United Nations A/HRC/56/68

General Assembly

Distr.: General
3 June 2024

Original: English

Human Rights Council
Fifty-sixth session
18 June–14 July 2024
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

Contemporary forms of racism, racial discrimination, xenophobia and related intolerance

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

Summary

In the present report, the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance, Ashwini K.P., summarizes the activities that she has undertaken over the past year and addresses the ways in which the predominant assumption that technology is objective and neutral is allowing artificial intelligence to perpetuate racial discrimination. She examines four cross-cutting ways in which artificial intelligence can contribute to manifestations of racial discrimination: data problems, algorithm design issues, the intentionally discriminatory use of artificial intelligence and accountability issues. She then provides examples of the application of artificial intelligence across various societal domains and its racially discriminatory impacts. She analyses emerging efforts to manage and regulate artificial intelligence before providing an overview of the relevant international human rights law standards. She concludes by presenting recommendations on how States should approach the management and regulation of artificial intelligence technologies to prevent and address racial discrimination.

* Agreement was reached to publish the present document after the standard publication date owing to circumstances beyond the submitter's control.

I. Introduction

1. The present report is submitted pursuant to Human Rights Council resolution 52/36, in which the Council requested the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance to submit to it an annual report. In the report, the Special Rapporteur describes her activities carried out under the mandate and explores the topic of artificial intelligence (AI) and racial discrimination.

2. To inform the report, the Special Rapporteur issued a call for submissions addressed to States Members of the United Nations and other stakeholders, including civil society organizations, international organizations and national human rights institutions. The Special Rapporteur extends her sincere gratitude to all Member States and other stakeholders who submitted information. She has drawn upon the input that she received, in the preparation of the present report, and remains open to an ongoing dialogue with all relevant stakeholders on this important topic.[1]

II. Summary of activities

3. In October 2023, the Special Rapporteur presented her reports on combating glorification of Nazism, neo-Nazism and other practices that contribute to fuelling contemporary forms of racism, racial discrimination, xenophobia and related intolerance and on online racist hate speech to the General Assembly at its seventy-eighth session.[2] Between 31 October and 14 November 2023, the Special Rapporteur undertook a country visit, to the United States of America.[3]

4. The Special Rapporteur took part in the ninth session of the Group of Independent Eminent Experts on the Implementation of the Durban Declaration and Programme of Action in August 2023. In January 2024, the Special Rapporteur attended the regional meeting for Asia and the Pacific on the International Decade for People of African Descent. In February 2024, she attended the International Conference on Food Justice from a Human Rights Perspective on the theme “Challenges of reality and future stakes”, held in Qatar. In April 2024, she attended the third session of the Permanent Forum on People of African Descent, where she gave a presentation on overcoming systemic racism and historical harm in education.

III. Artificial intelligence and racial discrimination

5. In the present report, the Special Rapporteur has chosen to focus on artificial intelligence and racial discrimination. This topic aligns with her strategic focus on the nexus between digital technologies and racial discrimination, as outlined in her report to the Human Rights Council, at its fifty-third session, setting out her strategic vision and initial priorities.[4] She builds on the work of the previous mandate holder on emerging digital technologies and racial discrimination[5] and responds to the interest of the Human Rights Council and the broader United Nations system in the governance of artificial intelligence.[6]

6. Recent developments in generative artificial intelligence and the burgeoning application of artificial intelligence continue to raise serious human rights issues, including

[1] The Special Rapporteur received research and analysis support from the International Human Rights Clinic of Harvard Law School and from the International Human Rights and Conflict Resolution Clinic and Stanford Center for Racial Justice of Stanford Law School. She sincerely thanks all those involved for their invaluable support in the preparation of the report.

[2] A/78/302 and A/78/538.

[3] See A/HRC/56/68/Add.1.

[4] A/HRC/53/60, paras. 50–53.

[5] See A/75/590, A/HRC/44/57 and A/HRC/48/76.

[6] See, for example, High-level Advisory Body on Artificial Intelligence, “Interim report: governing AI for humanity” (December 2023); Human Rights Council resolution 53/29; and General Assembly resolutions 78/213 and 78/265.

concerns about racial discrimination. Generative artificial intelligence is changing the world and has the potential to drive increasingly seismic societal shifts in the future. The rapid spread of the application of artificial intelligence across various fields is of deep concern to the Special Rapporteur. This is not because artificial intelligence is without potential benefits. In fact, it presents possible opportunities for innovation and inclusion. Such technologies are, however, growing and evolving with largely unbridled momentum. The Special Rapporteur is concerned that policy and legal measures that seek to manage and regulate artificial intelligence are not keeping pace with the growth of this technology and that emerging efforts to govern and regulate it are insufficiently attentive to its huge current capacity and future potential to both perpetuate and deepen systemic racial discrimination, as well as to widen inequality within and between regions, countries and communities.

7. As articulated by the previous mandate holder, there is an enduring and harmful notion that technology is neutral and objective:

> 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.[7]

8. In the present report, the Special Rapporteur addresses the ways in which the predominant assumption that technology is objective and neutral is allowing artificial intelligence to perpetuate racial discrimination.

