Advance Edited Version

Distr.: General
18 June 2020

Original: English

Human Rights Council
Forty-fourth session
15 June–3 July 2020
Agenda item 9
Racism, racial discrimination, xenophobia and related
forms of intolerance, follow-up to and implementation
of the Durban Declaration and Programme of Action

Racial discrimination and emerging digital technologies: a human rights analysis

Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance*

Summary

In the present report, the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance, E. Tendayi Achiume, analyses different forms of racial discrimination in the design and use of emerging digital technologies, including the structural and institutional dimensions of this discrimination. She also outlines the human rights obligations of States and the responsibility of corporations to combat this discrimination.

---

* The present report was submitted after the deadline so as to include the most recent information.

I. Introduction

1.   Emerging digital technologies have fundamentally altered the way we live our lives, and as such their human rights impact has been the subject of important analyses by the special procedures of the Human Rights Council.[1] Existing reports address how these technologies affect a broad spectrum of human rights, including the rights to freedom of opinion and expression, freedom of peaceful assembly and of association and the human rights of those subject to extreme poverty. The United Nations High Commissioner for Human Rights has contributed analysis of emerging digital technologies and the right to privacy.[2] In the present report, the Special Rapporteur aims to advance analogously robust analysis at the intersection of emerging digital technologies and racial equality and non-discrimination principles under international human rights law.

2.   The scope of the report is racism, intolerance, discrimination and other forms of harmful exclusion and differentiation on the basis of race, colour, descent, or national or ethnic origin, in keeping with the International Convention on the Elimination of All Forms of Racial Discrimination. This includes discrimination against indigenous peoples. In the present report, the Special Rapporteur urges an equality-based approach to human rights governance of emerging digital technologies. This requires moving beyond "colour-blind" or "race neutral" strategies.[3] A colour-blind analysis of legal, social, economic and political conditions commits to an even-handedness that entails avoiding explicit racial or ethnic analysis in favour of treating all individuals and groups the same, even if these individuals and groups are differently situated, including because of historical structures of intentional discrimination. What is required in the context of emerging digital technologies is careful attention to their racialized and ethnic impact, from government officials, the United Nations and other multilateral organizations, and the private sector. In the present report, the Special Rapporteur highlights intersectional forms of discrimination, including on the basis of gender and religion, and calls attention to the ongoing failure of States and other stakeholders to track and address compounded forms of discrimination at the intersections among race, ethnicity, gender, disability status, sexual orientation and related grounds.

3.   The Special Rapporteur only briefly addresses the racially discriminatory impacts of emerging digital technologies on migrants, refugees and other non-citizens, because these groups will be the focus of a separate report of the Special Rapporteur to the General Assembly.

4.   A key finding in the report is that emerging digital technologies exacerbate and compound existing inequities, many of which exist along racial, ethnic and national origin grounds. The examples highlighted in the report raise concerns about different forms of racial discrimination in the design and use of emerging digital technologies. In some cases, this discrimination is direct, and explicitly motivated by intolerance or prejudice. In other cases, discrimination results from disparate impacts on groups according to their race, ethnicity or national origin, even when an explicit intent to discriminate is absent. And in yet other cases, direct and indirect forms of discrimination exist in combination, and can have such a significant holistic or systemic effect as to subject groups to racially discriminatory structures that pervade access to and enjoyment of human rights in all areas of their lives.

5.   In the context of the coronavirus disease (COVID-19) pandemic, for example, early reports have shown the disparate effects of the pandemic on marginalized racial and ethnic groups, including because of the exclusion of these groups from the benefits of emerging digital technologies, or because emerging digital technologies are deployed in ways that put these groups at greater risk of human rights violations. Notwithstanding widespread perceptions of emerging digital technologies as neutral and objective in their operation, race and ethnicity shape access to and enjoyment of human rights in all of the fields in which these technologies are now pervasive. States have obligations to prevent, combat and

---

[1] See www.ohchr.org/Documents/HRBodies/SP/List_SP_Reports_NewTech.pdf.
[2] See A/HRC/39/29.
[3] See https://qz.com/1585645/color-blindness-is-a-bad-approach-to-solving-bias-in-algorithms.

remediate this racial discrimination, and private actors, such as corporations, have related responsibilities to do the same.

6.   Among emerging digital technologies, the Special Rapporteur focuses in the report on networked and predictive technologies, many involving big data and artificial intelligence, with some emphasis on algorithmic (and algorithmically assisted) decision-making. Much of the existing human rights analysis of racial discrimination and emerging digital technologies has shed light on a specific set of issues: online hate incidents and the use of digital platforms to coordinate, fund and build support for racist communities and their activities. In the report, the Special Rapporteur goes a step further, bringing racial equality and non-discrimination principles to bear on the structural and institutional impacts of emerging digital technologies, which researchers, advocates and others have identified as alarming. Among the concerns is the prevalence of emerging digital technologies in determining everyday outcomes in employment, education, health care and criminal justice, which introduces the risk of systemized discrimination on an unprecedented scale. A recent report from the European Union Agency for Fundamental Rights highlights examples of these concerns in the European Union and provides valuable recommendations for the required response.[4]

7.   As "classification technologies that differentiate, rank, and categorize", artificial intelligence systems are at their core "systems of discrimination".[5] Machine-learning algorithms reproduce bias embedded in large-scale data sets capable of mimicking and reproducing implicit biases of humans, even in the absence of explicit algorithmic rules that stereotype.[6] Data sets, as a product of human design, can be biased due to "skews, gaps, and faulty assumptions".[7] They can also suffer from "signal problems", demographic non- or under-representation because of the unequal ways in which data were created or collected.[8] In addition to inaccurate, missing and poorly represented data, "dirty data" include data that have been manipulated intentionally or distorted by biases.[9] Such data sets potentially lead to discrimination against or exclusion of certain populations, notably minorities along identities of race, ethnicity, religion and gender.

8.   Even where discrimination is not intended, indirect discrimination can result from using innocuous and genuinely relevant criteria that also operate as proxies for race and ethnicity. Other concerns include the use of and reliance on predictive models that incorporate historical data – data often reflecting discriminatory biases and inaccurate profiling – including in contexts such as law enforcement, national security and immigration. At a more fundamental level, the design of emerging digital technologies requires developers to make choices about how to best achieve their goals, and those choices will result in different distributional consequences.[10] A core concern of the Special Rapporteur in the report is with such choices that disparately affect the human rights of individuals and groups on the basis of their race, ethnicity and related grounds.

9.   With respect to class in particular, research shows that even where policymakers, civil servants and scientists have pursued automated decision-making with an intention to make more efficient and more fair decisions, the systems they used to achieve these ends have been shown to reinforce inequality, and result in punitive outcomes for persons living in poverty.[11] Given that racially and ethnically marginalized communities often disproportionately live under conditions of poverty, equality and non-discrimination principles should be central to human rights analyses of emerging digital technologies for social welfare and other socioeconomic systems. An important recent report by the Special

---

[4]   See https://fra.europa.eu/en/publication/2018/bigdata-discrimination-data-supported-decision-making.

[5]   Sarah Myers West, Meredith Whittaker and Kate Crawford, "Discriminating systems: gender, race and power in AI" (New York, AI Now Institute, 2019), p. 6.

[6]   See https://philmachinelearning.files.wordpress.com/2018/02/gabbriellejohnson_algorithmic-bias.pdf.

[7]   See https://foreignpolicy.com/2013/05/10/think-again-big-data.

[8]   Ibid.

[9]   See https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3403010.

[10]   See https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899.

[11]   Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York, Picador, 2018).

Rapporteur on extreme poverty and human rights describes the rise of the digital welfare state in countries in which systems of social protection and assistance are powered by emerging digital technologies, and in ways that have severe negative human rights implications for social welfare.[12] As elaborated in a later section, the Special Rapporteur's assessment is that digital welfare states, as they exist today, are better described as "discriminatory digital welfare states" because the trend is that they allow race and ethnicity (among other grounds) to shape access to human rights on a discriminatory basis. Urgent intervention is necessary to curb these discriminatory patterns.

