scholars can help raise awareness and prevent erosion of the rule of law, democracy, and respect for human rights.

Technology-driven life remains a regulation-free zone in many respects, leaving human rights at risk. As the Australian law professor Philip Alston states, “[t]he human rights community has thus far done a very poor job of persuading industry, Government, or seemingly, society at large of the fact that a technologically driven future will be disastrous if it is not guided by respect for human rights that is in turn grounded in law.”⁵² The author and history professor Yuval Harari agrees that global cooperation is needed to “regulate the explosive power of artificial intelligence” and to prevent the world’s data from falling into the hands of a powerful few.⁵³

The lack of legal principles and jurisprudence addressing human rights violations through ADM systems makes it difficult to evaluate these problems. “The courts do not provide a consistent and coherent approach to assessing prima facie discrimination. As a result, system developers, controllers, regulators, and users lack clear and consistent legal requirements that could be translated into system design and governance mechanisms to detect, remedy, and prevent automated discrimination.”⁵⁴ Dr. Jennifer Cobbe suggests that, “to properly confront the challenge of the algorithmic state,

⁵² PHILIP ALSTON, U.N. GEN. ASSEMBLY, REPORT OF THE SPECIAL RAPPORTEUR ON EXTREME POVERTY AND HUMAN RIGHTS 14 (2019), *available*

at <https://undocs.org/Home/Mobile?FinalSymbol=A%2F74%2F493&Language=E&DeviceType=Desktop&LangRequested=False>.

⁵³ Yuval Harari Warns Humans Will be “Hacked” if Artificial Intelligence Is Not Globally Regulated, CBS NEWS (Oct. 29, 2021), <https://www.cbsnews.com/news/yuval-harari-sapiens-60-minutes-2021-10-29/>.

⁵⁴ WACHTER, *supra* note 21, at 44.

new legal concepts, frameworks, and oversight mechanisms may be needed for the problems of an algorithmic system applied across many decisions, with limits on when algorithmic systems can and can't be used, and an oversight body empowered to investigate complaints."⁵⁵

The legislative process is often slow, and the creation of new rules may lag behind rapidly developing technologies. In this context, Frederik Borgesius' suggestion may prove useful:

The statutes could be phrased in a reasonably technology-neutral way. Technology-neutral legal provisions with broad principles have the advantage of not having to be changed every time a new technology is developed. A disadvantage is that broad principles can be difficult to apply in practice. Therefore, guidance by regulators can be useful. Guidelines can be amended faster and thus be more specific and concrete.⁵⁶

There is no doubt that we need new legal rules to ensure that ADM systems do not infringe upon human rights. We also need courts to interpret these rules so that we can confront these challenges.

⁵⁵ Jennifer Cobbe, *Confronting the Algorithmic State*, ADMIN. L. IN THE COMMON L. WORLD (Sept. 24, 2020), <https://adminlawblog.org/2020/09/24/jennifer-cobbe-confronting-the-algorithmic-state/>.

⁵⁶ BORGESIOUS, *supra* note 13, at 61-62.