A. Cross-cutting ways in which artificial intelligence can contribute to manifestations of racial discrimination

9. Artificial intelligence is not a monolith. Indeed, there are several types. Predictive artificial intelligence is considered a "traditional" form of the technology, and the models use historical data, patterns and trends to make informed predictions about future events or outcomes.

10. Artificial intelligence that is used to identify printed characters, human faces, objects and other information is another form of "traditional" artificial intelligence and encompasses various technologies for recognizing and distinguishing objects, individuals and patterns in the data with which it is provided.

11. Generative artificial intelligence systems are newer forms of artificial intelligence. They are versatile and can be used for a range of purposes. They encompass a class of artificial intelligence systems designed to produce diverse outputs on the basis of extensive training data sets, neural networks, deep learning architecture and user prompts. Generative artificial intelligence models can produce a wide range of output, including images, text, audio, video and synthetic data. Unlike artificial intelligence models that are focused on identifying patterns in existing data, generative artificial intelligence is trained to create new data points that mimic the patterns in and characteristics of the data used to train machine learning models. The advent of generative artificial intelligence will lead to many new applications, as well as many new human rights questions.[8]

12. These different types of artificial intelligence have a multitude of applications. The Special Rapporteur elaborates on more specific examples of the uses of artificial intelligence, and related implications in terms of racial discrimination, below. She wishes to stress, however, that it is very important to examine commonalities in the ways in which artificial intelligence can perpetuate racial discrimination, particularly within legal and policy debates relating to its management and regulation. In such debates, the effects of artificial intelligence must be viewed through the lens of systemic racism, defined as the "operation of a complex,

[7] A/HRC/44/57, para. 12.

[8] Australia Human Rights Commission submission. All submissions will be posted on the website of the Office of the United Nations High Commissioner for Human Rights (OHCHR).

interrelated system of laws, policies, practices and attitudes in State institutions, the private sector and societal structures that, combined, result in direct or indirect, intentional or unintentional, de jure or de facto discrimination, distinction, exclusion, restriction or preference on the basis of race, colour, descent or national or ethnic origin".[9] As reflected in that definition, systemic racism is a complex, often insidious and society-wide phenomenon. Manifestations of systemic racism in one domain are interrelated, interdependent and mutually reinforcing with those in others. Looking at the cross-cutting ways in which artificial intelligence contributes to racial discrimination can help to identify the ways in which it interacts with and reinforces manifestations of systemic racism and holistically reinforces systemic oppression in society along racial and ethnic lines.[10]

1. Data problems

13. The rise of artificial intelligence systems and machine learning algorithms has led to the digitization of data on a massive scale. Algorithms use those data to make decisions and engage in actions across several sectors. However, the data sets on which algorithms are trained are often incomplete or underrepresent certain groups of people. If particular groups are over- or underrepresented in the training sets, including along racial and ethnic lines, algorithmic bias can result. Similarly, if the training sets include already biased data, they can produce biased outcomes.

14. If the training data are insufficient, the algorithms may make predictions that are systematically discriminatory for groups that are unrepresented or underrepresented in the data. Not only can algorithmic bias occur with too little data; algorithms based on unrepresentative data can also produce skewed outcomes. For instance, a study focused on law enforcement image databases in the United States showed that people of African descent were more likely to be erroneously singled out in facial recognition networks used by law enforcement officers. This was due to errors in facial identification for that group and the overrepresentation of people of African descent in police photograph databases, which reflects historical patterns of systemic racism.[11]

15. Historical biases can affect the data themselves. A core element of machine learning is making predictions about the future on the basis of data from the past. However, if past data are biased against certain groups, including along racial and ethnic lines, the computer models can reproduce and amplify those biases. The use of biased or flawed data to inform real life decisions can further target and harm marginalized racial and ethnic groups because the use of those data in the context of artificial intelligence creates more data, which are then used to inform future decisions. Such self-reinforcing systems can replicate and deepen existing disparities.

16. The final issue with data is privacy. The data used in artificial intelligence systems often include the personal information of the individuals to whom the data belong. The collection and processing of data without consent violates the right to privacy. There are also incidents of data collected in one setting, such as health care, including through the use of health-care applications, being shared, without consent, for use in others, such as for law enforcement purposes. Data breaches and unauthorized access to personal information through hacking pose additional privacy concerns. For those from racially marginalized groups, human rights concerns relating to the right to privacy can be amplified. Privacy violations can put those groups at risk of ostracization, discrimination or physical danger.[12]

2. Algorithm design problems

17. A second common form of bias in artificial intelligence tools arises from the way in which algorithms are designed. If bias is embedded in design choices, an algorithm can contribute to biased outcomes, even if the data fed into the algorithm are perfectly