10.   In the preparation of the report, the Special Rapporteur benefited from valuable input from: expert group meetings hosted by the Global Studies Institute of the University of Geneva, the Promise Institute for Human Rights at the University of California, Los Angeles, (UCLA) School of Law and the UCLA Center for Critical Internet Inquiry; research by the Harvard Law School Cyberlaw Clinic at the Berkman Klein Center for Internet & Society, and the New York University School of Law Center on Race, Inequality and the Law; interviews with researchers; and submissions received by a range of stakeholders in response to a public call for submissions. Non-confidential submissions will be available on the webpage of the mandate.[13]

II. Drivers of discrimination and inequality in emerging digital technologies

11.   Any human rights analysis of emerging digital technologies must first grapple with the social, economic and political forces that shape their design and use, and with the individual and collective human interests and priorities at play that contribute to the racially discriminatory design and use of these technologies.

12.   The public perception of technology tends to be that it is inherently neutral and objective, and some have pointed out that this presumption of technological objectivity and neutrality is one that remains salient even among producers of technology. But technology is never neutral – it reflects the values and interests of those who influence its design and use, and is fundamentally shaped by the same structures of inequality that operate in society.[14] For example, a 2019 review of 189 facial recognition algorithms from 99 developers around the world found that "many of these algorithms were 10 to 100 times more likely to inaccurately identify a photograph of a black or East Asian face, compared with a white one. In searching a database to find a given face, most of them picked incorrect images among black women at significantly higher rates than they did among other demographics."[15] There can no longer be any doubt that emerging digital technologies have a striking capacity to reproduce, reinforce and even to exacerbate racial inequality within and across societies. A number of important academic studies have shown concretely that the design and use of technologies are already having this precise effect across a variety of contexts.[16] More research and funding are required to unpack fully how even the inductive processes at the core of some artificial intelligence techniques, such as machine learning, contribute to undercutting values such as equality and non-discrimination.[17]

---

[12] See A/74/493.

[13] See www.ohchr.org/EN/Issues/Racism/SRRacism/Pages/Callinformationtechnologies.aspx.

[14] Langdon Winner, *The Whale and The Reactor: A Search for Limits in an Age of High Technology* (Chicago, University of Chicago Press, 1986), p. 29.

[15] See www.scientificamerican.com/article/how-nist-tested-facial-recognition-algorithms-for-racial-bias.

[16] See, e.g., Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York, Penguin, 2016); Ruha Benjamin, *Race After Technology* (Cambridge, United Kingdom, Polity Press, 2019); and Safiya Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York, New York University Press, 2018).

[17] See, e.g., Gabrielle M. Johnson, "Are algorithms value-free? Feminist theoretical virtues in machine learning", *Journal of Moral Philosophy* (forthcoming).

13.   Within the fields and industries that produce emerging digital technologies, misplaced faith in the neutrality or objectivity of numbers and their power to overcome racism has been shown to contribute to discriminatory outcomes.[18] Even the field that has developed to promote fairness, accountability and transparency in the design and use of emerging digital technologies needs to pay greater attention to the broader societal structures of discrimination and injustice.[19] Indeed, among the biggest challenges to addressing racially discriminatory use and design of emerging digital technologies are approaches that treat this issue as purely or largely a technological problem for computer scientists and other industry professionals to solve by engineering bias-free data and algorithms. Technology is a product of society, its values, its priorities and even its inequities, including those related to racism and intolerance. Technological determinism – the idea that technology influences society but is itself largely neutral and insulated from social, political and economic forces – only serves to shield the forces that shape emerging digital technologies and their effects from detection and reform. "Techno-chauvinism" – an overreliance on the belief that technology can solve societal problems[20] – has a similar effect, and can complicate interrogating and changing the values and interests that shape technology and technological outcomes.

14.   Although there remains a great need for scrutiny of and accountability for the quality of engineering in ensuring equality and non-discrimination principles, securing these and other human rights principles must begin with an acknowledgment that the heart of the issue is a political, social and economic one, and not solely a technological or mathematical problem. Inequality and discrimination, even in those circumstances in which they are the product of the design and use of emerging digital technologies, will not be "cured" by more perfect modelling of equality and non-discrimination. Concretely, this means that thinking and action that seek to combat racial discrimination in the design and use of emerging digital technologies, both in the private and public sectors, should not be the exclusive or near-exclusive terrain of technology experts. Instead, such thinking and action must be more holistic, as researchers and others with expertise in emerging digital technologies have argued.[21] Governments and the private sector must commit to approaches that include experts on the political, economic and social dimensions of racial discrimination at all stages of research, debate and decision-making to mitigate racially discriminatory design and use of emerging digital technologies. Affected racial and ethnic minority communities must play decision-making roles in the relevant processes.

15.   Private corporations wield monumental influence in the design and use of emerging digital technologies. Among digital platforms, seven "super platforms" – Microsoft, Apple, Amazon, Google, Facebook, Tencent and Alibaba – account for two thirds of the total market value of the world's 70 largest platforms.[22] Notwithstanding the global reach of their emerging digital technologies, the corporations that exert the greatest influence over them are predominantly concentrated in Silicon Valley, in the United States of America, while Europe's share is 3.6 per cent, that of Africa 1.3 per cent and that of Latin America 0.2 per cent.[23] For example, Google has 90 per cent of the global market for Internet searches.[24] Occupying two thirds of the global social media market, Facebook is the top social media platform in more than 90 per cent of the world's economies. Amazon has an almost 40 per cent share of the world's online retail activity. As a result, the specific cultural, economic and political values of Silicon Valley fundamentally shape how many of

---

[18] See, e.g., West, Whittaker and Crawford, "Discriminating systems".

[19] See, e.g., www.tandfonline.com/doi/full/10.1080/1369118X.2019.1593484 and http://sorelle.friedler.net/papers/sts_fat2019.pdf.

[20] See, e.g., Meredith Broussard, *Artificial Unintelligence: How Computers Misunderstand the World* (Cambridge, Massachusetts, Massachusetts Institute of Technology, 2018).

[21] See, e.g., www.odbproject.org/2019/07/15/critiquing-and-rethinking-fairness-accountability-and-transparency.

[22] *Digital Economy Report 2019: Value Creation and Capture: Implications for Developing Countries* (United Nations publication, Sales No. E.19.II.D.17), p. xvi.

[23] Ibid., p. 2.

[24] Ibid., p. xvii.

the emerging digital technologies operate globally, including in contexts very far removed from this small region of North America.

16.   Beyond market dominance, corporations serve as key intermediaries between Governments and their nations, with the capacity to significantly transform the situation of human rights. Technology produced by powerful global North corporations is created in a very specific political, economic, social and governance context. It can have egregious effects in other contexts, such as those in the global South. One example is the role that Facebook played in Myanmar.[25] There are also concerns about the unregulated, and in some cases exploitative, terms on which data are extracted from individuals and nations in the global South, by profit-seeking corporate actors in the global North who cannot be held accountable.[26]

17.   Emerging digital technology sectors, such as those in Silicon Valley, are characterized by a "diversity crisis" along gender and race lines,[27] especially at the highest levels of decision-making. According to an important study of the field, "currently, large scale AI systems are developed almost exclusively in a handful of technology companies and a small set of elite university laboratories, spaces that in the West tend to be extremely white, affluent, technically oriented, and male. These are also spaces that have a history of problems of discrimination, exclusion, and sexual harassment."[28] The study further finds that "this is much more than an issue of one or two bad actors: it points to a systematic relationship between patterns of exclusion within the field of AI and the industry driving its production on the one hand, and the biases that manifest in the logics and application of AI technologies on the other."[29] Technology produced in such fields that disproportionately exclude women, racial, ethnic and other minorities is likely to reproduce these inequalities when it is deployed. Producing technology that works within complex social realities and existing systems requires understanding social, legal and ethical contexts, which can only be done by incorporating diverse and representative perspectives as well as disciplinary expertise.[30]

18.   Market and economic forces exert a powerful influence on the design and use of emerging digital technologies, which in turn have a transformational effect on markets, even on capitalism itself.[31] On the one hand, some economic influence seeks intentionally to promote discrimination and intolerance. Examples include wealthy individuals who fund online platforms advocating supremacist ideology.[32] On the other hand, the most powerful market forces may primarily seek profitable outcomes from emerging digital technologies without explicitly racist or intolerant intentions. But the evidence shows that profitable products can produce racial discrimination. Where economies are structured by racial and ethnic inequality – as is the case all over the world – profit maximization will typically be consistent with and in many cases reinforce or compound racial and ethnic inequality.

