BRILL
NIJHOFF

INTERNATIONAL HUMAN RIGHTS LAW REVIEW
10 (2021) 1-39

IHRL REVIEW
brill.com/hrlr

Meeting the Chimera: How the CEDAW Can Address Digital Discrimination

Tetyana (Tanya) Krupiy
Postdoctoral fellow, Tilburg Law School, Tilburg University, Tilburg, The Netherlands
t.krupiy@tilburguniversity.edu; krupiy.tanya@hotmail.com

Abstract

The article analyses what is distinct about the manner in which the delegation of the decision-making task to an artificial intelligence system produces harm from the standpoint of the prohibition of discrimination. It explores the manner in which the context of digital discrimination challenges the application of the Convention on the Elimination of all Forms of Discrimination Against Women (CEDAW). The article suggests how the subject matter of CEDAW may be rethought to enable it to respond to digital discrimination. It formulates a legal test which can be added to the existing toolbox without the need to amend the treaty. The article offers approaches to interpreting CEDAW teleologically in order to enable it to remain relevant in the face of technological innovation.

Keywords

international human rights law – digital discrimination – intersectionality – CEDAW – rights of women

1 Introduction

In June 2020 the United Nations (UN) Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance E Tendayi Achiume called on States to protect individuals from discrimination in the context of the use of artificial intelligence decision-making processes.[1] The

1 United Nations Human Rights Council, 'Racial Discrimination and Emerging Digital Technologies: a Human Rights Analysis' (18 June 2020) UN Doc A/HRC/44/57 para 52.

UN Special Rapporteur released the report after mounting evidence that the use of artificial intelligence systems disadvantages people of colour,[2] women[3] and individuals with a disability.[4] The term 'digital discrimination' denotes discrimination which occurs as an outcome of the use of 'automated decisions taken by algorithms.'[5] Hiring is one of the contexts in which organisations have an interest in using artificial intelligence systems.[6] They employ artificial intelligence systems to predict the future performance of candidates[7] and as an assistive tool in the decision-making process.[8]

States recognise the role which international human rights law has in guiding the regulation of emerging digital technologies.[9] To illustrate, the Council of Europe Committee of Experts on Human Rights recommended that Member States develop legislation which allows all actors to respect and to promote human rights.[10] However, there is variation in the degree to which

2 S Brownstone, 'This Data Tool Helps Homeless People Get Housing. If You're White, Your Chances Are Even Better' *The Seattle Times* (Seattle, 29 October 2019) <www.seattletimes.com/seattle-news/homeless/this-data-tool-helps-homeless-people-get-housing-if-youre-white-your-chances-are-even-better>.

3 S Natarajan and S Nasiripour, 'Viral Tweet About Apple Card Leads to Goldman Sachs Probe' (Bloomberg LP, 9 November 2019) <www.bloomberg.com/news/articles/2019-11-09/viral-tweet-about-apple-card-leads-to-probe-into-goldman-sachs>.

4 A Engler, 'For Some Employment Algorithms, Disability Discrimination by Default' (The Brookings Institution, 2019) <www.brookings.edu/blog/techtank/2019/10/31/for-some-employment-algorithms-disability-discrimination-by-default>.

5 N Criado and JM Such, 'Digital Discrimination' in Karen Yeung and Martin Lodge (eds), *Algorithmic Regulation* (Oxford University Press 2019) 1.

6 E Jacobs, 'How Artificial Intelligence Helps Companies Recruit Talented Staff' *Financial Times* (London, 25 February 2019) <www.ft.com/content/2731709c-3043-11e9-8744-e7016697f225>.

7 A Prince and D Schwarcz, 'Proxy Discrimination in the Age of Artificial Intelligence and Big Data' (2020) 105 Iowa LR 1257.

8 Ibid, 13.

9 C Bradley and R Wingfield, 'National Artificial Intelligence Strategies and Human Rights: a Review' (Stanford Cyber Policy Center 2020) 21; Committee of Experts on Human Rights Dimensions of Automated Data Processing and Different Forms of Artificial Intelligence MSI-AUT, 'Addressing the Impacts of Algorithms on Human Rights: Draft Recommendation of the Committee of Ministers to Member States on the Human Rights Impacts of Algorithmic System Doc MSI-AUT(2018)06rev3' (Council of Europe 2019) Appendix para 1; R Vought, 'Memorandum for the Heads of Executive Departments and Agencies M-20–34' (Executive Office of the President, Office of Management and Budget 2020) Principle 7; V Putin, 'Decree of the President of the Russian Federation No 490 of 10 October 2019 on the Development of Artificial Intelligence in the Russian Federation' (Office of the President of the Russian Federation 2019) para 19(a).

10 Committee of Experts on Human Rights Dimensions of Automated Data Processing and Different Forms of Artificial Intelligence <https://www.coe.int/en/web/freedom-expression/msi-aut>5.

States place international human rights law as a pillar around which to develop policy.[11] For instance, East and Southeast Asian States, such as Singapore[12] and South Korea,[13] frame regulatory interventions by reference to ethics. Their strategic documents refer to developing a 'human-centric' approach to regulation.[14] Against this background a group of civil society groups issued the Toronto Declaration.[15] This Declaration calls on States to use international human rights law norms as a foundation for developing ethical principles to guide the development and use of artificial intelligence technologies.[16] Yet, the Council of Europe Ad hoc Committee on Artificial Intelligence explains that existing legal instruments do not provide 'an adequate and specific response to the challenges brought by' artificial intelligence technology.[17] This stems from the fact that States did not draft them with this technology in mind.[18] The article will demonstrate that it is possible to develop existing treaty provisions through legal interpretation to enable them to remain relevant.

The article contributes to existing knowledge by exploring the distinct manner in which harms associated with discriminatory treatment arise in the context of the employment of a fully autonomous decision-making process (hereinafter AIDMP). It fills a gap in existing scholarship because there is little literature on how the CEDAW applies to the context of digital discrimination. To date, most authors have focused on the issue of digital discrimination from the standpoint of either domestic law[19] or the European Convention on Human Rights.[20] There is a need for scholarship which expressly engages with CEDAW. CEDAW is an example of a specialist treaty which adopts a more

11 Bradley and Wingfield (n 9) 21.

12 Smart Nation & Digital Government Office, 'National Artificial Intelligence Strategy: Advancing Our Smart Nation Journey' (Smart Nation & Digital Government Office 2019) 10.

13 Government of the Republic of Korea, 'Mid to Long-term Master Plan in Preparation for the Intelligent Information Society: Managing the Fourth Industrial Revolution' (Government of the Republic of Korea 2016) 56.

14 Ibid; Smart Nation & Digital Government Office (n 12) 10.

15 Amnesty International and Access Now, 'The Toronto Declaration' (Amnesty International & Access Now, 2021) <www.torontodeclaration.org>.

16 Ibid.

17 Council of Europe Ad hoc Committee on Artificial Intelligence, 'Feasibility Study CAHAI(2020)23' (Council of Europe 2020) para 82.

18 Ibid.

19 B Casey, 'Title 2.0: Discrimination in a Data-driven Society' (2019) 2019 J Law and Mobility 36, 47; M Raub, 'Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices' (2018) 71 Ark L Rev 529, 544.

20 FJZ Borgesius, 'Strengthening Legal Protection Against Discrimination by Algorithms and Artificial Intelligence' (2020) 24 IJHR 1572, 1576; E Lundberg, 'Automated Decision-making vs Indirect Discrimination: Solution or Aggravation? ' (Umea University 2019).

comprehensive approach to the protection against discrimination in comparison with general international human rights law treaties.[21] CEDAW placed an additional obligation on States to combat the structural causes of gender discrimination through introducing the concept of transformative equality.[22] The article advances knowledge by proposing how the CEDAW provisions may be developed through legal interpretation to enable the treaty to respond to digital discrimination. It formulates a new test which can be added to the existing toolbox to enable CEDAW to respond to technological innovation. This does not mean that the conclusion of a treaty addressing the context of artificial intelligence is undesirable. Past experience shows that individuals experience a 'giant leap' in the enjoyment of fundamental rights when States tailor human rights protections to a particular context.[23]

Following this introduction, section 2 introduces the core concepts which CEDAW utilises to construct a framework for protecting individuals from discrimination. It explains that the view that international human rights law is individualistic in its focus[24] is an oversimplification. Section 3 analyses the distinct manner in which digital discrimination occurs. The employment context serves as a case study to contextualise the discussion.

Section 4 discusses how the subject matter of the protection can be rethought to enable CEDAW to respond to digital discrimination. Section 5 considers what challenges the digital context presents for the application of CEDAW. It analyses how the CEDAW may be interpreted so as to broaden the subject matter of the protection to include the individuality of people and human diversity. It suggests how the provisions and concepts within CEDAW may be developed through legal interpretation in order to enable the treaty to address a broader array of harms associated with digital discrimination. The article articulates an additional test defining discrimination which can be added to the existing toolbox to enable the treaty to respond to the digital context.

21 R Holtmaat, 'The CEDAW: a Holistic Approach to Women's Equality and Freedom' in A Hellum and HS Aasen (eds), *Women's Human Rights: CEDAW in International, Regional and National Law* (Cambridge University Press 2013) 96.

22 Ibid.

23 OM Arnardottir, 'A Future of Multidimensional Disadvantage Equality?' in OM Arnardottir and G Quinn (eds), *The U.N. Convention on the Rights of Persons with Disabilities: European and Scandinavian Perspectives* (Martinus Nijhoff Publishers 2009) 41.

24 R Kapur, *Gender, Alterity and Human Rights: Freedom in a Fishbowl* (Edward Elgar Publishing 2018) 1.

2 CEDAW: an Introduction of Core Concepts

The Universal Declaration of Human Rights states that the core purpose of the prohibition of discrimination is to protect the individuals' self-respect[25] and dignity.[26] The most relevant provisions of CEDAW to the present discussion will be introduced now as a means of illustrating how it protects individuals from discrimination. There is a relationship between the concepts the CEDAW uses to protect individuals from discrimination and the cognitive foundation underpinning the treaty. Article 1 CEDAW requires the protection of equality on the basis of sex.[27] According to the Committee on the Elimination of Discrimination Against Women (hereinafter CEDAW Committee), provisions of CEDAW should be interpreted in light of its general recommendations, views, concluding observations and statements.[28] While the general comments are not legally binding on States,[29] States attach great weight to them.[30] The CEDAW Committee interprets the term sex in its General Recommendation 28 as referring to the biological characteristics of women.[31] The CEDAW Committee uses the term gender to refer to the manner in which society constructs women's identities, attributes and roles.[32] The scope of the treaty extends to sexual and gender identity.[33] Article 2 enumerates the steps which States have to take to eliminate discrimination against women.[34] The CEDAW Committee interpreted Article 2 as requiring duty bearers to recognise that women may suffer from additional grounds of discrimination including race, ethnicity, age, class, religious identity and disability.[35] Since the CEDAW Committee uses the

25 T Khaitan, *A Theory of Discrimination Law* (Oxford University Press 2015) 18.

26 Universal Declaration of Human Rights, GA res. 217A (III), UN Doc A/810 at 71, (10 December 1948), Preamble.

27 Convention on the Elimination of all Forms of Discrimination Against Women (adopted 18 December 1979, entered into force 3 September 1981) 1249 UNTS 13 (CEDAW) Art 1.