[9] A/HRC/47/53, para. 9.
[10] A/HRC/44/57, para. 43.
[11] Nicol Turner Lee, Paul Resnick and Genie Barton, "Algorithmic bias detection and mitigation: best practices and policies to reduce consumer harms", Brookings Institution, 22 May 2019.
[12] Samantha Lai and Brooke Tanner, "Examining the intersection of data privacy and civil rights", Brookings Institution, 18 July 2022. See also Privacy International submission.

representative. Decisions about the parameters and functioning of an algorithm can introduce biases. Algorithm designers make decisions about which variables an algorithm will use, how to define categories or thresholds for sorting information and what data will be used to build the algorithm. The choices made by designers include how to measure specific features and define algorithmic success. Sometimes, the backgrounds or perspectives of algorithm designers may cause them to embed unconscious biases, including racial biases, in their algorithm designs.[13] This lack of diversity in digital technology sectors is reportedly exacerbated by the absence of inclusive consultation processes in the development of artificial intelligence systems, which contributes to algorithmic design issues.[14]

18. Algorithmic design choices can have significant discriminatory impacts in real life. For example, when building a loan risk assessment algorithm, the way in which "risk" is defined and measured may lead to discriminatory results. If an algorithm designer decides to use credit scores as a proxy for risk, there could be discriminatory outcomes for groups of people who tend to have lower credit scores. Research has shown that there can be a strong correlation between credit score, race and other demographic indicators and that the use of credit scores disadvantages certain groups.[15] That correlation can, in many cases, be seen as a by-product of existing systemic racism and exclusion. Individuals may be disadvantaged by the choice made by an algorithm designer to use credit scores to assess loan risk, despite it ostensibly not being a discriminatory criterion.

3. Use for discriminatory purposes

19. Artificial intelligence can, in some cases, be used for explicitly racist purposes through its selective deployment against targeted groups, resulting in discriminatory outcomes. For example, there are reports of law enforcement agencies intentionally using artificial intelligence to survey and overpolice particular communities, along racially discriminatory lines.[16] Furthermore, intentional discrimination can occur when Governments and others exploit the technology's capabilities to monitor, profile and target specific groups or individuals on the basis of their racial or ethnic identities.[17]

20. The spread of disinformation is another way in which artificial intelligence can be used for explicitly racist purposes. Political actors can use artificial intelligence to generate texts, images and videos to manipulate public opinion and political processes in their favour and undermine trust in institutions, including along racial lines. Governments are also reported to have used artificial intelligence to sow discord and facilitate online censorship.[18]

4. Accountability problems

21. The fact that some artificial intelligence tools make decisions independently of humans means that the decision-making process is hidden, as if in an opaque "black box". In addition, an algorithm might make decisions independently because, once exposed to data, artificial intelligence algorithms are constantly updating themselves. Over time, an artificial intelligence tool may use, in its decision-making, factors on which it was not originally programmed to rely. Instead, these factors come from patterns that it has itself identified in the data. As the algorithm incorporates these new patterns into its code and decision-making, individuals relying on the algorithm may no longer be able to "look under the hood" and pinpoint the criteria that the algorithm has used to produce certain outcomes. Thus, the "black

[13] Ninareh Mehrabi and others, "A survey on bias and fairness in machine learning", *ACM Computing Surveys*, vol. 54, No. 6 (2022); The London Story submission; and A/HRC/44/57, para. 17.

[14] NetMission.Asia submission.

[15] A.R. Lange and Natasha Duarte, "Understanding bias in algorithmic design", Medium, 6 September 2017.

[16] See Amnesty International, *Decode Surveillance NYC: Methodology* (London, 2022); and NetMission.Asia submission.

[17] NetMission.Asia submission.

[18] Tate Ryan-Mosley, "How generative AI is boosting the spread of disinformation and propaganda", *MIT Technology Review*, 4 October 2023.

box" problem makes the artificial intelligence reasoning process insidious and opaque.[19] In addition, many algorithms developed by business entities cannot be scrutinized because of contract and intellectual property laws, exacerbating accountability issues.[20]

22. The "black box" problem has particularly concerning implications in the context of systemic racism. As described above, systemic racism is an insidious but deeply destructive, society-wide scourge. The forces driving systemic racism are not always recognized, a phenomenon that is exacerbated by persistent gaps in the collection of racially and ethnically disaggregated data.[21] Artificial intelligence, without effective accountability mechanisms, has significant capacity to be an additional driver of the already insidious and destructive phenomenon of systemic racism.

23. Artificial intelligence accountability issues have significant implications for the ability of those who experience acts of racial discrimination to seek effective remedies. Today, when those from marginalized racial and ethnic groups experience different outcomes because of human decision-making, courts and other accountability mechanisms can examine whether the actions were intentional and justifiable.[22] When people are the decision-makers, there is often evidence that can be used to make such assessments. In many cases, autonomous decision-making processes do not create evidentiary trails in the same way as human decision makers.[23] "Black box" issues will exacerbate the already significant barriers in access to justice for those who experience racial discrimination.