19.   To a great extent, inequalities in access to and enjoyment of the benefits of emerging digital technologies track (a) geopolitical inequalities at the international level, and (b) patterns of racial, ethnic and gendered inequality within individual countries.[33]

20.   At the international level, countries in the global South lack the digital infrastructure that exists in the global North: active broadband subscription in the global South is less

---

[25] See A/HRC/39/64.

[26] See, e.g., https://privacyinternational.org/long-read/3390/2020-crucial-year-fight-data-protection-africa.

[27] See www.technologyreview.com/2018/02/14/145462/were-in-a-diversity-crisis-black-in-ais-founder-on-whats-poisoning-the-algorithms-in-our; Noble, *Algorithms of Oppression*; and West, Whittaker and Crawford, "Discriminating systems".

[28] West, Whittaker and Crawford, "Discriminating systems", p. 6 (footnote omitted).

[29] Ibid. p. 7.

[30] Ibid.

[31] See, e.g., Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York, Public Affairs, 2019).

[32] See, e.g., https://journals.sagepub.com/doi/full/10.1177/1536504218766547.

[33] See, e.g., https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3403010.

than half that in the global North.[34] In Africa, 22 per cent of individuals use the Internet, compared with 80 per cent in Europe.[35] In the so-called least developed countries, only one in five persons is online, in contrast to four out of five in so-called developed countries.[36] Even as technology has been beneficial for national responses to the COVID-19 pandemic, these benefits have not been evenly distributed. The least developed countries are not only the most vulnerable to the human and economic consequences of COVID-19, but also the least digitally ready to access public health information online and to make use of digital schooling, working and shopping platforms.[37]

21.   Among countries, the United States and China dominate the global digital economy. These two countries account for 90 per cent of the market capitalization value of the world's 70 largest digital platforms, which include social media and content platforms, e-commerce platforms, Internet search services, mobile ecosystems and industrial cloud-based platforms.[38] Current predictions suggest that emerging digital technologies are poised to further widen the digital divide between countries that have and those that lack the capabilities to take advantage of such technologies.

22.   Digital divides also exist within countries. For example, notwithstanding the dominance of the United States within the global digital economy, racial and ethnic minorities in that country have disparate access to the benefits of emerging digital technologies. In many cases, they are subject to the most significant human rights violations associated with emerging digital technologies, as illustrated in part III below. According to a 2019 survey by the Pew Research Center, black and Hispanic adults remain less likely to own a computer or have high speed Internet at home in the United States.[39] While 82 per cent of whites report owning a desktop or laptop computer, only 58 per cent of blacks and 57 per cent of Hispanics do.[40] Substantial racial and ethnic differences in broadband adoption also exist, with whites being 13 to 18 per cent more likely to report having a broadband connection at home than blacks or Hispanics.[41] This digital divide along the axis of race and ethnicity is significant. As researchers have argued, however, interventions to promote digital inclusion of racial and ethnic minorities must not be pursued in ways that expose them to further rights violations, including as a result of privacy and surveillance concerns.[42] In the case of China, part III below exemplifies the severe human rights consequences of its design and use of emerging digital technologies. These concerns are further amplified by the growing influence of Chinese emerging digital technologies in the global South.

23.   Indigenous peoples are also subject to discriminatory exclusion from the benefits of emerging digital technologies.[43] Estimates in Canada show that approximately half of the predominantly indigenous northern population lack the high speed connections available to their southern counterparts.[44] Indigenous digital inclusion has also been low in Australia, especially outside cities, with only 6 per cent of residents in some remote Aboriginal communities having a computer in 2011.[45] By 2015, Aboriginal people were still 69 per cent less likely than non-Aboriginal people to have any Internet connection.[46]

---

[34] *Trade and Development Report 2018: Power, Platforms and the Free Trade Delusion* (United Nations publication, Sales No. E.18.II.D.7), p. viii.

[35] Ibid.

[36] See *Digital Economy Report 2019*, p. 13.

[37] See https://unctad.org/en/PublicationsLibrary/dtlinf2020d1_en.pdf.

[38] *Digital Economy Report 2019*, p. xvi.

[39] See www.pewresearch.org/fact-tank/2019/08/20/smartphones-help-blacks-hispanics-bridge-some-but-not-all-digital-gaps-with-whites.

[40] Ibid.

[41] Ibid.

[42] See https://journals.uic.edu/ojs/index.php/fm/article/view/3821.

[43] See https://unesdoc.unesco.org/ark:/48223/pf0000260781.

[44] See https://rightscon2018.sched.com/event/EHqs/addressing-the-digital-divide-in-indigenous-communities-in-north-america.

[45] See www.creativespirits.info/aboriginalculture/economy/internet-access-in-aboriginal-communities.

[46] Ibid.

III. Examples of racial discrimination in the design and use of emerging digital technologies

A. Explicit intolerance and prejudice-motivated conduct

24.   Actors seeking to spread racist speech and incitement to discrimination and violence have relied on emerging digital technologies, with social media platforms playing a pivotal role. The Special Rapporteur has highlighted these trends in previous reports on neo-Nazi and other white supremacist groups that rely on social media platforms to recruit, raise funds and coordinate.[47] Another prominent example of explicitly prejudice-motivated use of emerging digital technologies is the use of Facebook by radical nationalist Buddhist groups and military actors in Myanmar to exacerbate discrimination and violence against Muslims and the Rohingya ethnic minority in particular.[48] In 2018, the Chief Executive Officer of Facebook, Mark Zuckerberg, testified to the United States Senate that Facebook's artificial intelligence systems were unable to detect hate speech in such contexts.[49] These are not the only instances: a submission also highlighted the use of Facebook to amplify discriminatory and intolerant content, including content inciting violence against religious and linguistic minority groups in India.[50]

25.   Social media bots – automated accounts – have been used to shift political discourse and misrepresent public opinion. Out of a sample of 70 countries, bots were used in 50 countries for social media manipulation campaigns in 2019.[51] For groups that rely on emerging digital technologies as a strategy for promoting racial, ethnic and religious discord and intolerance, bots are central to their capacity to spread racist speech or disinformation online. Examples suggest that the coordinated use of bots has been especially prevalent before elections. For example, leading up to the Swedish election in 2018, researchers identified 6 per cent of Twitter accounts discussing national politics as bots, which posted about topics related to immigration and Islam more than genuine accounts.[52] Similarly, in the period before the 2018 election in the United States, 28 per cent of Twitter accounts posting antisemitic tweets were bots, which posted 43 per cent of all antisemitic tweets.[53] Emerging digital technologies in the Russian Federation have been used to promote ethnic and racial divisions on social media,[54] through hundreds of falsified online personas and pages on Twitter, Facebook and other social media sites. Although some posts were directed towards ethnic minority groups and called for racial equality, many denounced such groups in an effort to promote racial tensions. Some personas supported white nationalist groups, prompting discrimination and violence against racial minorities.[55]

B. Direct or indirect discriminatory design/use of emerging digital technology

26.   The design and use of emerging digital technologies can directly and indirectly discriminate along racial or ethnic lines in access to a range of human rights.