28 UN Committee on the Elimination of Discrimination against Women, 'General Recommendation No 28 on the Obligations of States Parties under Article 2 of the Convention on the Elimination of All Forms of Discrimination against Women' (16 December 2010) UN Doc CEDAW/C/GC/28 para 7 (General Recommendation No 28).

29 International Law Association, 'Final Report on the Impact of Findings of the United Nations Human Rights Treaty Bodies' (International Law Association 2004) 3–4.

30 Ibid 5.

31 General Recommendation No 28 (n 28) para 5.

32 Ibid.

33 UN Committee on the Elimination of Discrimination against Women, 'Concluding Observations of the Committee on the Elimination of Discrimination against Women: Albania' (16 September 2010) UN Doc CEDAW/C/ALB/CO/3 para 43.

34 CEDAW, Art 2.

35 General Recommendation No 28 (n 28) para 18.

language of 'or other factors'[36] in the General Recommendation 25, it interprets CEDAW as covering all aspects of personal identity which may trigger discriminatory treatment. By using distinct categories, the CEDAW Committee interprets the subject matter of the protection as being based on the possession of personal characteristics related to group membership.

Ratna Kapur argues that individualism is at the epicentre of human rights.[37] International human rights law places the individual at the centre of the analysis.[38] Although CEDAW places the individual as the focus of the analysis, this is only a part of the story. The prohibition of discrimination is individualistic in its focus because it centres on persons and on the possession of protected characteristics. Paragraph 1 of the Preamble reaffirms the 'dignity and worth of the human person.'[39] According to the CEDAW Committee, CEDAW conceives of equality in terms of women having the ability to make choices and to develop themselves.[40] Marsha Freeman, Christine Chinkin and Beate Rudolf explain that the CEDAW positions individuals as autonomous subjects who are responsible for their choices.[41] The individualistic focus of the treaty can be gleaned additionally from how CEDAW protects women who possess more than one protected characteristic. The CEDAW Committee explains in the General Recommendation 28 that Article 2 CEDAW requires States to recognise in law the concept of intersectionality.[42] Intersectionality recognises that women who possess additional protected characteristic(s) experience discrimination in a different way and to a different degree than men with the same protected characteristic(s).[43]

By focusing on the individual and how that individual experiences disadvantage, the concept of intersectionality exhibits an individualistic approach. Had the concept of intersectionality not been individualistic in its focus, it would have recognised the relational nature of human existence. According to social constructionism individuals create the world through using particular

36 UN Committee on the Elimination of Discrimination against Women, 'General Recommendation No 25, on Article 4, Paragraph 1, of the Convention on the Elimination of All Forms of Discrimination against Women, on Temporary Special Measures' (2000) UN Doc A/55/18 (General Recommendation No 25), Annex V 152 para 12.

37 Kapur (n 24) 1

38 Ibid 14.

39 CEDAW, Preamble para 1.

40 General Recommendation No 28 (n 28) para 22.

41 MA Freeman, C Chinkin and B Rudolf, *The UN Convention on the Elimination of All Forms of Discrimination Against Women: A Commentary* (Oxford University Press 2012) 154.

42 General Recommendation No 28 (n 28) para 18.

43 Ibid.

kind of language to describe it.[44] The manner in which individuals together generate meaning and demarcate legitimate ways of acting plays a role in discrimination. For instance, women in academia describe being dismissed when they raise a concern about problematic conduct.[45] Similarly, LGBTQ+ scientists describe how their colleagues do not want to learn about why their behaviour is perceived as creating a non-inclusive environment.[46]

It is an oversimplification to describe CEDAW as being solely premised on individualism. The CEDAW Committee interprets the prohibition of discrimination in a manner which recognises the collective dimension of human relations. The prohibition of discrimination in Article 1 CEDAW covers the exclusion of individuals on the basis of sex.[47] The term exclusion refers to patterns of belief and social practices which deny individuals opportunities.[48] The CEDAW Committee interprets Articles 1, 2, 3, 4, 5 and 25 in conjunction as requiring States to address 'prevailing gender relations' and gender-based stereotypes in order to eliminate discrimination on the basis of sex.[49] The CEDAW Committee recognises in part the role which the interactions between individuals at the collective level play in producing meaning and in demarcating behavioural expectations. It acknowledges in the General Recommendation 28 that unequal gender relations vest power in the dominant group to define meaning.[50] The dominant group defines how the dominated group is expected to behave.[51] This creates gender stereotypes.[52]

CEDAW addresses the intermediate level at which individuals negotiate meaning with each other on a daily basis indirectly. CEDAW conceives of stereotypes and gender relations as being the result of the operation of legal norms, legal institutions, societal institutions and societal structures.[53] Culture,

44 JM Watkins, BJ Mohr and R Kelly, *Appreciative Inquiry: Change at the Speed of Imagination* (John Wiley & Sons 2011) 16.

45 Female Graduate Students and Research Staff in the Laboratory for Computer Science and the Artificial Intelligence Laboratory at MIT, *Barriers to Equality in Academia: Women in Computer Science at MIT* (Massachusetts Institute of Technology 1983) 20.

46 E Gibney, 'Discrimination Drives LGBT+ Scientists to Think About Quitting' (2019) 571 Nature 16.

47 CEDAW, Art 1.

48 RJ Cook and S Cusack, *Gender Stereotyping: Transnational Legal Perspectives* (University of Pennsylvania Press 2009) 109.

49 General Recommendation No 25 (n 36) paras 6–7.

50 General Recommendation No 28 (n 28) para 5.

51 Ibid.

52 Ibid para 9.

53 General Recommendation No 25 (n 36) para 7.

community and society define the position of women.[54] Political, economic, cultural, social, religious, ideological and environmental factors play a role in this process.[55] Societal and legal institutions are a product of individuals acting collectively through governmental organs. The references to culture, community and society point to the levels of both society and groups within society. The same can be said for decisions affecting religion, ideology and societal arrangements. The treaty obligations to counter gender stereotypes encompass the interactions between individuals within a group. This stems from the fact that Article 4(1) CEDAW extends the application of the duty to address the causes of discrimination to both public and private actors.[56] Moreover, since the interactions at the intermediate level add up to produce change at the level of culture, the obligation to address prevailing gender relations extends to the intermediate level of interpersonal interactions.

In protecting women from discrimination CEDAW makes visible the protected characteristic and how human conduct at the collective level interacts with the protected characteristic. Paradoxically, this approach renders the individual both visible and invisible simultaneously. The focus on the protected characteristic does not bring to light the complex manner in which the individual is experiencing discriminatory conduct. As a result, the prohibition does not capture all problematic conduct. CEDAW prohibits direct and indirect discrimination.[57] Direct discrimination occurs when an individual treats another individual differently explicitly based on grounds of sex and gender differences.[58] Since CEDAW focuses on differential treatment based on sex, it does not cover conduct where an individual invokes a reason unrelated to the protected characteristic as a basis for the decision. This is problematic because individuals can adopt tactics to conceal the fact that they are treating an individual differently on the basis of a protected characteristic. A Latina American scholar Francisca de la Riva-Holly[59] describes her Latina male colleagues labelling her as non-collegial to justify subjecting her to negative treatment.[60] In fact, they were motivated by bias in relation to her sex, race and

54 General Recommendation No 28 (n 28) para 5.
55 Ibid.
56 General Recommendation No 25 (n 36) para 10.
57 Ibid, para 7.
58 General Recommendation No 28 (n 28) para 16.
59 Francisca de la Riva-Holly, 'Igualadas' in GG y Muhs, YF Niemann and CG González (eds), *Presumed Incompetent: The Intersections of Race and Class for Women in Academia* (Utah State University Press, 2012) 289.
60 Ibid, 292–93.

socio-economic background.[61] Francisca de la Riva-Holly would find it difficult to invoke the prohibition of direct discrimination. Employers can justify treating an individual as a less effective employee by reference to collegiality. This is because teamwork is one of the criteria for good job performance.

Indirect discrimination occurs 'when a law, policy, programme or practice appears to be neutral' but has a discriminatory effect in practice on the group because the apparently neutral measure does not address pre-existing inequalities.[62] Indirect discrimination renders it invisible how a particular individual with protected characteristics experiences a particular rule and practice. Only those disadvantages which manifest themselves on a group level and affect a sufficient number of individuals gain visibility. The Committee on Economic, Social and Cultural Rights clarifies that there should be 'relative disadvantages for some groups' for there to be indirect discrimination.[63] By requiring States to achieve substantive equality,[64] the CEDAW Committee mitigates in part the concern that in order to be recognised as unlawful the discriminatory practices have to affect a sufficient number of individuals. This is because it is concerned with whether there is equality of opportunity rather than with whether a neutral measure disproportionately affects a group in a negative manner. The CEDAW Committee defines substantive equality in the General Recommendation 25 as requiring that 'women be given an equal start and that they be empowered by an enabling environment to achieve equality of results.'[65]

The principle of substantive equality does not make it possible to fully address discrimination. The fact that an organisation has an equal number of men and women does not mean that it achieved equality. For instance, women can be exposed to a hostile working environment unlike men notwithstanding their equal representation. Numerical parity can hide the fact that women who possess additional protected characteristics are underrepresented. Finally, substantive equality makes invisible the microaggressions and other types of problematic conduct individuals with additional protected characteristics may experience from their colleagues. Derald Wing Sue coined the term 'microaggressions' to describe the following types of intentional and unintentional treatment:

61 Ibid.

62 General Recommendation No 28 (n 28) para 16.

63 UN Committee on Economic, Social and Cultural Rights, 'General Comment No 20: Non-discrimination in Economic, Social and Cultural Rights (art 2, para 2 of the International Covenant on Economic, Social and Cultural Rights)' (2 July 2009) UN Doc E/C.12/GC/20 para 12.