B. Use of artificial intelligence and its discriminatory impact

24. In the present section, the Special Rapporteur provides examples of the uses of artificial intelligence across different societal domains and its racially discriminatory impacts. These examples are illustrative and non-exhaustive and are provided as clear evidence that artificial intelligence is already contributing to racial discrimination. The Special Rapporteur perceives these examples as interconnected and mutually reinforcing manifestations of racial discrimination, which contribute to the holistic reinforcement of systemic, society-wide oppression, along racial and ethnic lines.

25. The Special Rapporteur has chosen three domains to exemplify the discriminatory impact of artificial intelligence: law enforcement, security and the criminal justice system; education; and health care. In relation to the use of artificial intelligence in other contexts, the Special Rapporteur recommends consulting the reports of the previous mandate holder on the rise of digital borders and mapping racial and xenophobic discrimination in digital border and immigration enforcement and on the use of digital technologies in border and immigration enforcement.[24] The Special Rapporteur also refers readers to her report to the General Assembly, at its seventy-eighth session, on online racist hate speech, which addresses the use of artificial intelligence in social media content moderation,[25] and to the report of the Special Rapporteur on extreme poverty and human rights to the General Assembly at its seventy-fourth session, which provides an analysis of the use of artificial intelligence in social protection systems.[26]

[19] Yavar Bathaee, "The artificial intelligence black box and the failure of intent and causation", *Harvard Journal of Law and Technology*, vol. 31, No. 2 (2018); A/HRC/44/57, para. 34; and Renata M. O'Donnell, "Challenging racist predictive policing algorithms under the Equal Protection Clause", *New York University Law Review*, vol. 94, No. 3 (June 2019).

[20] A/HRC/44/57, para. 44.

[21] A/HRC/47/53, para. 16.

[22] Bathaee, "The artificial intelligence black box".

[23] Ibid.

[24] A/75/590 and A/HRC/48/76.

[25] A/78/538.

[26] A/74/493.

1. Law enforcement, security and the criminal justice system

(a) Automated identification

26. Law enforcement agencies use automated identification tools to connect what they observe in a particular environment to a potential "match" in a database. One of the most common types of automated identification tools is facial recognition technology. Facial recognition tools take video footage or photographs of a person and feed them into algorithms. The algorithms then compare the images against a database of police photographs, driver's licence photographs or other images with the goal of identifying the person.[27] The designers of such tools train the models on which they are based by showing them images of faces, through a process of machine learning. The goal is to train the models to identify the distinguishing features of human faces.[28] However, the image data sets used to train these models are not always demographically representative.[29] In one study of a popular image database, researchers found an overrepresentation of men between the ages of 18 and 40 and an underrepresentation of people with dark skin.[30] According to another study of commercially released facial recognition systems, gender classification algorithms are trained on data sets with overwhelmingly white male faces.[31] The lack of racial, gender and cultural diversity in artificial intelligence tools' training sets leads to one of the classic data problems described above. Groups that are underrepresented in the training data, including those that experience intersectional forms of discrimination, are more likely to be erroneously matched by the algorithm.

27. It has been reported that the misidentification of faces by these technologies has led to an increased number of arrests of people of African descent.[32] The Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression and the United Nations High Commissioner for Human Rights have noted that facial recognition tools often contribute to unlawful discrimination and racial profiling.[33] Despite such human rights concerns, facial recognition systems have been deployed by law enforcement agencies in a number of countries. For example, the Government of India has reportedly invested significantly in such systems. The facial recognition system used by the Delhi Police was reported to be accurate in only 2 per cent of cases and to put minority communities at a disproportionate risk of misidentification and false arrest.[34] Brazilian law enforcement officials have reportedly falsely accused and arrested individuals on the basis of faulty facial recognition tools. According to a 2019 study, 90 per cent of people arrested in Brazilian cities on the basis of facial recognition technology are of African descent.[35]

28. Gunshot detection systems are another common type of automated identification tool used by law enforcement officials in a number of countries. One system, named ShotSpotter, involves placing sensors containing a microphone, a GPS system, memory and processing, and cell capability in neighbourhoods.[36] When the sensors detect a noise that could be a gunshot, an algorithm triangulates the location of whatever caused the noise. The algorithm

[27] Marissa Gerchick and Matt Cagle, "When it comes to facial recognition, there is no such thing as a magic number", American Civil Liberties Union, 7 February 2024.

[28] Julia Dressel and Andrew Warren, "Breaking down data analytics and AI in criminal justice", Recidiviz, 8 March 2022.

[29] AI for the People submission.

[30] Khari Johnson, "ImageNet creators find blurring faces for privacy has a 'minimal impact on accuracy'", VentureBeat, 16 March 2021.

[31] Joy Buolamwini and Timnit Gebru, "Gender shades: intersectional accuracy disparities in commercial gender classification", *Proceedings of Machine Learning Research*, vol. 81 (2018). See also Gerchick and Cagle, "When it comes to facial recognition, there is no such thing as a magic number"; AI for the People submission; and Internet Lab submission.

[32] Gerchick and Cagle, "When it comes to facial recognition, there is no such thing as a magic number".

[33] See A/HRC/41/35 and A/HRC/48/31.