---

[47] See A/73/312 and A/HRC/41/55.
[48] A/HRC/42/50, paras. 71–75.
[49] See www.commerce.senate.gov/2018/4/facebook-social-media-privacy-and-the-use-and-abuse-of-data.
[50] Submission by Avaaz.
[51] See https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf.
[52] See www.semanticscholar.org/paper/Political-Bots-and-the-Swedish-General-Election-Fernquist-Kaati/2af3d1e16d5553dc489d8b44321ea543d571a4a9.
[53] See www.adl.org/resources/reports/computational-propaganda-jewish-americans-and-the-2018-midterms-the-amplification.
[54] See https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3304223.
[55] Ibid., p. 180.

27. With respect to the right to work, in one submission it was reported that Paraguay had implemented a digital employment system that allowed employers to sort and filter prospective employees by various categories, some of which served as proxies for race.[56] Furthermore, the system is only available in Spanish, when less than half of rural indigenous peoples in Paraguay speak Spanish. Such limited language accessibility effectively restricts this system's availability to jobseekers on an ethnic basis, even if this is not the intention of policymakers.

28. Algorithms used for selecting successful job candidates in North America and Europe have also been criticized for making discriminatory recommendations. These systems are trained to identify candidates based on data sets of existing "successful" employees that include information on protected characteristics, such as gender, ethnicity or religion. As a result, the respective algorithmic systems reproduce and reinforce existing racial, ethnic, gender or other bias by making decisions that reflect existing inequities in employment. Such systems effectuate direct and indirect forms of racial discrimination.[57] On the other hand, when these systems prohibit any consideration of protected statuses, such as race and ethnicity, they can undercut special measures or affirmative action that States may have adopted to promote equal employment opportunities.

29. In other cases, the introduction of automated systems that do not rely directly on discriminatory inputs or processes can nonetheless indirectly discriminate against marginalized groups in their access to work by reducing or eliminating available positions. A submission provided the example of a new artificial intelligence-based project for sanitation management in India that would eliminate the need for many jobs typically performed by those in the lowest, or Dalit, caste.[58] Dalits, especially women, can often only find employment in the sanitation sector, and some Indian states have prioritized Dalits for sanitation jobs. Implementation of smart sanitation systems would likely affect the jobs and livelihoods of Dalits disproportionately, especially Dalit women. In light of the broader socioeconomic and political marginalization of Dalits in India, automation in the sanitation sector might fundamentally undercut access to work for those who rely on employment in the sanitation sector.

30. Emerging digital technologies also have a discriminatory impact on the right to health. The top 10 health-care algorithms on the United States market use patients' past medical costs to predict future costs, which are used as a proxy for health-care needs.[59] A recent study of such an algorithm used by a leading health services company found that it had been unintentionally yet systematically discriminating against black patients in the United States.[60] Intended to help enrol high-risk patients in care management programmes, the algorithm was found to encode racial bias by using patients' health-care costs as a proxy for their health needs in order to predict their level of risk.[61] Considered by its developers as "race-blind" because race was not an input,[62] the algorithm consistently assigned lower risk scores to black patients who were equally sick as their white counterparts.[63] The algorithm failed to identify less than half the number of black patients at risk of complicated medical needs as white patients. As a result, black patients were less likely to be referred to programmes for interventions to improve their health. Hospitals, insurers and government agencies use this algorithm and similar ones to help manage care for over 200 million people in the country each year.[64]

31. In another example from the United States, a recent case study examined a predictive model developed by Epic Systems Corporation, the leading global developer of

[56] Submission by the Equal Rights Trust.
[57] See https://fra.europa.eu/en/publication/2018/bigdata-discrimination-data-supported-decision-making.
[58] Submission by the Association of Progressive Communications. See also www.apc.org/sites/default/files/gisw2019_artificial_intelligence.pdf.
[59] See www.sciencenews.org/article/bias-common-health-care-algorithm-hurts-black-patients.
[60] See https://science.sciencemag.org/content/366/6464/447.
[61] Ibid.
[62] See www.thelancet.com/journals/landig/article/PIIS2589-7500(19)30201-8/fulltext.
[63] See https://science.sciencemag.org/content/366/6464/447.
[64] Ibid.

electronic health records.[65] Integrated directly into existing electronic health records, Epic's artificial intelligence tool estimates the likelihood that a patient will no-show by using the patient's personal information, including ethnicity, class, religion and body mass index, as well as the patient's record of prior no-shows. In pointing out the obvious potential to discriminate against vulnerable patient populations, the researchers note that "removing sensitive personal characteristics from a model is an incomplete approach to removing bias."[66] Prior no-shows, for example, likely correlate with socioeconomic status mediated by the patient's inability to cover transportation or childcare costs, or to take time off work for the appointment. They also likely correlate with race and ethnicity because of correlations between socioeconomic status, and race and ethnicity.[67] It was revealed in another recent study that black patients were more likely to be scheduled into overbooked appointment slots and thus had to wait longer when they did show up.[68]

32.   In the housing context, studies in the United States have shown ethnic discrimination in Facebook's targeted advertising. Facebook used to allow advertisers to "narrow audience" by excluding Facebook users with certain "ethnic affinities" under the "demographics" category of its ad-targeting tool.[69] This targeted advertising could be used to prevent black people from viewing specific housing advertisements, which is prohibited under United States anti-discrimination law. Facebook controls an estimated 22 per cent of the market share for digital advertisements in the United States,[70] and its targeted advertising, which is the core of the company's business model,[71] has been shown to be racially exclusionary.[72] These practices are best understood as a form of digital redlining, which is defined as "the creation and maintenance of technology practices that further entrench discriminatory practices against already marginalized groups".[73] Facebook uses targeted advertising in the employment context as well, raising similar concerns.

33.   In yet other cases, access to technology – and the information available through it – are denied in ways that have disparate impacts, or that target specific racial, ethnic or religious groups, sometimes on a discriminatory basis. In 2019, multiple States, including Bangladesh, the Democratic Republic of the Congo, Egypt, India, Indonesia, Iran (Islamic Republic of), Myanmar, the Sudan and Zimbabwe, completely restricted Internet access to specific regions, with the effect of preventing nearly all communication in or out of those regions.[74] Researchers have linked more targeted Internet shutdowns to regions with higher densities of minority groups.[75]

34.   With respect to the right to a fair trial, multiple courts in Latin America have begun using Prometea, a software that uses voice recognition and machine learning prediction, to streamline judicial proceedings. The district attorney's office and courts in Buenos Aires use this artificial intelligence system to automate judicial decision-making in simple cases, such as disputes about taxi licences and complaints from teachers about not being compensated for school supplies.[76] In such cases, Prometea interprets the facts given to it and suggests a legal outcome based on prior jurisprudence in similar cases. A judge must approve the decision before it is made official, which is the case 96 per cent of the time.[77] A

[65] See www.healthaffairs.org/do/10.1377/hblog20200128.626576/full.
[66] Ibid.
[67] See https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3467047.
[68] Ibid.
[69] See www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race.
[70] See www.emarketer.com/content/us-digital-ad-spending-will-surpass-traditional-in-2019.
[71] See www.motherjones.com/politics/2019/12/facebook-agreed-not-to-let-its-ads-discriminate-but-they-still-can.
[72] See www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin.
[73] See www.congress.gov/116/meeting/house/110251/witnesses/HHRG-116-BA00-Wstate-GillardC-20191121.pdf.
[74] See www.hrw.org/news/2019/12/19/shutting-down-internet-shut-critics.
[75] See www.accessnow.org/cms/assets/uploads/2020/02/KeepItOn-2019-report-1.pdf.
[76] See www.bloombergquint.com/businessweek/this-ai-startup-generates-legal-papers-without-lawyers-and-suggests-a-ruling.
[77] See www.giswatch.org/2019-artificial-intelligence-human-rights-social-justice-and-development.

real concern is that this high approval rate may well result from a presumption of technological objectivity and neutrality as discussed. The Constitutional Court of Colombia uses Prometea to filter *tutelas*, or individual constitutional rights complaints, and decide which to hear.[78] The concern with Prometea and many other such artificial intelligence systems is the "black box" effect – the basis of their decision-making is opaque, and it is difficult or impossible for judges, other court officials, and litigants (and even public authorities who commission these systems) to determine bias in design, input or output. While it is impossible to know the impact that Prometea has or could have on racial and ethnic minorities, the risk is high that such systems will reinforce or exacerbate existing racial and ethnic disparities in the justice systems in which they are deployed.