64 General Recommendation No 25 (n 36) para 7.

65 Ibid, para 8.

> everyday verbal, nonverbal, and environmental slights…which communicate hostile, derogatory, or negative messages to target persons based solely upon their marginalized group membership. In many cases, these hidden messages may invalidate the group identity or experiential reality of target persons, demean them on a personal or group level, communicate they are lesser human beings, suggest they do not belong with the majority group, threaten and intimidate, or relegate them to inferior status and treatment.[66]

This stems from the fact that the principle of substantive equality focuses on whether an individual has access to opportunities rather than on the challenges which an individual experiences after having gained such access.

The principle of transformative equality mitigates the gap in the principle of substantive equality. The CEDAW Committee interprets state obligations under Article 4(1) in the General Recommendation 25 as including the duty to achieve transformative equality by remedying the underlying causes of inequality.[67] The States should transform 'opportunities, institutions and systems so that they are no longer grounded in historically determined male paradigms of power and life patterns.'[68] To achieve this, States should consider the lives of women and men 'in a contextual way.'[69] Transformative equality addresses inequality stemming from how organisational processes are constituted. For instance, long work hours, work culture[70] and lack of consideration for the need to be inclusive disadvantage women.[71] To address workplace discrimination it is necessary to change structures, systems, practices, work culture and how individuals interact.[72] Transformative equality addresses these elements by obliging duty bearers to change those aspects of the design and operation of the institutions which are the root causes of inequality.

66 DW Sue, 'Microaggressions: More Than Just Race' (*Sussex Publishers LLC,* 2010) <www.psychologytoday.com/us/blog/microaggressions-in-everyday-life/201011/microaggressions-more-just-race>.

67 General Recommendation No 25 (n 36) para 10.

68 Ibid.

69 Ibid.

70 TK Green, *Discrimination Laundering: the Rise of Organisational Innocence and the Crisis of Equal Opportunity Law* (Cambridge University Press 2017) 150.

71 Ibid 123.

72 Ibid, 116.

3 The Distinct Character of Digital Discrimination

There is no accepted definition of artificial intelligence in the scientific community.[73] For the purpose of the present discussion an AIDMP is defined as encompassing all stages involved in the construction and operation of a decision-making process which operates on artificial intelligence software.[74] The AIDMP begins with the developer formulating the problem to be solved and the goal to be achieved.[75] For instance, the goal to be achieved can refer to identifying individuals whose future performance is likely to meet a variety of criteria associated with being a good employee. The definition encompasses the process of 'gathering, combining, cleaning, sorting and classifying data' for the purpose of creating a model representing a pool of candidates.[76] It includes the 'process of selection, prioritisation, recommendation and decision-making' which leads to the allocation of a positive decision to the candidate.[77] This definition is broader than that of the Committee of Experts of the Council of Europe. The Experts of the Council of Europe treat the algorithmic process as starting with the stage of gathering data.[78] The basis for using a broader definition is that how a computer scientist defines criteria for a good employee influences what relationships between the data the AIDMP detects.[79] In turn, the type of detected relationships bear on what predictions the AIDMP generates and whether a candidate receives a positive decision.[80]

Sandra Wachter maintains that the use of AIDMPs produces new classes of disadvantaged groups which do not correspond to the protected characteristics the law traditionally protected.[81] Examples of such categories includes dog ownership and being a video gamer.[82]

Wachter does not fully capture the distinct manner in which the operation of the AIDMP produces legally relevant harms. In many cases it will not be

73 Council of Europe Ad hoc Committee on Artificial Intelligence (n 17) para 5.

74 Tetyana (Tanya) Krupiy, 'A Vulnerability Analysis: Theorising the Impact of Artificial Intelligence Decision-Making Processes on Individuals, Society and Human Diversity from a Social Justice Perspective' (2020) 38 Computer Law and Security Review 1, 7.

75 Ibid.

76 Committee of Experts on Human Rights Dimensions of Automated Data Processing and Different Forms of Artificial Intelligence (n 10) para 3.

77 Ibid.

78 Ibid.

79 S Barocas and AD Selbst, 'Big Data's Disparate Impact' (2016) 104 CLR 671, 679.

80 Ibid, 679–80.

81 S Wachter, 'Affinity Profiling and Discrimination by Association in Online Behavioural Advertising' (2019) 35 Berkeley Tech LJ 1, 56.

82 Ibid.

possible to establish a causal link between adverse treatment and membership of a distinct group. This aspect poses a challenge for applying CEDAW because the treaty conceives of the subject matter of the protection in terms of the possession of a distinct identifiable common characteristic. Current scholarship engages with this issue to a limited extent.

Consider the scholarship of Anya Prins and Daniel Schwarcz. They coined the term 'proxy discrimination' to describe the mechanism through which the operation of the AIDMP produces a prohibited outcome.[83] Prins and Schwarcz point out that when processing data the software seeks out information correlated with group membership.[84] This stems from the fact that it uses all the information with predictive capacity.[85] Specifically, the practice of relying on the AIDMP appears neutral but it in fact disproportionally harms individuals with a protected characteristic.[86] A more detailed description of how AIDMPs generate a model of the environment and produce predictions contextualises the argument of Prins and Schwarcz. The AIDMP uses data about applicants to represent them in a mathematical model.[87] The model detects patterns or correlations in the data[88] as part of identifying the relation between pieces of information.[89] The system allocates individuals into groups based on shared characteristics.[90] The AIDMP predicts an applicant's performance in light of the individuals' data whom it treats as being sufficiently similar to the applicant.[91] Since the AIDMP groups data on the basis of similarity[92] it necessarily seeks out criteria corresponding to group membership.

The concept of 'proxy discrimination'[93] does not capture the fact that it will not be possible to determine the relationship between the input information which the AIDMP uses to predict the candidate's performance to the possession of the protected characteristic. Scientific research and in some cases empirical evidence is needed to establish the existence of a relationship between the protected characteristic and the input. To illustrate, in the future the AIDMPs may use inputs relating to the off-duty conduct of the employees

83 Prince and Schwarcz (n 7) 4.
84 Ibid 7–8.
85 Ibid.
86 Ibid, 4.
87 Foster Provost and Tom Fawcett, *Data Science for Business* (O'Reilly Media Inc 2013) 39.
88 Ibid, 25.
89 Ibid, 23.
90 Ibid, 24.
91 Ibid, 146.
92 Ibid, 24.
93 Prince and Schwarcz (n 7) 4.

in order to predict their leadership potential and future career success.[94] The inputs include 'taste in beer,' music preferences and what type of newspaper the candidate reads alongside thousands of other variables.[95] According to sociologist Pierre Bourdieu, the upbringing and social environment determine the preferences of individuals and how others perceive their choices.[96] Individuals from a similar socio-economic and cultural background exhibit similar tastes.[97] The use of information about personal tastes and habits creates a situation where a dominant group defines a good candidate in a manner which confers on it an advantage. However, it is difficult in practice to trace how the inequality in accessing opportunities is linked to the possession of a protected characteristic. One would need to know how music preferences are related to the socio-economic background of the candidate for instance. Given that the AIDMP uses thousands of variables as inputs,[98] it may be financially impossible to carry out scientific research on the relationship between each input variable and the possession of a protected characteristic. A relevant factor is that it is challenging to determine how a particular input variable is linked to all protected characteristics. For instance, individuals of all genders from affluent socio-economic backgrounds could favour beer with a particular taste.

Moreover, it is not apparent how each variable which the AIDMP uses as an input for the purpose of producing a decision is connected to the possession of a protected characteristic. There may be degrees to which an input variable correlates with a protected characteristic. To illustrate, organisations use the artificial intelligence software HireVue to make hiring decisions based on the data about the candidate's tone of voice[99] and sentence structure.[100] Features of speech provide information about the 'status and group membership of the speakers.'[101] Research demonstrates that people regard individuals who have

94 Julie Manning Magid, 'Does Your AI Discriminate?' (*The Conversation,* 3 July 2020) <https://theconversation.com/does-your-ai-discriminate-132847>.

95 Ibid.

96 P Bourdieu, 'Les Trois Etats du Capital Culturel' (1979) 30 Actes De la Recherche en Sciences Sociales 3, 4; P Bourdieu, 'Men and Machines' in KK Cetina and AV Cicourel (eds), *Advances in Social Theory and Methodology: Toward an Integration of Micro-and Macro-Sociologies* (Routledge and Kegan Paul, 1981) 308; MV Reiss and M Tsvetkova, 'Perceiving Education from Facebook Profile Pictures' (2020) 22 New Media & Society 550, 552.

97 Reiss and Tsvetkova, ibid, 553.

98 Magid (n 94).

99 Richard Feloni, 'Consumer-goods Giant Unilever has been Hiring Employees Using Brain Games and Artificial Intelligence—and It's a Huge Success' (*Business Insider France,* 2017) <www.businessinsider.fr/us/unilever-artificial-intelligence-hiring-process-2017-6>.

100 Magid (n 94).

101 Michael Hogg and Graham Vaughan, *Social Psychology* (4th edn, Prentice Hall 2005) ch 15.

standard pronunciation as possessing greater competency and intelligence than individuals with non-standard accents.[102] There is a connection between the possession of a protected characteristic and the degree to which an individual exhibits a standard way of speaking. For instance, individuals with a migrant background have speech patterns which differ from native speakers. Individuals who attended private schools are likely to exhibit standard patterns of speaking to a greater degree than individuals who attended a school in a poorly funded school district due to disparities in resources. To illustrate, private schools in the United Kingdom have three times more resources than state schools and have small classes.[103] However, individuals have a degree of control over the extent to which their speech pattern corresponds to that of a dominant group. For instance, bicultural individuals are able to tailor how they communicate a message to the culture of the dominant group.[104] Circumstances influence the extent to which individuals can mitigate the impact of coming from a poor socio-economic or migrant background. The degree of familiarity which individuals have with the culture of the dominant group varies depending on the length and intensity of their experience with the dominant culture.

The nature of technical systems compounds the possibility that it will be impossible to detect the relationship between the use of a particular input and the protected characteristic notwithstanding the fact that there is a relationship between the decision and the possession of a protected characteristic. Ludwig von Bertalanffy developed the general system theory to describe principles which are valid for a 'system.'[105] The principles of the theory apply to all systems irrespective of the nature of the system's elements and the relations between the elements.[106] His theory is relevant to analysing the AIDMP. Systems which regulate their internal states through receiving ongoing external feedback are a subclass of 'general systems.'[107] The AIDMP falls into this

102 P Powesland and H Giles, 'Persuasiveness and Accent-message Incompatibility' (1975) 28 Human Relations 85, 89–90; D Andrews, 'Subjective Reactions to Two Regional Pronunciations of Great Russian: A Matched-Guise Study' (1995) 37 Canadian Slavonic Papers 89, 99.