[34] Amnesty International, "Ban the scan: Hyderabad", available at https://banthescan.amnesty.org/hyderabad/.

[35] Group of experts from Brazil submission.

[36] Alisha Ebrahimji, "Critics of ShotSpotter gunfire detection system say it's ineffective, biased and costly", CNN, 24 February 2024.

filters out other possible sounds before sending the audio to a person for review.[37] The available information suggests that gunshot detection systems are deployed disproportionately in communities inhabited by racially marginalized groups,[38] and they can have a very high error rate. The placement of gunshot detection systems in communities in which marginalized racial and ethnic groups live and the inaccuracies of gunshot detection systems exacerbate systemic biases within law enforcement.

29. There are many examples of how the use of automated identification technology has had life-altering consequences. In 2019, in the United States, a Black man in New Jersey was reportedly falsely arrested and held in jail for 10 days because of a facial recognition error. Despite the existence of exonerating evidence, the authorities did not drop his case for almost a year, and he faced up to 25 years of imprisonment for the charges brought against him. The incident had a significant impact on the man's life.[39] In February 2024, law enforcement officers in Chicago reportedly opened fire on a child who was lighting fireworks after responding to a false alert from ShotSpotter.[40] Another example of the use of this type of artificial intelligence technology is the reported adoption by the Israel Defense Forces of Wolf Pack, a vast database containing images and all available information on Palestinians from the West Bank, further integrating various surveillance programmes such as Blue Wolf and Red Wolf.[41] Across the Old City of Hebron, the Israel Defense Forces reportedly installed artificial intelligence-powered cameras capable of identifying human faces, which are connected to the Blue Wolf programme, a mobile application that allows soldiers to detect and categorize Palestinians across the West Bank by means of an extensive biometric database in which most have not consented to enrol, resulting in ongoing surveillance of Palestinians. The rigorous application of the Wolf Pack system by the Israel Defense Forces exacerbates the apartheid perpetuated against Palestinians.[42] These examples show the serious human rights implications resulting from the use of artificial intelligence systems to make consequential decisions in high-risk settings.

(b) Predictive policing algorithms

30. Another form of artificial intelligence technology that is commonly used by law enforcement is predictive policing. Predictive policing tools make assessments about who will commit future crimes, and where any future crime may occur, on the basis of location and personal data.

31. Predictive policing can exacerbate the historical overpolicing of communities along racial and ethnic lines.[43] Because law enforcement officials have historically focused their attention on such neighbourhoods, members of communities in those neighbourhoods are overrepresented in police records. This, in turn, has an impact on where algorithms predict that future crime will occur, leading to increased police deployment in the areas in question.[44]

---

[37] Jay Stanley, "Four problems with the ShotSpotter gunshot detection system", American Civil Liberties Union, 24 August 2021.

[38] Ibid.; and MacArthur Justice Center, "ShotSpotter is deployed overwhelmingly in Black and Latinx neighborhoods in Chicago", available at https://endpolicesurveillance.com/burden-on-communities-of-color/.

[39] Gerchick and Cagle, "When it comes to facial recognition, there is no such thing as a magic number"; and Khari Johnson, "How wrongful arrests based on AI derailed 3 men's lives", *Wired*, 7 March 2022.

[40] Adam Schwartz, "Responding to ShotSpotter, police shoot at child lighting fireworks", Electronic Frontier Foundation, 22 March 2024.

[41] Amnesty International, *Automated Apartheid: How Facial Recognition Fragments, Segregates and Controls Palestinians in the OPT* (London, 2023), pp. 41–45.

[42] Sophia Goodfriend, "Algorithmic State violence: automated surveillance and Palestinian dispossession in Hebron's Old City", *International Journal of Middle East Studies*, vol. 55, No. 3 (2023).

[43] Tim Lau, "Predictive policing explained", Brennan Center for Justice, 1 April 2020; and Jon Fasman, "The black box of justice: how secret algorithms have changed policing", *Fast Company*, 9 February 2021.

[44] Kristian Lum and William Isaac, "To predict and serve?", *Significance*, vol. 13, No. 5 (2016); and Australian Human Rights Commission submission.

Predictive policing can also reflect aspects of the "black box" problem, as the algorithms lack transparency, including about what data are analysed and how the predictions are used.[45]

32. Location-based predictive policing algorithms draw on links between places, events and historical crime data to predict when and where future crimes are likely to occur.[46] Police forces then plan their patrols accordingly. When officers in overpoliced neighbourhoods record new offences, a feedback loop is created, whereby the algorithm generates increasingly biased predictions targeting these neighbourhoods. In short, bias from the past leads to bias in the future. In the United Kingdom of Great Britain and Northern Ireland, a Government-commissioned study of algorithmic bias in policing showed that identifying geographical locations as "hotspots" for crime could prime officers to expect more crime in those areas. As a result, the officers were more likely to stop or arrest people in "hotspots" on the basis of bias than on the basis of genuine public safety imperatives.[47] In Uruguay, researchers have found that data used in location-based predictive policing algorithms could be biased. The location variable could function as a proxy for socioeconomic or ethnic background, triggering discrimination.[48]