35.   In the criminal justice context, police departments in different parts of the world use emerging digital technologies for predictive policing, in which artificial intelligence systems pull from multiple sources of data, such as criminal records, crime statistics and the demographics of neighbourhoods.[79] Many of these data sets reflect existing racial and ethnic bias, thus operating in ways that reinforce racial discrimination despite the presumed "objectivity" of these technologies or even their perceived potential to mitigate the bias of the human actors they supplement or replace. Furthermore, police departments tend to deploy predictive technologies disproportionately in impoverished communities of predominantly ethnic or racial minorities.

36.   The United Kingdom of Great Britain and Northern Ireland, for example, uses a database, known as the Gangs Violence Matrix, that has been demonstrated to be discriminatory.[80] Police officers reportedly make assumptions about individuals based on their race, gender, age and socioeconomic status, which further reinforce those stereotypes.[81] The result is that 78 per cent of individuals on the Matrix are black, and an additional 9 per cent are from other ethnic minority groups, while the police's own figures show that only 27 per cent of those responsible for serious youth violence are black. The police also share the Matrix with other agencies, such as job centres, housing associations, and educational institutions, leading to discrimination against individuals on the basis of their supposed gang affiliation. Depending on the nature of the way this information is shared, this poses an opportunity for possible violations of the right to privacy and may affect housing and employment rights on a discriminatory basis. Those whose names are on the Matrix experience "multiple stop and search encounters which seemingly lack any legal basis".[82] Some report that police have stopped and searched them 200 times, others report up to as many as 1,000 times, with some reporting multiple stops every day. This has an impact on individuals' rights to freedom from interference with their privacy and their freedom from arbitrary arrest on an ethnically discriminatory basis.

37.   By way of another example, it was highlighted in one submission that predictive policing was becoming the methodology used in local policing in so-called crime prevention strategies in cities in the United States such as Los Angeles.[83] Until recently, the Los Angeles Police Department had been using technology called PredPol to examine 10 years of crime data, including the types, dates, locations and frequency of crimes, to predict when and where crimes would likely occur over the next 12 hours. These data, gathered and categorized by police officers, are both the product and the cause of heightened surveillance in black and Latinx communities. Predictive policing reiterates and exacerbates the existing biases in the policing system, while providing the guise of objectivity because of the use of supposedly neutral algorithmic decision-making. Although the Los Angeles Police Department has suspended its use of PredPol, it has not disavowed use of other predictive policing products that are likely to raise similar concerns.

---

[78] See www.ambitojuridico.com/noticias/informe/constitucional-y-derechos-humanos/prometea-inteligencia-artificial-para-la (in Spanish).
[79] Submission by the Association for Progressive Communications.
[80] A/HRC/41/54/Add.2, para. 40.
[81] See www.amnesty.org.uk/files/reports/Trapped%20in%20the%20Matrix%20Amnesty%20report.pdf.
[82] See www.stop-watch.org/uploads/documents/Being_Matrixed.pdf.
[83] Submission by the Stop LAPD Spying Coalition.

C. Racially discriminatory structures

38.   Examples from different parts of the world show that the design and use of different emerging digital technologies can be combined intentionally and unintentionally to produce racially discriminatory structures that holistically or systematically undermine enjoyment of human rights for certain groups, on account of their race, ethnicity or national origin, in combination with other characteristics. In other words, rather than only viewing emerging digital technologies as capable of undercutting access to and enjoyment of discrete human rights, they should also be understood as capable of creating and sustaining racial and ethnic exclusion in systemic or structural terms. Under this subheading, the Special Rapporteur reviews examples of existing and potentially discriminatory structures, emphasizing the prevalence of biometric data systems, racialized surveillance and racialized predictive analytics in maintaining these structures.

39.   China uses biometric identification and surveillance to track and restrict the movements and activities of the Uighur ethnic minority group, violating members of this group's rights to equality and non-discrimination, among others.[84] Uighurs experience frequent baseless police stops and are subjected to having their telephones scanned at police checkpoints, which violates their right to privacy. There is a mandatory collection of extensive biometric data, including DNA samples and iris scans, for Uighurs. According to credible reports, the State, "using a combination of facial recognition technology and surveillance cameras throughout the country, looks exclusively for Uighurs based on their appearance and keeps records of their comings and goings for search and review".[85] It is also noted in reports that this surveillance and data collection activity is occurring alongside large numbers of ethnic minorities being held incommunicado in political "re-education camps" under the pretext of countering religious extremism, without detainees being charged or tried.[86] The picture that emerges is one of systemic ethnic discrimination, supported and indeed made possible by a number of emerging digital technologies, which violates a broad spectrum of human rights for Uighurs.

40.   Kenya and India have implemented biometric identification for accessing public services, known as Huduma Namba and Aadhaar, respectively.[87] The programmes include collection of various forms of biometric data, including fingerprints, retina and iris patterns, voice patterns and other identifiers. When trying to access public services through these systems, certain racial and ethnic minority groups in both countries find that they are excluded from them, while others face logistical barriers and long vetting processes that in effect can result in de facto exclusion from accessing public services to which they are entitled. These public services include pensions and unemployment benefits in India, and all essential government services in Kenya, including voting, registering birth certificates and civil marriages, paying taxes and receiving deeds to property. The Supreme Court of India has upheld the statute requiring the Aadhaar number for receiving government welfare. Despite the same judgment prohibiting private entities from using Aadhaar for non-governmental purposes, like banking, employment and mobile telecommunications, such a requirement remains prevalent in practice. Furthermore, persons with disabilities – including among ethnic and racial minorities – experience discrimination for not being able to provide fingerprint or iris scans. Though the law provides special mechanisms for such persons, they continue to face logistical hurdles because many centres have no training in enrolling them without the biometric data.[88] Without stringent protections, digital identification systems for public services disproportionately exclude racial and ethnic minorities, especially those whose citizenship status is insecure.[89]

---

[84] See CERD/C/CHN/CO/14-17.

[85] See www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html; and A/HRC/41/35, para. 12.

[86] See CERD/C/CHN/CO/14-17.

[87] See A/74/493.

[88] See https://timesofindia.indiatimes.com/city/kolkata/court-relief-in-disabled-womans-aadhaar-battle/articleshow/68961357.cms.

[89] For a human rights analysis of racial discrimination in access to citizenship, see A/HRC/38/52.

41.   Many States are experimenting with incorporating emerging digital technologies into their welfare systems,[90] in ways that reinforce racially discriminatory structures. Australia has implemented the Online Compliance Intervention system, colloquially known as robo-debt.[91] This automated decision-making system uses machine learning algorithms to identify suspected overpayment of government welfare benefits and demands documentation from those recipients marked as having received more than they were entitled to in welfare payments. The system sent out approximately 20,000 debt letters each week for a six-month period in 2016 and 2017. An investigation estimated that between 20 and 40 per cent of debt letters were false positives based on flaws in the system processes and the data. The State shifted the onus onto welfare recipients to prove that they did not owe the State a debt. Recipients of welfare benefits at higher rates than white Australians,[92] indigenous Australians bear the greatest cost of this system's flaws, while being the worst equipped to challenge them given the barriers that they face. A recent human rights intervention in judicial proceedings highlights similar concerns in the Netherlands, where use of emerging digital technologies in the provision of social welfare has resulted in human rights violations against the poorest and most vulnerable in that country.[93] There, too, racial and ethnic minorities face disproportionate socioeconomic marginalization, raising pressing concerns that class discrimination is also racial discrimination.