103 F Green and Kynaston, 'Britain's Private School Problem: It's Time to Talk' *The Guardian* (London, 13 January 2020).

104 C Chiu and YY Hong, 'Cultural Competence: Dynamic Processes' in AJ Elliot and CS Dweck (eds), *Handbook of Competence and Motivation* (Guildford Press, 2005) 498.

105 L von Bertalanffy, 'An Outline of General System Theory ' (1950) 1 The British Journal for the Philosophy of Science 134, 139.

106 Ibid.

107 L von Bertalanffy, 'General Systems Theory and Psychiatry – An Overview' in W Gray, F Duhl and N Rizzo (eds), *General Systems Theory and Psychiatry* (Little Brown and Company 1969) 37.

category because it is a self-regulating system which operates through receiving ongoing feedback. An artificial intelligence system continuously updates its model based on the inputs it receives.[108] The general system theory stipulates that when components within a system interact they produce emergent effects which are more than the sum of their parts.[109] The system 'behaves as a whole.'[110] Every element performs differently when operating as part of a system in comparison to when it is isolated.[111]

The application of the general system theory to the context of the AIDMP yields the following insights. When developers program the AIDMP based on a neural network, the neural network represents a simplified mathematical model of the environment.[112] When the neurons in a neural network interact with each other and pass on information to higher levels in the neural network,[113] emergent effects occur. Emergent effects take place when the stages involved in processing the data interact. These stages include the AIDMP mapping data in the mathematical space, detecting correlations between the data and using data to make predictions about a particular candidate. A multitude of interactions occur and produce new emergent effects. The presence of emergent effects points to the fact that it is insufficient to protect individuals from digital discrimination by defining the subject matter of the protection by reference to the possession of a protected characteristic. The concept of a protected characteristic does not capture the impact which arises from the interaction between the inputs as well as between stages in the decision-making process. There is a need to rethink the subject matter of the protection in the context of digital discrimination.

Furthermore, it is limiting to conceive of the subject matter of the protection by reference to the stereotypical representations of a group with a protected characteristic in the context of the use of AIDMPs. According to Solon Barocas and Andrew Selbst, 'data mining holds the potential to unduly discount members of legally protected classes and to place them at systemic relative

108 A Agrawal, J Gans and A Goldfarb, 'How to Win with Machine Learning' *Harvard Business Review* (Brighton, September-October 2020); IBM, 'Data Science and Machine Learning ' (*IBM,* 2021) <https://ibm.com/analytics/machine-learning>.

109 Bertalanffy 'An Outline of' (n 105) 142.

110 Ibid, 146.

111 Ibid, 148.

112 Interview with Joelle Pineau, Associate Professor, School of Computer Science, McGill University (Montreal, Canada, 23 May 2017).

113 J Yosinski and others, 'Understanding Neural Networks Through Deep Visualisation' [2015] 150606579 ARXIV 1, 2.

disadvantage.'[114] The AIDMP uses decision-making criteria which appear neutral and which do not correspond to a protected characteristic but which privilege one group over another. For instance, consider the case of the HireVue software. The tone of voice does not directly correspond to sex, race, age or other protected characteristics. Yet, the use of these criteria places individuals who experience discrimination at work and who experience mental health difficulties at a disadvantage. There is research indicating that individuals who belong to underrepresented communities are at a higher risk of suffering from mental health difficulties due to experiencing discrimination and lack of acceptance by their communities.[115] As a result of experiencing discrimination at work, individuals who belong to groups which have historically experienced discrimination are likely to exhibit emotions that the AIDMP treats as indicators of poor performance. Examples include less smiling and having a sad tone of voice. Individuals with pre-existing mental health difficulties will be particularly affected. The privileging of certain modes of expression and being in the world over others functions in the same way as the construction of gender. In both cases a decision-maker with more resources and power imposes a definition on individuals regarding how they should define and express themselves in order to gain employment. The corporation replaces the role which men had historically in defining expectations for how women had to behave.

4 The Need to Think Differently About the Subject Matter of the Protection

One possible solution to enabling a legal rule to address situations where it is impossible to establish a causal link between the use of an input variable and the harmful effect is to define the subject matter of the protection in terms of safeguarding human diversity rather than in terms of drawing a distinction between individuals based on the possession of a protected characteristic. Richard Crisp defines diversity as the ability of individuals to define their identities in a multitude of ways.[116] The identities are complex and cannot be

114 Barocas and Selbst (n 79) 677.

115 B Ao, 'Black Trans Communities Suffer a Greater Mental-health Burden from Discrimination and Violence' *The Philadelphia Inquirer* (Philadelphia, 25 June 2020) <www.inquirer.com/health/black-transgender-trans-mental-health-therapy-20200625.html>.

116 R Crisp, 'Introduction' in R Crisp (ed), *The Psychology of Social and Cultural Diversity* (Blackwell Publishing, 2010) 1.

captured by reference to a particular characteristic, such as gender.[117] More broadly, diversity refers to the individuality of each person and to the infinite variability between individuals. Individuals should be able to develop complex identities and to express their identities in a way of their choosing. Alice Walker's image of flowers in a garden[118] is useful for capturing this conception of diversity. Alice Walker formulated womanism as a response to the failure of feminists to engage with the concerns of the women of colour.[119] According to Alice Walker, 'Womanist is to feminist as purple to lavender.'[120] Womanism treats humanity as one.[121] It regards people, communities and the world as interconnected.[122] Womanism opposes the idea of classifying human beings into categories and partitioning them into groups.[123] Such processes of categorisation and partitioning diminish the humanity of the individuals.[124] Ann Goodley arguably captures this notion of human diversity when describing each individual as a 'complex mosaic.'[125] The protection of human diversity necessitates that the subject matter of the prohibition of discrimination be defined by reference to two elements. First, the uniqueness and individuality of each person. Second, the human diversity should be protected as an intrinsic self-standing value. This entails protecting all human beings as a group.

5 How CEDAW Can Address Digital Discrimination

Attention will now turn to considering how CEDAW may respond to the phenomenon of digital discrimination. CEDAW can be interpreted progressively so as to define the subject matter of the protection in terms of safeguarding human diversity. A new definition of discrimination can be added to CEDAW to enable it to address digital discrimination. It is possible to interpret the provisions in a manner which enables the CEDAW to reflect the distinct manner in which the operation of the AIDMP brings about problematic outcomes.

117 Ibid.

118 A Walker, 'Womanist (1983)' in Layli Phillips (ed), *The Womanist Reader* (Routledge, 2006) 19.

119 PH Collins, 'Sisters and Brothers: Black Feminists on Womanism' in L Phillips (ed), *The Womanist Reader* (Routledge, 2006) 62.

120 Walker (n 118) 19.

121 L Maparyan, *The Womanist Idea* (Routledge, 2011) 21.

122 Ibid, 19.

123 Ibid, 21.

124 Ibid.

125 See interview in S Monro, 'Beyond Male and Female: Poststructuralism and the Spectrum of Gender' (2005) 8 International Journal of Transgenderism 3, 15.

5.1 *Reinterpreting the Subject Matter of the Protection*

Marsha Freeman, Christine Chinkin and Beate Rudolf maintain that the principle of diversity is essential for understanding the content and scope of CEDAW.[126] Rather than being an ancillary concept for understanding the scope and content of CEDAW,[127] the protection of human diversity is the subject matter of the treaty. The subject matter of the protection in CEDAW may be interpreted as encompassing the safeguarding of the uniqueness of each person and human diversity as such. This meaning of treaty text emerges when one interprets the term sex as having multiple dimensions. The CEDAW's provisions define discrimination by reference to the possession of a protected characteristic. Article 1 CEDAW defines the scope of the protection by reference to sex.[128] Article 31(1) of the Vienna Convention on the Law of Treaties 1969 states:

> A treaty shall be interpreted in good faith in accordance with the ordinary meaning to be given to the terms of the treaty in their context and in the light of its object and purpose.[129]

The context comprises the treaty's text, the preamble and the annexes.[130] The general recommendations of the CEDAW Committee indicate that it recognises that the ordinary meaning of the term sex encompasses multiple layers. In the General Recommendation 28 it interpreted Articles 1, 2(f) and 5(a) CEDAW as covering 'gender-based discrimination.'[131] The term sex refers to the biological characteristics of women.[132] The term gender refers to 'socially constructed identities, attributes and roles.'[133] The term sex can be interpreted as having more layers of meaning. It refers not only to the biological embodiment of individuals and to how society ascribes traits to particular biological embodiments, but additionally to how individuals psychologically experience themselves in their bodies. Interpreting the term sex in this manner is in accordance with the ordinary meaning of this word. People experience themselves not only through their biological embodiment but additionally through their emotional embodiment in their bodies. Since people have cognition,

126 Freeman, Chinkin and Rudolf (n 41) 240.
127 Ibid.
128 CEDAW, Art 1.
129 Vienna Convention on the Law of Treaties 1969 (adopted 23 May 1969, entered into force 27 January 1980), 1155 UNTS 331 Art 31(1).
130 Ibid Art 31(2).
131 General Recommendation No 28 (n 28) para 5.
132 Ibid.
133 Ibid.

emotions and wishes, it is artificial to draw a line between the physical body and the experience of inhabiting a body.

Although the CEDAW Committee never identified the psychological embodiment of a body as an element of sex, this interpretation is consistent with its approach to interpreting the treaty obligations. The CEDAW Committee recognised that there is a need to protect the ability of women to make their life choices freely. It required Uganda to adopt comprehensive anti-discrimination legislation that includes 'the prohibition of multiple forms of discrimination against women on all grounds, including on the grounds of sexual orientation and gender identity.'[134] By confirming that the prohibition of discrimination extends to adverse treatment based on how people define and express their identities, the CEDAW Committee acknowledged implicitly that the protection covers both the physical and the psychological embodiment of individuals. The interpretation of the term sex as encompassing the physical, psychological and social dimensions advances the purpose of CEDAW. The CEDAW's Preamble makes a reference to the affirmation of 'the dignity and worth of the human person.'[135] Paragraph 2 of the Preamble states that all individuals are 'born free and equal in dignity and rights.'[136] The purpose of the treaty of protecting human dignity is furthered by interpreting the term sex in Article 1 as including how individuals experience themselves psychologically and how they express themselves. When individuals experience exclusion from opportunities due to manifesting their individuality and living their life according to their personal preferences, their human dignity is violated.