33. Person-based predictive policing tools provide a way of predicting who might commit a future crime on the basis of background data about individuals. Background data can include a person's age, gender, marital status, history of substance abuse and criminal record. As with location-based tools, past arrest data, which are often tainted by systemic racism in the criminal justice system, can skew the future predictions of those algorithms. The use of variables such as socioeconomic background, education level and location can act as proxies for race and perpetuate historical biases.[49] In Australia, the New South Wales Police Force used the algorithm-based Suspect Target Management Plan to identify individuals at risk of committing criminal offences. Its use reportedly led to a disproportionately high number police interactions with members of Aboriginal and Torres Strait Islander communities before it was discontinued.[50]

(c) Recidivism assessment algorithms

34. Recidivism assessment tools are used to inform decisions at different stages of the criminal justice system, including about bail, bond, sentencing and parole.[51] Recidivism assessment tools use historical data to assess defendants' likelihood of acting in certain ways, in particular whether they are likely to commit a new crime in the future. The tools produce risk scores, using information from sources such as criminal records and defendant surveys.[52]

35. Recidivism prediction tools exhibit multiple artificial intelligence challenges that contribute to racial discrimination. First, the tools have data challenges. The criminal justice system data used to train their algorithms reflect systemic inequities based on a history of racist policing behaviour.[53] In addition, design choices, such as how variables are measured or assessed, can contribute to algorithmic discrimination.[54] Moreover, the way in which an

[45] Lau, "Predictive policing explained".

[46] Will Douglas Heaven, "Predictive policing algorithms are racist. They need to be dismantled", *MIT Technology Review*, 17 July 2020.

[47] Ibid. See also Government of the United Kingdom of Great Britain and Northern Ireland, "Report commissioned by CDEI calls for measures to address bias in police use of data analytics", 16 September 2019.

[48] Juan Ortiz Freuler and Carlos Iglesias, "Algorithms and artificial intelligence in Latin America: a study of implementation by governments in Argentina and Uruguay", World Wide Web Foundation, September 2018; and Eticas Foundation, "Uruguay's Ministry of the Interior invests in predictive policing", 13 September 2021.

[49] Heaven, "Predictive policing algorithms are racist".

[50] Australian Human Rights Commission submission.

[51] Julia Angwin and others, "Machine bias", ProPublica, 23 May 2016.

[52] Ibid.

[53] See Heaven, "Predictive policing algorithms are racist"; and Michael Mayowa Farayola and others, "Fairness of AI in predicting the risk of recidivism: review and phase mapping of AI fairness techniques", in *Proceedings of the 18th International Conference on Availability, Reliability and Security* (Association for Computing Machinery, 2023).

[54] Mehrabi and others, "A survey on bias and fairness in machine learning".

algorithm designer chooses to define "success" can have an impact on what the algorithm seeks to maximize. If an algorithm is set to optimize for a minimum number of new offences, it may correlate longer sentences with lower reoffending rates, because people cannot reoffend while incarcerated. It can then use those patterns to recommend longer sentences.

36. Researchers have suggested that recidivism predictors are not accurate and that their errors have a disproportionate impact on racially marginalized groups. For example, a study in the United States found that risk scores were very unreliable in their forecasting of violent crime. People of African descent were reportedly mislabelled as future criminals at almost twice the rate of white individuals.

(d) Autonomous weapon systems

37. Autonomous weapon systems include any weapon systems with autonomy in their critical functions, including lethal autonomous weapons and less-lethal weapons. They have applications in law enforcement, as well as military, contexts and remain largely unchecked. These systems can select, detect, identify and attack targets without human intervention. An autonomous weapon is triggered by sensors and software that match a person with a "target profile" as determined by the system's algorithm. Autonomous weapon systems have very serious human rights implications, including relating to the right to life, the prohibition of torture and other ill-treatment and the right to security of person.[55]

38. The First Committee of the General Assembly heard that the window of opportunity to enact guardrails against the perils of autonomous weapons and artificial intelligence's military applications was rapidly closing as the world prepared for a "technological breakout".[56] The Special Rapporteur on extrajudicial, summary or arbitrary executions has previously recommended that the Human Rights Council call upon all States to declare and implement national moratoriums on at least the testing, production, assembly, transfer, acquisition, deployment and use of lethal autonomous robotics.[57]

39. There is a serious risk of grave and, in some circumstances, deadly racial discrimination resulting from the use of autonomous weapon systems. The criteria used to select targets likely include gender, age and race.[58] Target profiles also include seemingly neutral criteria, such as weight or heat signatures, but the machines often reflect the biases of their programmers and society. They can also be programmed with intentionally discriminatory target profiles.[59] For example, Israel is reportedly using lethal autonomous and semi-autonomous weapon systems. This reportedly includes the use of remote-controlled quadcopters to target Palestinians, in addition to automated target generation systems, operating at unparalleled speed and volume, to produce "kill lists".[60] The Gospel and Lavender, two artificial intelligence technology systems used by the Israel Defense Forces, are reported to have intensified the levels of destruction in Gaza, resulting in significant causalities, in particular among Palestinian women and children.[61]

[55] Amnesty International, "Autonomous weapons systems: five key human rights issues for consideration" (April 2015), p. 5.