42.   As States increasingly use emerging digital technologies to calculate risk and classify need, as exemplified by countries such as Denmark, New Zealand, the United Kingdom and the United States,[94] greater scrutiny of their potential to have a disparate impact on racial or ethnic minorities must be a State priority. Because digitalization of welfare systems occurs in societies in which groups are marginalized, discriminated against and excluded on a racial and ethnic basis, these systems are almost guaranteed to reinforce these inequities, unless States actively take preventive steps. Without urgent intervention, digital welfare states risk entrenching themselves as discriminatory digital welfare states.

43.   In some cases, although the racially discriminatory structures are sectoral, for example criminal justice, they nonetheless holistically undercut the human rights of those affected and reinforce their structural oppression in society. Such is the case in the United States, where emerging digital technologies sustain and reproduce racially discriminatory structures in the administration of criminal justice. There, emerging digital technologies are common not only in policing but also in the justice system, where they have been associated with discriminatory outcomes for racial and ethnic minorities. Several states in the United States use artificial intelligence risk assessment tools in every step of the criminal justice process. The developers intend these systems to provide objective, data-driven justice outcomes,[95] but the algorithms often rely on "data produced during documented periods of flawed, racially biased, and sometimes unlawful practices and policies".[96] As these algorithms affect sentencing, they can violate an individual's rights to equality before the law, to a fair trial, and to freedom from arbitrary arrest and detention. Such risk assessments often weigh such factors as prior arrests and convictions, parental criminal record, postal code and so-called "community disorganization".[97] As the authors of one study find: "These factors reflect over-policing, the behaviours of law enforcement in Black and brown communities, larger patterns of socioeconomic disadvantage resulting from the racial caste system, rather than anything about the behaviours of people who are

---

[90]   See A/74/493.

[91]   See www.unswlawjournal.unsw.edu.au/forum_article/new-digital-future-welfare-debts-without-proofs-authority and www.ombudsman.gov.au/__data/assets/pdf_file/0022/43528/Report-Centrelinks-automated-debt-raising-and-recovery-system-April-2017.pdf.

[92]   See www.aihw.gov.au/reports/australias-welfare/australias-welfare-2019-data-insights/contents/summary.

[93]   See www.ohchr.org/Documents/Issues/Poverty/Amicusfinalversionsigned.pdf.

[94]   A/74/493, para. 27.

[95]   See www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[96]   See https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3333423.

[97]   Submission by the New York University Center on Race, Inequality, and the Law.

targeted. In other words, the data is more predictive of racialized disadvantage and police presence in an accused person's community than a person's behaviour."[98]

IV. A structural and intersectional human rights law approach to racial discrimination in the design and use of emerging digital technologies: analysis and recommendations

44.   Emerging digital technologies pose a mammoth regulatory and governance challenge from a human rights perspective. In many cases, the data, codes and systems responsible for discriminatory and related outcomes are complex and shielded from scrutiny, including by contract and intellectual property laws. In some contexts, not even computer programmers may themselves be able to explain the way that their algorithmic systems function. This "black box"[99] effect makes it difficult for affected groups to overcome steep evidentiary burdens of proof typically required to prove discrimination through legal proceedings, assuming that court processes are even available in the first place. On the other hand, the companies responsible for creating and implementing emerging digital technologies face few if any legal requirements to prove that their systems comply with human rights principles and will not produce racially discriminatory outcomes.

45.   International human rights law will by no means be a panacea for the problems identified in this report but it stands to play an important role in identifying and addressing the social harms of artificial intelligence, and ensuring accountability for these harms,[100] as has been highlighted by the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression.[101] Ethical approaches to governing emerging digital technologies must be pursued in line with international human rights law, and States must ensure that these ethical approaches do not function as a substitute for development and enforcement of existing legally binding obligations. In thus section, the Special Rapporteur explains the concepts and doctrines of direct, indirect and structural racial discrimination under international human rights law and outlines the obligations they impose on States where emerging digital technologies are concerned. These obligations also have implications for non-State actors, such as technology corporations, which in many respects exert more control over these technologies than States do. This section also includes a non-exhaustive list of recommendations for concrete implementation of the norms and obligations presented.

A. Scope of legally prohibited racial discrimination in the design and use of emerging digital technologies

46.   The Special Rapporteur recalls that international human rights law is based on the premise that all persons, by virtue of their humanity, should enjoy all human rights without discrimination on any grounds. The prohibition on racial discrimination has achieved the status of a peremptory norm of international law[102] and as an obligation *erga omnes*.[103] Under international human rights law, States have further elaborated racial equality and non-discrimination obligations across a number of different treaty regimes; the principles of

---

[98]   Ibid.

[99]   Frank Pasquale, *The Black Box Society: the Secret Algorithms that Control Money and Information* (Cambridge, Massachusetts, Harvard University Press, 2015).

[100]   See https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_ Intelligence_Upholding_Human_Rights.pdf.

[101]   A/73/348, paras. 19–60.

[102]   Human Rights Committee, general comment No. 29 (2001) on derogations from provisions of the Covenant during a state of emergency, paras. 8 and 13 (c). See also A/CN.4/727, para. 59.

[103]   *Barcelona Traction, Light and Power Company, Limited, Judgment, I.C.J. Reports 1970*, p. 3, at p. 32, para. 33.

equality and non-discrimination are codified in all core human rights treaties.[104] Article 26 of the International Covenant on Civil and Political Rights states that the law shall prohibit any discrimination and guarantee to all persons equal and effective protection against discrimination on any ground such as race, colour**,** sex, language, religion, political or other opinion, national or social origin, property, birth or other status. The International Covenant on Economic, Social and Cultural Rights also prohibits discrimination on these grounds.[105]

47.   Article 1 (1) of the International Convention on the Elimination of All Forms of Racial Discrimination defines racial discrimination as any distinction, exclusion, restriction or preference based on race, colour, descent, or national or ethnic origin which has the purpose or effect of nullifying or impairing the recognition, enjoyment or exercise, on an equal footing, of human rights and fundamental freedoms in the political, economic, social, cultural or any other field of public life. The Convention aims at much more than a formal vision of equality. Equality in the international human rights framework is substantive and requires States to take action to combat intentional or purposeful racial discrimination, as well as to combat de facto, unintentional or indirect racial discrimination.[106]

48.   **States must deploy a structural understanding of the prohibition on racial discrimination in line with international human rights law in the context of emerging digital technologies. Human rights law definitions must be employed in the vital function of shaping how States determine the meaning of racial discrimination produced through certain uses of emerging digital technologies, and States should require that these definitions inform the approaches of the private sector. This means that they must address not only explicit racism and intolerance in the use and design of emerging digital technologies, but also, and just as seriously, indirect and structural forms of racial discrimination that result from the design and use of such technologies. Obligations to combat racial discrimination extend to the racially discriminatory structures and other forms of direct and indirect discrimination described in part III above. States must reject a "colour-blind" approach to governance and regulation of emerging digital technologies, one that ignores the specific marginalization of racial and ethnic minorities and conceptualizes problems and solutions relating to such technologies without accounting for their likely effects on these groups.**

49.   **Recalling paragraphs 92 to 98 of the Durban Programme of Action, the Special Rapporteur urges States to collect, compile, analyse, disseminate and publish reliable statistical data disaggregated on racial or ethnic grounds, in order to address individual and group racial inequities associated with the design and use of emerging digital technologies. The Special Rapporteur urges States to adopt an approach to data grounded in human rights, by ensuring disaggregation, self-identification, transparency, privacy, participation and accountability in the collection and storage of data.**[107] Identifying and addressing direct and indirect forms of discrimination requires the collection of data (in compliance with human rights standards) capable of revealing the disparate impacts of emerging digital technologies. Yet many States fail to collect or require the collection of such data. In fact, some European Union countries prohibit the collection of disaggregated data that would allow for identification and remediation of discrimination on the basis of ethnicity or race.[108] Such prohibitions bar these States' fulfilment of their obligations to prevent and combat racial discrimination, and they should adopt reforms. An example of a recent positive development in this regard is the Race Disparity Audit of the United Kingdom.[109]

50.   The elimination of racial discrimination, as mandated by international human rights law, requires an intersectional analysis. The following definition of intersectionality captures well its significance:

---

[104]   A/HRC/32/50, paras. 10–14.