The interpretation of the subject matter of the protection as referring to safeguarding human diversity is supported by reading Articles 1 and 5 CEDAW jointly. The definition of discrimination in Article 1 CEDAW prohibits 'any distinction, exclusion or restriction' made on the basis of sex which has as its purpose or impact the impairment or nullification of the enjoyment of human rights on an equal basis.[137] The term exclusion refers to patterns of belief and to social practices which deny individuals opportunities.[138] The term exclusion therefore covers situations where an organisation treats certain identities or personal preferences as more desirable and penalises people who do not adhere to such expectations. Since Article 1 prohibits excluding people from

134 UN Committee on the Elimination of Discrimination against Women, 'Concluding Observations of the Committee on the Elimination of Discrimination against Women: Uganda' (22 October 2010) UN Doc CEDAW/C/UGA/CO/7 para 44.

135 CEDAW, Preamble para 1.

136 Ibid, para 2.

137 Ibid, Art 1.

138 Cook and Cusack (n 48) 109–110.

opportunities based on how they define and express their identities, it is concerned with protecting human diversity as such. The obligations in Article 5 are closely linked with the prohibition of excluding individuals from opportunities on the basis of sex or gender identity. It is concerned with eliminating the root causes of exclusionary practices which stem from organisations determining how individuals should define and manifest their identity. Article 5 requires States to eliminate prejudices relating to the superiority of sexes by modifying social and cultural patterns of conduct.[139] Since Article 5 obliges States to eliminate social expectations of how a person of a particular gender identity is to behave, it protects the ability of individuals to define their identities and to manifest them without hindrance. Given that Articles 1 and 5 CEDAW require States to ensure that individuals do not exclude others from opportunities on the basis of ascribing particular attributes to them on the basis of their gender identity, the provisions are concerned with protecting human diversity as such. Articles 1 and 5 protect the ability of individuals to develop and to manifest distinct identities.

The interpretation of the subject matter of CEDAW as protecting human diversity and the individuality of individuals has far-reaching consequences for the context of the use of AIDMPs. When employers treat particular identities, preferences, lifestyle choices and modes of self-expression as being less worthy, they are drawing a distinction between individuals on the basis of gender identity. It is irrelevant that the applicants cannot trace the relationship between the decision and their physical embodiment. This is because there is no difference between excluding individuals based on the bodies they inhabit and based on how they inhabit their bodies psychologically. Furthermore, CEDAW applies to cases where the manner in which the AIDMP represents individuals in the model is not based on gender stereotypes but nevertheless excludes individuals from gaining employment. Attention will now turn to examining how existing tests prohibiting discrimination apply to the context of the AIDMP and how the gaps in the legal protection may be filled.

5.2 *The Prohibition of Direct Discrimination*

The following example will be used to contextualise the discussion regarding the adequacy of the prohibition of direct discrimination. The AIDMP uses data which the employer collected relating to the employee's off-duty conduct.[140] The AIDMP uses data relating to the employee's Facebook posts and Fitbit

139 CEDAW, Art 5.

140 Magid (n 94).

usage to predict successful performance.[141] The program treats the Facebook post about visiting a relative who had a heart attack as an indicator that there is a likelihood of an employee suffering a heart attack in the future. It uses the Fitbit data relating to the presence of periodic irregular heartbeat to conclude that the employee has a risk of suffering a heart attack. Based on this data the AIDMP denies an employee a promotion for a leadership position. Since the employee does not have a health condition at the point of the decision-making, the employee cannot argue that the decision violates the prohibition of direct discrimination on the basis of having a health condition. Neither is the employee treated differently due to having caring responsibilities for a sick relative. Yet, since the decision is based on predicting the employee's future health status, it is no different from a decision which is solely based on the employee's present health status. The type of harm the employee suffers in this case arguably differs from the phenomenon of 'proxy discrimination.'[142] The AIDMP does not base the decision on the information which reveals the employee's present illness.

In order for the employee to be protected in this scenario the prohibition of direct discrimination would need to encompass situations where an individual experiences differential treatment on the basis of traits, habits and conduct associated with the possession of a protected characteristic. There can be a temporal gap between the event, such as exhibiting irregular heartbeat, and the protected characteristic manifesting itself in the future. Sandra Wachter argues in the context of the European Convention on Human Rights that there is no difference between using a protected characteristic as a basis for the decision and information which has an 'affinity' with the possession of that characteristic.[143] Turning to the example at hand there is no basis for distinguishing between the possession of a protected characteristic and the pattern of activity associated with that characteristic. Individuals experience disadvantage in the same manner irrespective of whether they experience exclusion on the basis of possessing a protected characteristic or on the basis of exhibiting conduct which is related to their possession of a protected characteristic.

CEDAW may be interpreted dynamically to prohibit direct discrimination where the treatment is based on the conduct, pattern of activity, mode of self-expression and traits associated with possessing a protected characteristic.

141 Ibid; C Gohd, 'How the CIA is Using Artificial Intelligence to Collect Social Media Data' (*Futurism,* 2017) <https://futurism.com/how-the-cia-is-using-artificial-intelligence-to-collect-social-media-data>.

142 Prince and Schwarcz (n 7) 4.

143 Wachter (n 81) 37.

The purpose of the treaty to ensure equal treatment[144] will be frustrated if organisations can exclude individuals from opportunities based on characteristics associated with the possession of the protected characteristic. This interpretation is supported by interpreting Article 1 alongside Articles 2 and 3. The CEDAW Committee interpreted Article 2 in the General Recommendation 28 as obliging States to recognise in law that discrimination based on sex and gender is intertwined with 'other factors.'[145] Since the term 'other factors' is non-exhaustive, it is broad enough to include characteristics which are connected to the possession of a protected characteristic. Article 3 CEDAW is designed to enable the treaty to anticipate that new forms of discrimination may emerge.[146] It provides support for interpreting Article 1 in a manner which closes the gap in the legal protection. Consequently, Articles 1, 2 and 3 point to the fact that the prohibition of discrimination covers differential treatment on the basis of conduct, expression, pattern of activity and traits associated with the possession of a protected characteristic. This interpretation enables CEDAW to respond to the fact that the AIDMP produces harm in a distinct manner. This is in line with the approach of the CEDAW Committee to interpret the treaty in a progressive manner.[147]

5.3 *The Prohibition of Indirect Discrimination*

Philipp Hacker argues in the context of the European Union legislation that the use of decision-making factors to construct the AIDMP's model which appear neutral but which in fact correlate with the membership of a protected group can amount to indirect discrimination provided that the protected group is 'put at a specific disadvantage.'[148] The same holds if the decision is closely correlated to the membership of a protected group or if the decision puts the protected group at a 'specific disadvantage.'[149] The principle of indirect discrimination does not provide protection in all cases in the context of digital discrimination. This legal norm assumes that it is possible to identify a decision-making factor which can be linked to the creation of 'relative disadvantage'[150] for a group with a protected characteristic. It will be impossible in all cases to identify how decisions impose a 'relative disadvantage' on a group. The AIDMP processes data

144 CEDAW, Preamble para 2.
145 General Recommendation No 28 (n 28) para 18.
146 CEDAW, Art 3.
147 General Recommendation No 25 (n 36) para 3.
148 P Hacker, 'Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law' (2018) 55 CML Rev 1143, 1153.
149 Ibid.
150 UN Committee on Economic, Social and Cultural Rights (n 63) para 12.

based on detecting correlations between the data.[151] As has already been shown, each input may have varying degrees of correspondence to the possession of a protected characteristic. Since the AIDMP bases decisions on thousands of inputs,[152] the decision could be the result of taking into account different characteristics which are connected to the possession of a protected characteristic to varying degrees. To illustrate, the candidate's preference for viewing videos online about meditation could be related to practising Buddhism. The candidate's preference for particular music could have a degree of correspondence to ethnicity.[153] The candidate's tone of voice could be related to mood. In turn, the candidate's mood can correspond to whether the candidate can afford to see a psychotherapist and thus to the candidate's socio-economic background.

Yet, it may be impossible to detect a connection between the disadvantage and group membership. It is possible to envisage situations where the AIDMP uses inputs which have a degree of correspondence to protected characteristics but where the decision does not impose a 'relative disadvantage' on a particular group. The AIDMP places individuals into groups based on similarity[154] and uses group data to make the prediction about an applicant.[155] Thus, it is not concerned with evaluating the individual.[156] The fact that the AIDMP uses data about the group to produce a prediction about an individual[157] is significant. The connection which each point of data in the cluster of individuals whom the system treats as being similar has to the possession of a protected characteristic becomes obscured. This stems from the fact that the grouping of individuals as similar is related to the predictive capacity of the data rather than to the possession of a particular protected characteristic. Another relevant factor is that the AIDMP employs a distinct logic. The existence of a correlation does not indicate causation.[158] For instance, there was an instance when an algorithm learned how to distinguish between polar and brown bears based on the presence of snow rather than based on recognising the unique features of the

151 Provost and Fawcett (n 87) 25.
152 Magid (n 94).
153 Wachter (n 81) 39.
154 Provost and Fawcett (n 87) 24.
155 Ibid, 107.
156 L Taylor, 'On the Presumption of Innocence in Data-Driven Government. Are We Asking the Right Question?' in I Baraliuc and others (eds), *Being Profiled: Cogitas Ergo Sum* (Amsterdam University Press, 2018) 105.
157 Provost and Fawcett (n 87) 107.
158 Virginia Eubanks, *Automating Inequality: How High-tech Tools Profile, Police and Punish the Poor* (St Martin's Press, 2018) 144.

polar bear.[159] Since the AIDMP may use various strategies to achieve predictive capacity which is not based on detecting a causal relationship, it is difficult to establish a relationship between the use of an input variable and the impact on a specific group. Another relevant factor is that emergent effects occur as a result of different inputs, stages in the development process and stages within the decision-making process of the AIDMP interacting with one another. This aspect makes it harder to trace the connection between each input and the decision outcome.

Although there is a difficulty with detecting the connection between the use of the AIDMP and the negative impact on a group, the disadvantage which individuals experience due to facing social barriers will impact on how the AIDMP groups them together. For instance, an applicant could fail to get a particular work experience on the ground of experiencing discriminatory attitudes based on race or sexual orientation. The data serves as a proxy for the intersectional manner in which applicants experience inequality.[160] Even if the AIDMP ignores the applicants' race and sexual orientation, it may group such applicants together due to them lacking the requisite work experience. The data relating to the work experience is a proxy for the intersectional manner in which individuals experience discrimination. The systems theory suggests that this data will interact with other data when the AIDMP places individuals into groups based on similarity. The emergent effects which occur as a result of these interactions will manifest themselves when the AIDMP uses group data to predict the applicant's performance.