[56] United Nations, "Without adequate guardrails, artificial intelligence threatens global security in evolution from algorithms to armaments, speaker tells First Committee", 24 October 2023.

[57] A/HRC/23/47, para. 113.

[58] Ray Acheson, "Gender and bias", available at https://www.stopkillerrobots.org/wp-content/uploads/2021/09/Gender-and-Bias.pdf.

[59] Bonnie Docherty, "Expert Panel on the Social and Humanitarian Impact of Autonomous Weapons at the Latin American and Caribbean Conference on Autonomous Weapons", Human Rights Watch, 8 March 2023.

[60] Marwa Fatafta and Daniel Leufer, "Artificial genocidal intelligence: how Israel is automating human rights abuses and war crimes", Access Now, 9 May 2024.

[61] Yuval Abraham, "'Lavender': the AI machine directing Israel's bombing spree in Gaza", *+972 Magazine*, 3 April 2024.

2. Health care

(a) Health risk scores

40. Artificial intelligence can be utilized to create health risk scores for a variety of purposes in health care, including medical diagnosis and care planning. Racially discriminatory effects can occur when such algorithms are used to allocate health-care resources, because of algorithmic design and the data used to train artificial intelligence systems. There are cases in which such algorithms have been designed to identify who should qualify for extra care, using previous health-care costs as a proxy for medical needs. The data on which such decisions are based can be influenced by previous lack of adequate access to health care among those from marginalized racial and ethnic groups in the context of systemic racism, as well as different disease patterns due to disparities in the socioeconomic determinants of health.

41. In the United States, a calculator was developed to assist health-care providers in assessing the likelihood of successful vaginal birth after caesarean delivery. A study in 2019 revealed bias in the calculator's foundational algorithm. The calculator had two race-based correction factors, which resulted in lower predicted vaginal birth success rates for women of African descent and Hispanic women compared with white women with similar characteristics. Because of these correction factors, the calculator potentially worsened racial disparities in maternal health outcomes by discouraging clinicians from offering vaginal birth to women of African descent and Hispanic women, leading to higher rates of caesarean sections.[62]

(b) Disease detection

42. Another application of artificial intelligence technologies is disease detection, including cancer detection.[63] Artificial intelligence systems trained on extensive data sets comprising thousands or millions of images, including radiological scans, pathology images and photographs, can learn to distinguish between normal and cancerous lesions.[64] This deployment of artificial intelligence can significantly aid in early cancer detection, potentially saving lives while improving health-care system efficiency. However, those from marginalized racial and ethnic groups may not benefit equally from such advancements due to the algorithms not being generalizable to patient populations that are not adequately represented in the training data. Researchers have suggested that the use of artificial intelligence algorithms for skin cancer detection shows poorer performance for individuals with darker skin tones because many of the publicly available image data sets used to train them are biased, with a lack of diversity in skin tones and ethnic backgrounds.[65] For example, a survey of 21 open-access skin lesion data sets, containing over 100,000 images, revealed a significant underrepresentation of images of darker skin.[66]

(c) Artificial intelligence-enabled medical devices

43. Artificial intelligence is being developed and utilized alongside other developments in health-care technologies, including health-care devices. Many of these devices are artificial intelligence-enabled, and racial bias can become embedded in their operation. For example, in the United Kingdom, a report showed that bias was baked into the operation of medical devices at all stages of their development, including during phases involving algorithm development and machine learning. During the coronavirus disease (COVID-19)

[62] Darshali A. Vyas and others, "Challenging the use of race in the Vaginal Birth after Cesarean Section Calculator", *Women's Health Issues*, vol. 29, No. 3 (2019).

[63] Privacy International submission.

[64] Likhitha Kolla and Ravi B. Parikh, "Uses and limitations of artificial intelligence for oncology", *Cancer*, 30 March 2024.

[65] David Wen and others, "Characteristics of publicly available skin cancer image datasets: a systematic review", *The Lancet Digital Health*, vol. 4, No. 1 (2022).

[66] Ibid. See also Privacy International submission.

pandemic, the use of pulse oximetry devices to measure low oxygen levels in the blood led to overestimations of the levels of oxygen in the blood of people with darker skin tones.[67]

3. Education

(a) Academic and career success algorithms

44. In countries such as Finland and the United States, predictive analytics tools are used in education to determine the likelihood of future success on the basis of data, statistical algorithms and machine learning.[68] The data used in these algorithms include data on attendance, grades, behaviour and online activity. They are designed to help educators to guide students in decisions about their educational and career journeys. While the predictive analytics tools are intended to assist educators in improving outcomes for students, they often rate racial minorities as less likely to succeed academically and in their careers, because of algorithm design and data choices. On the basis of these ratings, educators may steer students from marginalized racial and ethnic groups away from educational and career choices that would maximize their potential and offer the best opportunities to break cycles of exclusion or invest fewer resources in these students.