[105]   See art. 2 (2).

[106]   Committee on the Elimination of Racial Discrimination, general recommendation No. 32 (2009) on the meaning and scope of special measures in the Convention, paras. 6–7.

[107]   See A/HRC/42/59.

[108]   See https://fra.europa.eu/en/publication/2018/bigdata-discrimination-data-supported-decision-making.

[109]   A/HRC/41/54/Add.2, paras. 16–19.

The idea of "intersectionality" seeks to capture both the structural and dynamic consequences of the interaction between two or more forms of discrimination or systems of subordination. It specifically addresses the manner in which racism, patriarchy, economic disadvantages and other discriminatory systems contribute to create layers of inequality that structures the relative positions of women and men, races and other groups. Moreover, it addresses the way that specific acts and policies create burdens that flow along these intersecting axes contributing to create a dynamic of disempowerment.[110]

51. The Committee on the Elimination of Racial Discrimination has clarified that the International Convention on the Elimination of All Forms of Racial Discrimination applies to multiple and intersecting forms of discrimination.[111] Furthermore, application of the Convention's prohibition on racial discrimination should be pursued alongside the Convention on the Elimination of All Forms of Discrimination against Women (art. 1), the Convention on the Rights of Persons with Disabilities (art. 2) and the United Nations Declaration on the Rights of Indigenous Peoples (art. 2), which similarly prohibit or condemn direct and indirect forms of discrimination.

52. **States should simultaneously seek to combat other forms of discrimination that intersect with race and ethnicity, and State obligations should be understood to require collection and analysis of disaggregated data that enable a better understanding of the human rights situation of groups and persons subject to multiple and intersecting structures of discrimination. In the context of emerging digital technologies, this means that anti-racial discrimination interventions must include meaningful attention to gender, disability status and other protected categories.** In a recent report, the Working Group on Experts of People of African Descent provides an example of the nature of intersectional analysis that is essential in this area.[112]

B. Obligations to prevent and combat racial discrimination in the design and use of emerging digital technologies

53. The International Convention on the Elimination of All Forms of Racial Discrimination articulates a number of general State obligations that must be brought to bear in the specific context of emerging digital technologies. It establishes a legal commitment for all States parties to engage in no act or practice of racial discrimination against persons, groups of persons or institutions and to ensure that all public authorities and public institutions, national and local, shall act in conformity with this obligation. Instead, States parties must pursue by all appropriate means and without delay a policy of eliminating racial discrimination in all its forms.[113] The Convention also requires States parties to take effective measures to review governmental, national and local policies, and to amend, rescind or nullify any laws and regulations which have the effect of creating or perpetuating racial discrimination wherever it exists. [114] Furthermore, when the circumstances so warrant, States parties shall take, in the social, economic, cultural and other fields, special and concrete measures to ensure the adequate development and protection of certain racial groups or individuals belonging to them, for the purpose of guaranteeing them the full and equal enjoyment of human rights and fundamental freedoms.[115]

54. Under article 7 of the Convention, States have undertaken to adopt immediate and effective measures, particularly in the fields of teaching, education, culture and information, with a view to combating prejudices which lead to racial discrimination. In other recent reports, the Special Rapporteur has articulated the human rights obligations that States have

---

[110] See www.un.org/womenwatch/daw/csw/genrac/report.htm.
[111] Committee on the Elimination of Racial Discrimination, general recommendation No. 32.
[112] See A/HRC/42/59.
[113] Art. 2 (1) (a).
[114] Art. 2 (1) (c).
[115] Art. 2 (2).

to combat racist and xenophobic speech and conduct, including online.[116] These obligations apply equally to the issues analysed in the present report: in the context of emerging digital technologies, States must take effective measures to detect and combat racially discriminatory design and use of such technologies in access to civil, political, economic, social and cultural rights.[117]

55.   **States' obligations to prevent and eliminate racial discrimination in the design and use of emerging digital technologies require addressing the "diversity crisis" in the sectors discussed in part II above. States must work together with private corporations, including on the basis of legally binding frameworks, to develop the necessary special measures to ensure that racial and ethnic minorities are meaningfully represented in all aspects of decision-making relating to the design and use of emerging digital technologies. A genuine shift in power is required in the various sectors of emerging digital technologies, and not merely tokenism of women and racial and ethnic minority groups. Central to shifting power – even within the private sector – are deeper engagement with and funding for research and knowledge production that specifically aim to deepen understanding of discrimination in the design and use of emerging digital technologies from an interdisciplinary perspective.** Researchers affiliated with the Center for Critical Race and Digital Studies offer examples.[118]

56.   **States must take swift and effective action to prevent and mitigate the risk of the racially discriminatory use and design of emerging digital technologies, including by making racial equality and non-discrimination human rights impact assessments a prerequisite for the adoption of systems based on such technologies by public authorities. These impact assessments must incorporate meaningful opportunity for co-design and co-implementation with representatives of racially or ethnically marginalized groups. A purely or even mainly voluntary approach to equality impact assessments will not suffice; a mandatory approach is essential**. Recent developments in this direction from the Council of Europe,[119] for example, are laudable. They must neither neglect racial discrimination nor exclude racial and ethnic minorities from decision-making. Sometimes prevention of racially discriminatory outcomes and other human rights violations in the design and use of emerging digital technologies by public authorities may require outright bans on their use until the risk of their harm is sufficiently mitigated. The city of San Francisco's decision to ban local government use of facial recognition software is a good example in this regard.

57.   **In order to comply with their equality and non-discrimination obligations, States must ensure transparency and accountability for public sector use of emerging digital technologies, and enable independent analysis and oversight, including by only using systems that are auditable.** Recent reforms by Canada to implement public sector accountability for the use of emerging digital technologies provide examples of an important first step in this regard.[120]

58.   **States must ensure that ethical frameworks and guidelines developed to provide flexible, practical and effective regulation and governance of emerging digital technologies are grounded in legally binding international human rights principles, including those prohibiting racial discrimination. The Toronto Declaration on protecting the right to equality and non-discrimination in machine learning systems exemplifies the symbiotic relationship that should exist between binding international human rights law and ethical guidelines or principles for artificial intelligence governance.**[121] The Toronto Declaration stresses the binding nature of equality and non-

---

[116] See A/73/305, A/73/312 and A/HRC/38/53.

[117] Article 5 of the Convention makes explicit the application of the prohibition on racial discrimination to access and enjoyment of civil, political, social, economic and cultural rights.

[118] See https://criticalracedigitalstudies.com.

[119] Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, 8 April 2020.

[120] See www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592.

[121] See www.torontodeclaration.org/declaration-text/english.

discrimination under international human rights law and provides actionable guidelines for practical implementation of these laws.

Private corporations, and the United Nations and other multilateral bodies

59.   Although international human rights law is only directly legally binding on States, in order to discharge their legal obligations in this regard, States are required to ensure effective remedies for racial discrimination attributable to private actors, including corporations.[122] Under the International Convention on the Elimination of All Forms of Racial Discrimination, States must enact special measures to achieve and protect racial equality throughout the public and private spheres.[123] This should include close regulatory oversight of companies involved in emerging digital technologies.