To illustrate, a student who received top grades in her Spanish course found out that the artificial intelligence software predicted that she would fail the International Baccalaureate examination.[161] Meredith Broussard comments that this was likely to be due to the algorithm using the historical performance of the students at the same school as one of the inputs.[162] The students mostly came from low-income families and many of them were racialised.[163] This example illustrates that the operation of the AIDMP has the effect of compounding the structural inequality which an individual experiences. This

159 J Khalili, 'Artificial Intelligence is Hopelessly Biased-and That's How It Will Stay' (*Future US,* 24 May 2020)<www.techradar.com/news/playing-god-why-artificial-intelligence-is-hopelessly-biased-and-always-will-be>.

160 Prince and Schwarcz (n 7) 7–8.

161 M Broussard, 'When Algorithms Give Real Students Imaginary Grades' *The New York Times* (New York, 8 September 2020) <www.nytimes.com/2020/09/08/opinion/international-baccalaureate-algorithm-grades.html>.

162 Ibid.

163 Ibid.

discussion corroborates the assertion of Virginia Eubanks that the operation of the AIDMPs deepens inequality.[164]

The issue becomes how individuals can be protected given the fact that it is not possible to show a causal link between an input variable and the harmful impact on an identifiable group. One possible way of addressing this issue is to introduce an additional test for discrimination under Article 3 CEDAW. Doing so would enable States to fulfil their obligation under Article 3 to take measures in all fields "to ensure the full development and advancement of women" for the purpose of guaranteeing them enjoyment of human rights on an equal basis with men.[165] This is because Article 3 is residual in character.[166] The drafters intended that Article 3 should enable CEDAW to apply to new forms of discrimination which may emerge.[167] The addition of new tests for discrimination advances the intent of the drafters that CEDAW be able to respond to new developments regarding how individuals experience discrimination.

The proposed test overcomes the challenge of the difficulty of detecting the emergent effects by focusing on how the design and operation of the AIDMP constructs relationships. Martha Albertson Fineman's vulnerability theory serves as a conceptual framework for formulating the test.[168] The vulnerability theory stipulates that individuals are situated in different economic, social, cultural and institutional relationships.[169] These relationships cannot be categorised as being either private or public.[170] The position which individuals occupy within these relationships influences their opportunities.[171] These institutions operate in conjunction and determine the individuals' resilience.[172] They shape how easily an individual can recover from life's setbacks and take advantage of opportunities.[173] A test of discrimination based on the vulnerability theory defines discrimination by reference to

164 Eubanks (n 158) 204.
165 CEDAW, Art. 3.
166 General Recommendation No 25 (n 36) para 6.
167 General Recommendation No 28 (n 28) para 8.
168 MA Fineman, 'Equality and Difference – the Restrained State' (2015) 66 Ala LR 609, 614 (Equality and Difference).
169 MA Fineman, 'Equality, Autonomy and the Vulnerable Subject in Law and Politics' in A Grear and MA Fineman (eds), *Vulnerability: Reflections on a New Ethical Foundation for Law and Politics* (Ashgate Publishing 2013) 22 (Equality, Autonomy).
170 MA Fineman, 'Injury in the Unresponsive State: Writing the Vulnerable Subject into Neo-Liberal Legal Culture' in A Bloom, DM Engel and M McCann (eds), *Injury and Injustice: The Cultural Politics of Harm and Redress* (Cambridge University Press 2018) 19.
171 Fineman, 'Equality, Autonomy' (n 169) 23.
172 Ibid, 22.
173 Fineman, 'Equality and Difference' (n 168) 622–23.

how the software developers construct relationships when they design the AIDMP.[174] Specifically, it is relevant how the developers construct the relationship between themselves and the subjects of the decision-making process, between the employer and the subjects of the decision-making as well as between individuals whose data is present in the mathematical model. In the course of applying the proposed test one evaluates how an unequally situated candidate in relationships affects that candidate's ability to obtain a positive decision. One also needs to consider how each decision the developer makes during the construction of the AIDMP bears on the manner in which the candidate is positioned in relationships. Third, one should assess how different stages involved in the execution of the decision-making process bear on the position of the candidate in a set of relationships. Fourth, how the decisions the AIDMP generates change the situation of each candidate in relationships over the long term should be examined.

Under the proposed test the use of the AIDMP is unlawful whenever it constructs a relationship between the developer, the employer and the subjects of the decision-making in a manner which:

i) is unequal OR
ii) impedes the ability of candidates to define and express their identities OR
iii) impedes the ability of the applicants to access opportunities OR
iv) places the candidate at a disadvantage OR
v) compounds a pre-existing disadvantage by making the harm caused by prior injustice worse[175] OR
vi) does not account for how structural inequality in society affects the ability of the applicant to access the opportunity[176]

To illustrate, this definition covers a situation where the operation of the AIDMP denies an individual access to employment due to creating an unequal relationship between the applicants. The unequal relationship could be due to the AIDMP treating certain identities, activities or lifestyle preferences as having less value. It includes denying employment to applicants on the basis of their pronunciation and other modes of expression.

174 Fineman, 'Equality, Autonomy' (n 169) 23.
175 D Hellman, 'Indirect Discrimination and the Duty to Avoid Compounding Injustice' in H Collins and T Khaitan (Hart Publishing, 2018) 113.
176 Khaitan (n 25) 31.

5.4 *The Principle of Substantive Equality*

The obligation to achieve substantive equality may be interpreted purposively in order to address the context of digital discrimination. This obligation arises from the obligations in Article 4(1) CEDAW.[177] The principle of substantive equality requires the employers to use a recruitment and promotion process which results in the same number of men and women occupying relevant positions.[178] This principle obliges the developers to calibrate the AIDMP in a manner which enables the employer to achieve gender parity in the workforce. It is necessary to interpret the principle of substantive equality jointly with Article 2.[179] The CEDAW Committee explained in the General Recommendation 28 that duty bearers must recognise the intersectional manner in which individuals experience discrimination.[180] The joint application of the principle of substantive equality with the principle of intersectionality requires developers to program the selection procedure within the AIDMP with a view to recognising the additional barriers which individuals with intersectional identities face. The developers should program the AIDMP in a manner which enables the organisation to have a composition of the labour force which reflects the prevalence of individuals with intersecting identities in society.

In practice, it is challenging to implement the duty to achieve de facto equality due to the manner in which the AIDMP operates. The operation of the AIDMP has the potential to result in individuals with protected characteristics experiencing 'systemic relative disadvantage.'[181] The logic underlying the operation of the AIDMP is conducive to conferring positive decisions on candidates with a protected characteristic who experience relative advantage in relation to other candidates with a protected characteristic. This stems from the fact that artificial intelligence software operates based on the logic of optimisation.[182]

Optimisation involves choosing criteria for the decision-making which allow the end goal to be achieved as well as possible while satisfying the constraints.[183] It involves identifying characteristics which are linked to the

177 General Recommendation No 25 (n 36) para 8.

178 Ibid, paras 8–9.

179 Ibid, para 6.

180 General Recommendation No 28 (n 28) para 18.

181 Barocas and Selbst (n 79) 677.

182 A Badar, BS Umre and AS Junghare, 'Study of Artificial Intelligence Optimisation Techniques applied to Active Power Loss Minimisation' (2014) IOSR Journal of Electrical and Electronics Engineering 39, 39.

183 M Koopialipoor and A Noorbakhsh, 'Applications of Artificial Intelligence Techniques in Optimising Drilling' (IntechOpen Limited, 2020) <www.intechopen.com/books/

variable to be predicted, such as good performance.[184] Moreover, it entails selecting individuals whose scores on that variable are the highest.[185] The optimisation logic underlying the AIDMP[186] will result in the selection of women who most closely fit the profile. Yet, the greater the disadvantage which the woman experiences, the more difficult it will be for her to fit a profile which is based on the logic of optimisation. The AIDMP will select women who have the greatest access to various resources. The extent to which women can take advantage of opportunities is influenced by their position in relationships.[187]

For instance, socio-economic disadvantage impairs the ability of the individuals to secure places at elite universities.[188] Racialised women who possess relative advantage over other racialised female candidates due to having access to well-funded schools or due to not having to work alongside studying are likely to find it easier to match the profile of an effective employee. Since the AIDMP bases the decisions by reference to how closely a candidate matches the profile of optimum performance, it will favour hiring racialised women who have a greater degree of advantage in comparison to other racialised candidates with intersectional identities. Since the AIDMP uses numerical calculations to determine whether there is adherence to the principle of substantive equality, it will not account for how each applicant experiences inequality in society. As a result, the AIDMP has the potential to select applicants with a protected characteristic in a manner which excludes candidates who experience disadvantage to a greater degree.

The provision of disaggregated data about the number of hired women by reference to the possession of more than one protected characteristic does not fully address this issue. This stems from the fact that disadvantage is a matter of degree. A contextual analysis is needed to evaluate the relative degree of disadvantage which a candidate experiences. The use of numerical analysis which employs aggregated data does not permit such a contextual analysis. The following example illustrates the difficulty. The AIDMP only recognises information which the developers program it to detect.[189] Consequently, the

---

emerging-trends-in-mechatronics/applications-of-artificial-intelligence-techniques-in-optimizing-drilling>.

184 Barocas and Selbst (79) 679.

185 Ibid.

186 Badar, Umre and Junghare (n 182) 39.

187 Fineman, 'Equality, Autonomy' (n 169) 22.

188 LR Pruitt, 'Who's Afraid of White Class Migrants? On Denial, Discrediting, and Disdain (and Toward a Richer Conception of Diversity)' (2015) 31 Colum J Gender & L 196, 225.

189 B Holzer, 'The Best Algorithms Struggle to Recognise Black Faces Equally' *Wired* (Boone, 22 July 2019) <www.wired.com/story/best-algorithms-struggle-recognize-black-faces-equally>.