60.   As articulated in the Guiding Principles on Business and Human Rights, private companies bear a responsibility to respect human rights, including through human rights due diligence. Human rights due diligence requires: assessing actual and potential human rights impacts; integrating and acting upon the findings; tracking responses; and communicating how these impacts are addressed.[124] As highlighted in the Business and Human Rights in Technology Project (B-Tech Project), which applies the Guiding Principles to digital technologies, due diligence should apply to the conceptualization, design and testing phases of new products – as well as the underlying data sets and algorithms that support them.[125] The Toronto Declaration identifies three core elements or steps for corporate human rights diligence for machine learning systems: (a) identification of potential discriminatory outcomes; (b) prevention and mitigation of discrimination and tracking of responses; and (c) transparency regarding efforts to identify, prevent and mitigate discrimination. As highlighted in a recent report, preventive human rights due diligence approaches must be built "in multi-disciplinary teams that can identify blind-spots in AI and find systemic biases in context-specific environments along all lifecycle stages, starting in product development".[126]

61.   **States must ensure that human rights ethical frameworks for corporations involved in emerging digital technologies are linked with and informed by binding international human rights law obligations, including on equality and non-discrimination**. There is a genuine risk that corporations will reference human rights liberally for the public relations benefits of being seen to be ethical, even in the absence of meaningful interventions to operationalize human rights principles. Although references to human rights, and even to equality and non-discrimination, proliferate in corporate governance documents,[127] these references alone do not ensure accountability. **Similarly, implementation of the framework of Guiding Principles on Business and Human Rights, including through initiatives such as the B-Tech Project, must incorporate legally binding obligations to prohibit – and provide effective remedies for – racial discrimination.**

62.   An inherent problem with the ethics-based approaches that are promulgated by technology companies is that ethical commitments have little measurable effect on software development practices if they are not directly tied to structures of accountability in the

---

[122] See, e.g., Human Rights Committee, general comment No. 31 (2004) on the nature of the general legal obligation imposed on States parties to the Covenant, para. 8.

[123] Committee on the Elimination of Racial Discrimination, general recommendation No. 32, para. 23. See also Committee on Economic, Social and Cultural Rights, general comment No. 20 (2009) on non-discrimination in economic, social and cultural rights, para. 11; and Committee on the Elimination of Racial Discrimination, general recommendation No. 29 (2002) on descent in the context of article 1 (1) of the Convention, sect. 7.

[124] A/HRC/17/31, annex, principle 17.

[125] See www.ohchr.org/Documents/Issues/Business/B-Tech/B_%20Tech_Project_ revised_scoping_final.pdf.

[126] See www.institut-fuer-menschenrechte.de/fileadmin/user_upload/Publikationen/ANALYSE/ Analysis_Business_and_Human_Rights_in_the_Data_Economy.pdf.

[127] See https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518482.

workplace.[128] From a human rights perspective, relying on companies to regulate themselves is a mistake, and an abdication of State responsibility. The incentives for corporations to meaningfully protect human rights (especially for marginalized groups, which are not commercially dominant) can stand in direct opposition to profit motives. When the stakes are high, fiduciary obligations to shareholders will tend to matter more than considerations concerning the dignity and human rights of groups that have no means of holding these corporations to account. Furthermore, even well-intentioned corporations are at risk of developing and applying ethical guidelines using a largely technological lens, as opposed to the broader society-wide, dignity-based lens of the human rights framework.

63. **States must rely upon international human rights law prohibitions on racial discrimination in ensuring corporate human rights due diligence.** An example of a promising development has been proposed by the European Commission concerning mandatory due diligence for companies,[129] if such a requirement can ensure substantive human rights implementation and enforcement.

C. Obligations to provide effective remedies for racial discrimination in the design and use of emerging digital technologies

64. The international human rights system operates from the premise that violations of international human rights law incur an obligation on violators to provide adequate and effective remedy to victims of those violations.[130] Victims of human rights violations, including racially discriminatory violations, hold a corresponding right to full remediation, including through judicially or governmentally determined reparations. Article 6 of the International Convention on the Elimination of All Forms of Racial Discrimination is clear in this regard: States parties shall assure to everyone within their jurisdiction effective protection and remedies, through the competent national tribunals and other State institutions, against any acts of racial discrimination which violate human rights and fundamental freedoms contrary to the Convention, as well as the right to seek from such tribunals just and adequate reparation or satisfaction for any damage suffered as a result of such discrimination. This requirement arises because, for rights to have meaning, effective remedies must be available to redress violations.

65. **In the context of effective remedies for racial discrimination in the design and use of emerging digital technologies, States must ensure the full spectrum of effective remedies, including access to justice, protection against possible violations, and guarantees of cessation and non-recurrence of violations, while also combating impunity.[131] The Basic Principles and Guidelines on the Right to a Remedy and Reparation for Victims of Gross Violations of International Human Rights Law and Serious Violations of International Humanitarian Law set out five main elements of remedy and reparations for human rights violations: restitution, compensation, rehabilitation, satisfaction and guarantees of non-repetition.[132] Each of these elements plays a different role in ensuring a holistic and effective remedy, one closely related to the notion of transitional justice.[133]**

66. Restitution aims to restore the victim to the original situation before the gross violations of international human rights law occurred.[134] Compensation entails the payment for economically assessable damage, including physical and mental harms, lost social

---

[128] See https://ainowinstitute.org/AI_Now_2018_Report.pdf.

[129] See https://responsiblebusinessconduct.eu/wp/2020/04/30/european-commission-promises-mandatory-due-diligence-legislation-in-2021.

[130] See, e.g., International Covenant on Civil and Political Rights, art. 2; and International Convention on the Elimination of All Forms of Racial Discrimination, art. 6.

[131] Committee on the Elimination of Racial Discrimination, general recommendation No. 31 (2005) on the prevention of racial discrimination in the administration and functioning of the criminal justice system; and Human Rights Committee, general comment No. 31.

[132] General Assembly resolution 60/147, annex, para. 18.

[133] A/69/518, para. 20.

[134] General Assembly resolution 60/147, annex, para. 19.

benefits, material damages, moral damage and costs incurred.[135] Rehabilitation includes provision of medical and psychological care as well as legal and social services.[136] Satisfaction is a wide-ranging element of reparations and remedies. Where appropriate, satisfaction may encompass measures to stop violations, disclose truth, restore dignity, accept responsibility, memorialize harms, and ensure sanctions against responsible parties.[137] Lastly, guarantees of non-repetition are measures of reparations and remedies that contribute to non-recurrence. These are most closely associated with structural reform and strengthening of State institutions, ensuring sufficient civilian oversight and proper respect for human rights.[138]

67. **States must ensure restitution, compensation, rehabilitation, satisfaction and guarantees of non-repetition to victims of racial discrimination in the design and use of emerging digital technologies. States should also refer to the guidance of the Special Rapporteur on the promotion of truth, justice, reparation and guarantees of non-recurrence, on the formulation and implementation of reparations measures, and to the guidance of the Expert Mechanism on the Rights of Indigenous Peoples.[139]**

68. **The existing human rights framework on remedies and reparations is also an important resource for private corporations committed to combating racial discrimination in the use and design of emerging digital technologies.** In other contexts, private actors have played an important role in providing reparations for racial discrimination, including by taking responsibility for their role in such discrimination.[140] **Private corporations such as Microsoft, Apple, Amazon, Google, Facebook, Tencent and Alibaba have an important role to play in providing restitution, compensation, rehabilitation, satisfaction and guarantees of non-repetition to victims of racial discrimination related to their technologies and products.**

---

[135] Ibid., para. 20.
[136] Ibid., para. 21.
[137] Ibid., para. 22.
[138] Ibid., para. 23.
[139] See A/HRC/EMRIP/2019/3.
[140] A/74/321, para. 62.