AIDMP will score applicants lower who do not communicate their achievements in a format it recognises. Some candidates could receive a higher score because they had access to mentoring or to training opportunities regarding how to present their application in a format which increases the likelihood of the AIDMP recognising their achievements. Individuals who live in communities which organise themselves based on the principle of mutual support may have access to more information than individuals who grew up in the dominant individualistic culture prevalent in Western societies. Because the disaggregated data relates to a pool of candidates, the data does not reveal whether the candidates with more than one protected characteristic had equal opportunities. This points to the fact that disaggregated data based on the possession of a protected characteristic does not provide sufficient information about whether the AIDMP provided all applicants with equal opportunities. Moreover, the use of disaggregated data as a means of ensuring equal access to opportunities ignores the fact that individuals lack equal access to computers and to an internet connection. The digital divide persists even in countries such as the United States. Specifically, tens of millions of American citizens lacked access to home computers and affordable, high-speed internet in the year 2020.[190]

The difficulty with using disaggregated data to provide legal recognition of the fact that candidates experience discrimination in an intersectional manner in the context of the application of the principle of substantive equality stems from the fact that advantage is a matter of degree. The operation of the AIDMP is not sensitive to the variation in relative disadvantage among the applicants. The purpose of algorithms is to make the decision-making uniform.[191] The aggregated information about the number of women with protected characteristics whom the organisation hired erases the experience of relative disadvantage. This is because the AIDMP does not take into account the contextual situation of each applicant. Linnet Taylor explains that the aggregation of data 'conceals' 'meaning.'[192] This stems from the fact that 'the individual is neither identifiable nor analytically important in a dataset.'[193] The problem is compounded by the fact that it is impossible to express the relative disadvantage which each individual experiences using numbers. John Powell

190 J Simama, 'It's 2020. Why Is the Digital Divide Still with Us?' *eRepublic* (Folsom, 8 September 2020) <www.governing.com/now/Its-2020-Why-Is-the-Digital-Divide-Still-with-Us.html>.

191 A Narayan, 'FAT* 2018 Translation Tutorial: 21 Definitions of Fairness and Their Politics' (Youtube NL, 2018) < https://www.youtube.com/watch?v=wqamrPkF5kk > 32:15–33:23.

192 Taylor (n 156) 105.

193 Ibid.

explains that social assumptions[194] and institutional arrangements create disparities between groups.[195] They operate in an intersectional manner[196] and in a non-linear fashion.[197] Since the impact of social arrangements on individuals is intangible and cumulative in nature, it is impossible to measure the degree of disadvantage which an applicant experiences. This results in the application of the principle of substantive equality failing to remedy fully the situation which Kimberle Crenshaw described two decades ago. She described dominant groups as leaving a small hatch through which they let people in who were closest to their position.[198]

James Allen articulates a related concern. He uses the term algorithmic redlining to refer to a set of instructions which execute procedures limiting the access of people of colour to housing and financial services.[199] The coding of the principle of substantive equality into the software leads to a situation where the most disadvantaged persons are excluded without it being obvious how the line around the excluded group gets to be drawn. In order to comply with the principle of substantive equality organisations should use human decision-makers to select employees. Human decision-makers will be able to better take into account the intersectional identities of the candidates through their knowledge of social life and abstract thinking skills. The human decision-maker would need to evaluate how the intersectional identities of the candidates influenced the degree of disadvantage which they experience. The candidates should have an opportunity to explain the value of their non-traditional work experience for the company and how their transferrable skills will enable them to perform the job duties. The principle of substantive equality obliges the decision-maker to select candidates with a view to reflecting the prevalence of individuals with intersectional identities in the population.

5.5 *The Principle of Transformative Equality*

The principle of transformative equality places a duty on public and private actors to transform 'opportunities, institutions and systems.'[200] It is

194 JA Powell, 'Understanding Structural Racialisation' (2013) 47 Clearinghouse Rev 146, 147.
195 Ibid, 146.
196 ET Achiume, 'Beyond Prejudice: Structural Xenophobic Discrimination Against Refugees' (2016) 45 Georgetown J IL 323, 327.
197 Powell (n 194) 151.
198 K Crenshaw, 'Demarginalising the Intersection of Race and Sex: a Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics' (1989) 1 U Chi Legal F 139, 151.
199 JA Allen, 'The Colour of Algorithms: An Analysis and Proposed Research Agenda for Deterring Algorithmic Redlining' (2019) 46 Fordham Urb LJ 219, 222.
200 General Recommendation No 25 (n 36) para 10.

appropriate to analyse digital discrimination through the lens of the principle of transformative equality. The AIDMP is both an institution[201] and a technical system. The AIDMP is an institution because its operation changes how individuals are positioned in relation to other individuals and institutions.[202] In the course of detecting correlations between the data[203] the AIDMP creates a web of relations between individuals. Moreover, the AIDMP operates as an institution because it determines the access of individuals to opportunities, such as employment.

Additionally, the principle of transformative equality is applicable to how the AIDMP represents applicants within the mathematical model and displays the decision output. The CEDAW Committee explained in the General Recommendation 25 that prevailing gender relations and gender-based stereotypes are the root causes of discrimination.[204] The obligation to address the root causes of discrimination in Article 4(1)[205] together with the obligation in Article 5(a) CEDAW to remedy gender-based stereotypes[206] require organisations to use representations of individuals during the decision-making process which do not create new stereotypes or solidify existing stereotypes.

Suzana Dalul notes that the operation of the AIDMPs can generate stereotypical representations of groups with protected characteristics in the model.[207] The Office of the United Nations High Commissioner for Human Rights defines a stereotype as a 'preconception about attributes or characteristics that are or ought to be possessed by members of a particular social group or the roles that are or should be performed' by members of that group.[208] One of the situations when the model generates stereotypical representations is when the developers use biased historical data.[209] Vicente Ordóñez found that the artificial intelligence system generated a model which had a strong association between being a woman and doing cooking.[210] This was due to developers training the system on photographs which depicted women as cooking

201 Krupiy (n 74) 2.
202 Ibid.
203 Provost and Fawcett (n 87) 25.
204 General Recommendation No 25 (n 36) para 7.
205 Ibid.
206 CEDAW, Art 5(a).
207 S Dalul, 'AI is not Neutral, It is Just as Biased as Humans' (*AndroidPIT,* 2019) <www.nextpit.com/ai-is-not-neutral-biased-as-humans>.
208 Office of the United Nations High Commissioner, 'Gender Stereotypes and Stereotyping and Women's Rights' (September 2014).
209 Dalul (n 207).
210 T Simonite, 'Machines Taught by Photos Learn a Sexist View of Women' *Wired* (Boone, 2017) <www.wired.com/story/machines-taught-by-photos-learn-a-sexist-view-of-women>.

more frequently than men.[211] The stereotypical representation of women in the model has an impact on how the AIDMP processes the data and whether it confers a positive decision on the applicant. For instance, if the AIDMP associates the profession of a computer scientist with being a male then it will select male candidates for the position.[212] More broadly, Rachel Adams and Nora Ni Loideain maintain that the gendered representations of women through technology cause societal harm by disadvantaging them as a group.[213] The principle of transformative equality prohibits the developers from selecting the data and labelling the data in a manner which perpetuates stereotypes. Similarly, it prohibits the use and labelling of data which results in the AIDMP detecting correlations that replicate stereotypes. This is because the principle of transformative equality requires private actors to change systems and institutions 'so that they are no longer grounded in historically determined male paradigms of power and life patterns.'[214] When the developers select, label and process data in a manner which replicates stereotypes they construct institutions which replicate male paradigms of power. Moreover, they breach the duties to remedy the root causes of inequality[215] and to eliminate practices which ascribe stereotypical roles based on gender.[216]

What is more, the process underlying the operation of the AIDMP is conducive to amplifying existing stereotypes and to creating new stereotypes. The cognitive processes of 'excessive' categorisation and oversimplification play a role in human beings developing stereotypes.[217] The AIDMP mirrors the cognitive processes involved in stereotyping. It categorises and oversimplifies the environment in the course of operating. The AIDMP relies on the creation of categories for operating. The developers convert qualitative data into quantitative format in order to make it usable for the AIMDP.[218] Categories are needed to designate what it is that numerical values represent. According to Foster

211 Ibid.

212 Y Cooper, 'Amazon Ditched AI Recruiting Tool That Favored Men For Technical Jobs' *The Guardian* (London, 11 October 2018) <www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine>.

213 R Adams and NN Loideain, 'Addressing Indirect Discrimination and Gender Stereotypes in AI Virtual Personal Assistants: The Role of International Human Rights Law' (2019) 8 CILJ 241, 242.

214 General Recommendation No 25 (n 36) para 10.

215 Ibid.

216 CEDAW, Art 5(a).

217 PL Bellodi, 'The General Practitioner and the Surgeon: Stereotypes and Medical Specialties' (2004) 59 Revista do Hospital das Clínicas 15, 15.

218 Provost and Fawcett (n 87) 30.

Provost and Tom Fawcett, the model abstracts complexity from real life.[219] This is because the model focuses on the indicators correlating with the variable to be predicted.[220] Additionally, the model oversimplifies reality because it abstracts individuals from the relationships in which they are embedded. When the AIDMP groups individuals based on similarity,[221] it creates new relationships between individuals.[222] These relationships are based on similarity between their data rather than on the actual relationships in which they are embedded. For instance, the model of the AIDMP does not account for the fact that the developer occupies a position of power in relation to the subject of the decision-making. Angelina Fisher and Thomas Streinz coined the term 'power to datafy' to reflect inequality in

> the power to decide what kind of data is being generated and in what form or format, how and where it is amassed and used, by whom, for what purpose, and for whose benefit.[223]

The discretionary choices which a developer makes impact on the degree to which the applicant will adhere to the construct of an effective employee.[224] A relevant consideration is that the AIDMP classifies individuals when it generates groupings of data based on the presence of correlations between the data. The processes of oversimplification for the purpose of categorising individuals and assigning individuals to a group resembles the cognitive process involved in stereotyping. Thus, the decision-making procedure within the AIDMP has parallels with the processes in the human brain associated with forming stereotypes.

The representation of individuals in terms of discrete characteristics within the AIDMP perpetuates a simplified understanding of individuals with protected characteristics. The same holds for when the AIDMP provides an explanation of how it produced a decision. This is because the AIDMP will provide information on how closely the scores an applicant received on different metrics correspond to the construct of an ideal candidate. Marshall McLuhan argues that technology has effects on human awareness.[225] The processes of

219 Ibid 6.
220 Ibid.
221 Ibid, 24.
222 Krupiy (n 74) 12.
223 A Fisher and T Streinz, 'Confronting Data Inequality' (2021) 2021/1 New York University School of Law International Law and Justice Working Papers 2, 4.
224 Barocas and Selbst (n 79) 678.
225 M McLuhan, *Understanding Media: Extensions of Man* (Gingko Press, 2013) 20.

oversimplification, categorisation and classification inherent to the operation of the AIDMP will interplay with the processes in the human brain involved in the production of stereotypes. This interplay with existing mechanisms of simplification in the human brain will entrench and amplify existing stereotypes. Individuals will come to have particular associations with the protected characteristics and group identities. Furthermore, the manner in which the AIDMP influences human awareness can result in individuals developing new stereotypes. Individuals can develop new associations between a protected characteristic and the possession of an attribute in the course of interpreting decision outputs. The same is the case for affiliation or association with a group.

The principle of transformative equality prohibits representing individuals in a simplified manner in the model of the AIDMP. It prohibits basing decisions on simplified representations of individuals with protected characteristics in the model. Moreover, it prohibits the generation of simplified accounts during the stage of the AIDMP providing an explanation for the reason for the decision. The CEDAW Committee in the General Recommendation 3 interpreted Article 5(1) CEDAW as requiring the elimination of 'stereotyped conceptions of women' which 'owing to sociocultural factors' perpetuate discrimination.[226] Since the way in which the AIDMP represents individuals in a mathematical model and operates parallels the mechanisms involved in the formation of stereotypes, Article 5(1) CEDAW prohibits full automation of the decision-making.

The logic underlying the AIDMP creates difficulty for organisations to comply with the duty inherent to the principle of transformative equality to remedy the structural causes of inequality. The processes underlying the operation of the AIDMP render it an institution[227] which creates a hierarchical relationship between individuals in the course of operating. The expression of attributes using quantitative metrics necessarily gives rise to a hierarchical relationship between groups. This stems from the fact that the decision-making process collapses the variability of human beings and their expression into designations of good and bad performance. The subjective choices involved in determining what is a desirable quality for an employee can place one group at a disadvantage.[228] For instance, culture influences how individuals express themselves while communicating. African American individuals gaze more when speaking than when listening in comparison to white individuals.[229]

226 UN Committee on the Elimination of Discrimination against Women (6th Session), 'General Recommendation No 3: Education and Public Information Campaigns' (1987) UN Doc A/42/38.

227 Krupiy (n 74) 2.

228 Barocas and Selbst (n 79) 678.

229 Hogg and Vaughan (n 101) ch 15.

A white speaker may interpret an African American's lack of eye contact as rudeness and vice versa.[230] Since the same behaviour can have different meanings within two cultures, and since the AIDMP designates one mode of bodily expression as superior, it necessarily advantages one group over another. The developer's decision whether a lot of eye contact is linked to high or low performance will determine whether the operation of the AIDMP favours white or African-American applicants.

The operation of the AIDMP contravenes CEDAW whenever it creates a hierarchical relationship between applicants and the dominant group and as a result prevents them from accessing opportunities. The principle of transformative equality which the CEDAW Committee deduced from Article 4 requires developers to change the institutions so that they do not reflect 'historically determined male paradigms of power and life patterns.'[231] The CEDAW Committee required Fiji to change laws with a view to remedying the hierarchical relationships by recognising women as heads of households.[232] The reference to the 'male paradigms of power' in the General Recommendation 25[233] should be interpreted broadly to encompass the creation of all hierarchical relationships. The reference to the life patterns of men should be interpreted as encompassing the life patterns of all dominant groups. This interpretation furthers the purpose of the treaty to impose obligations on States in order to end discrimination 'in all its forms and manifestations.'[234] It is difficult to see how duty holders could transform institutions with a view to remedying the causes of discrimination[235] if they were to only focus on unequal relationships stemming from patriarchy. Interpreting the principle of transformative equality as covering all hierarchical relationships is supported by reading Article 4 in conjunction with the duty to prohibit gender-based discrimination.[236] Because in defining discrimination Article 1 CEDAW uses the language of 'making a distinction, exclusion or restriction' on the basis of sex which has the effect of impairing the enjoyment of human rights on an equal basis,[237] it is concerned with the impact on the applicant rather than with the reason why the decision-maker was able to exclude the applicant.

230 Ibid.
231 General Recommendation No 25 (n 36) para 10.
232 UN Committee on the Elimination of Discrimination against Women, 'Concluding Comments of the Committee on the Elimination of Discrimination against Women: Fiji' (14 January-1 February 2002) Supp No 38 UN Doc A/57/38 para 55.
233 General Recommendation No 25 (n 36) para 10.
234 CEDAW Preamble para 15.
235 General Recommendation No 28 (n 28) para 5.
236 Ibid.
237 CEDAW, Art 1.

Additional support of the interpretation of the scope of the obligations as covering the production of inequality through the creation of a hierarchical relationship characterised by an imbalance of power is found in how the CEDAW Committee interprets the term gender. In defining the term gender in the General Recommendation 28 the CEDAW Committee recognised the roles which the unequal distribution of power and hierarchical relationships play in causing inequality. According to the CEDAW Committee the term gender encompasses social constructs which give rise to a hierarchical relationship between women and men.[238] The social constructs operate to distribute power unequally and to disadvantage women.[239] Since hierarchical relationships disadvantage individuals through distributing power unequally, in order to achieve compliance with the duty to remedy the causes of discrimination[240] the duty bearers should modify all social arrangements which produce inequality in this manner. Further support for this interpretation is found in the General Recommendation 27. Since the CEDAW Committee noted that social hierarchies play a role in the inequitable distribution of aid resources,[241] it recognised that what matters is the role of the hierarchical relationship in producing inequality rather than the source of the hierarchical relationship. It is irrelevant whether the social hierarchy stems from socio-economic inequality, the gender of the decision-maker or other reasons.

The logic of optimisation underlying the operation of the AIDMP[242] violates the principle of transformative equality due to enacting a hierarchical relationship between individuals and due to penalising applicants who do not exhibit the life patterns of the dominant groups. The hierarchy manifests itself in the decision-making process being structured around the life pattern of dominant groups. This results in an imbalance of power and access to opportunities. Anja Bechmann argues that AIDMPS harm underrepresented groups by subjecting them to 'normalisation logics.'[243] They select individuals whose data most closely resembles that of the majority.[244] The AIDMP penalises individuals with protected characteristics due to relying on the logic of

238 General Recommendation No 28 (n 28) para 5.

239 Ibid.

240 Ibid.

241 UN Committee on the Elimination of Discrimination against Women, 'General Recommendation No 27 on Older Women and Protection of their Human Rights' (16 December 2010) UN Doc CEDAW/C/GC/27 para 25.

242 Badar, Umre and Junghare (n 182).

243 A Bechmann, 'Data as Humans' in RF Jorgensen (ed), *Human Rights in the Age of Platforms* (The MIT Press, 2019) 88.

244 Ibid, 86.

optimisation. It excludes individuals from employment opportunities by creating a narrow definition of a good employee. This stems from the fact that the logic of optimisation entails using a fixed number of measurable parameters as selection criteria. Furthermore, the requirement to attain the highest score on each parameter creates exclusion. There are many reasons why individuals may not be able to map their skills and contributions onto quantifiable parameters. An individual may have highly developed analytic skills but may have a lower number of written outputs due to having dyslexia. The AIDMP which evaluates employees based on continuous monitoring is likely to designate employees who require the taking of frequent breaks as lacking motivation or as unsuitable. There is a relationship between productivity and the number of breaks which an individual takes. Similarly, employees whose speed of working is influenced by side effects from medication will be disadvantaged. Given that each individual is unique and the tremendous variability between individual circumstances, there always remains a possibility that the AIDMP does not recognise the individual's capabilities. Even if companies were to deliberately construct a model which reflected human diversity, it would take a 'tremendous effort' to gather data to train the machine which represented the diverse range of disabilities individuals have.[245]

A potential limitation of CEDAW stems from how the CEDAW Committee has approached interpreting the treaty. The CEDAW Committee treats inequality as stemming from an antagonistic relationship between groups. For this reason, the more the structures of inequality deviate from the model of male oppression, the more difficult it becomes to apply CEDAW. For instance, since the CEDAW Committee focuses on the male oppressing the woman, it does not expressly engage with a situation where women treat other women who possess additional protected characteristics poorly. Francisca de la Riva-Holly describes how women from affluent backgrounds in Mexico treat women they employ to do domestic work.[246] The affluent women do not treat the domestic workers with respect or equality.[247] They take all the credit for any progress in the socio-economic position of the family which may result from the domestic worker's hard work.[248] The domestic worker can never be sufficiently grateful for being hired.[249] The fact that the CEDAW Committee does not engage in depth with different causes of inequality results in its toolbox being limited.

245 Engler (n 4).
246 Riva-Holly (n 59) 287.
247 Ibid.
248 Ibid.
249 Ibid.

A possible solution to this problem is for the CEDAW Committee to draw on a wide range of theories of how societies produce inequality in order to formulate new concepts relating to discrimination. This approach is supported by interpreting Articles 2 and 3 CEDAW jointly. Article 2(e) CEDAW obliges duty bearers to 'take all appropriate measures to eliminate discrimination against women by any person, organisation or enterprise.'[250] The fact that Article 2 uses the generic term discrimination indicates that it is meant to capture all social mechanisms involved in discrimination. Article 2 should be read in light of Article 3. Since States adopted Article 3 to cover unanticipated ways in which individuals may experience discrimination,[251] it supports the conclusion that the CEDAW Committee should develop new concepts of discrimination to reflect new knowledge of the mechanisms involved in producing inequality.

6 Conclusion

Digital discrimination can be difficult to detect because the mechanisms through which the operation of the AIDMP produces exclusion is diffuse and distributed. It is not always apparent how the harmful result occurs. Individuals can experience exclusion without the possibility of establishing a clear link between an input which the AIDMP uses, the decision-making process and the decision output. The same is the case for groups comprising individuals with intersectional identities. Even if computer scientists manage to make the workings of the AIDMP visible, this will not address the issue of the opaqueness. This stems from the fact that the effects are necessarily emergent and complex. The impact is greater than the sum of its parts. In order for CEDAW to address digital discrimination the subject matter of the protection should be reconceptualised as protecting the individuality of human beings and human diversity. For the large part this reframing can be achieved by interpreting CEDAW teleologically.

In order for the prohibition of direct discrimination to be relevant for the digital context it should be interpreted to encompass treating individuals differently on the basis of conduct, expression, pattern of activity and traits which have a degree of relationship to the possession of a protected characteristic. A new test should be developed to complement the prohibition of indirect discrimination. The test should focus on how the AIDMP constructs relationships between the subjects of the decision-making, the developer and

250 CEDAW, Art 2(e).

251 General Recommendation No 28 (n 28) para 8.

the employer. The principle of substantive equality provides sufficient protection provided that it is interpreted as requiring human decision-making. The principle of transformative equality is well-suited for protecting individuals in the digital context under two conditions. First, there is a need to broaden the understanding of the mechanisms involved in the production of inequality to all hierarchical relationships. Second, there is a need to use a broader range of theories of how societies produce inequality to understand the causes of inequality.