(b) Grading algorithms

45. Grading algorithms typically use historical grading data to evaluate student performance. Such data can be biased by historical patterns of systemic racism in educational institutions. The bias in the data will be replicated by predictive scoring algorithms for students, especially when teacher input is excluded.[69] Grading algorithms can be hugely consequential in determining the opportunities available to students, including in relation to access to university education or employment opportunities after education. Racially biased automated decisions may therefore limit opportunities for students from marginalized racial and ethnic groups and undercut the potential of education to be a tool to disrupt systemic racism.

46. The United Kingdom provides a cautionary example of the deployment of a grading algorithm. In 2020, Advanced Level (A-level) examinations were cancelled due to the COVID-19 pandemic. As a substitute for examination grades, teachers were asked to predict students' results. The national regulatory agency for grading then deployed an algorithm to standardize the predicted scores on the basis of each school's historical grading data. Forty per cent of students, many of whom attended schools in lower-income areas, had their scores downgraded as a result. Conversely, the algorithm upgraded a disproportionally high number of students from independent, fee-paying schools. The Government responded to the controversy by reversing the algorithm's standardization. However, the episode caused significant disruptions to university admissions processes.[70]

(c) Large language models in education

47. Generative artificial intelligence tools rely on large language models to produce novel content, including text, music, images and videos. Large language models are being used in educational settings and can assist with improving academic outcomes for students of all ages. Studies have shown that language models are biased towards English, which is the most widely used language on the Internet and the language in which most artificial intelligence researchers and technologists work. Moreover, only a handful of the approximately

[67] Privacy International submission.

[68] Stina Westman and others, "Artificial intelligence for career guidance – current requirements and prospects for the future", *International Academic Forum Journal of Education*, vol. 9, No. 4 (2021); and Kelli A. Bird, Benjamin L. Castleman and Yifeng Song, "Are algorithms biased in education? Exploring racial bias in predicting community college student success", *Journal of Policy Analysis and Management*, 31 January 2024.

[69] Benjamin Herold, "Why schools need to talk about racial bias in AI-powered technologies", *Education Week*, 12 April 2022.

[70] Bryan Walsh, "How an AI grading system ignited a national controversy in the U.K.", Axios, 19 August 2020; and Daan Kolkman, "'F**k the algorithm'? What the world can learn from the UK's A-level grading fiasco", London School of Economics Impact Blog, 26 August 2020.

6,000 languages in the world have high-quality data resources that can be used to train artificial intelligence models. To address that gap, companies have begun to develop multilingual language models. However, multilingual models do not perform as well as English-language models. The use of large language models in educational settings could disadvantage students from linguistic backgrounds that are not represented in the underlying data resources, which may have racially disproportionate impacts.[71]

48. There are debates about whether generative artificial intelligence tools based on large language models should be banned among students rather than integrated into curricula. There are also steps in some educational settings to try to restrict the use of generative artificial intelligence tools that rely on large language models among students. Some educational institutions are using artificial intelligence tools to detect the use of artificial intelligence by students. The use of such tools, which may contain algorithmic bias, to patrol cheating may introduce further biases that harm students from marginalized racial and ethnic groups. Such harm is bound to be exacerbated in cases in which institutions have not set up equitable appeals processes.[72]

(d) Facial recognition in educational institutions

49. Facial recognition technologies have been introduced in many educational settings around the world, despite evidence of racial bias in their operation, as described above. Facial recognition systems are being used to automate attendance-taking, to enhance school security, to perform examination proctoring functions and even to record the emotions of children in schools to monitor how much they are learning. This is often without adequate human rights due diligence or regulatory oversight. For example, in Brazil, an increasing number of schools are adopting facial recognition tools to streamline operations, track attendance and enhance security.[73] However, it has been reported that neither municipalities nor states conducted human rights impact assessment studies or analysed the risks of discrimination associated with facial recognition software before implementing these projects.[74]

50. The use of facial recognition software in educational settings is having racially discriminatory impacts. There have been cases, including one reported in the Kingdom of the Netherlands, in which students of African descent have had to shine lights in their faces to be recognized by the artificial intelligence systems used to mediate access to important examinations. Such experiences impact students' equal right to education but also create friction and exclusion when students from marginalized racial and ethnic groups are given the impression that the system was not designed for them. The recording and monitoring of children's emotions in schools has significant privacy implications for all students and can perpetuate racial bias. These systems have been found to interpret the facial expressions of individuals of African descent and white individuals differently, attributing negative feelings, such as contempt and anger, more frequently to those of African descent.[75]

[71] Felix Richter, "The most spoken languages: on the Internet and in real life", Statista, 21 February 2024; Emily M. Bender, "The #BenderRule: on naming the languages we study and why it matters", The Gradient, 14 September 2019; Gabriel Nicholas and Aliya Bhatia, "Lost in translation: large language models in non-English content analysis", Center for Democracy and Technology, 23 May 2023; A. Bergman and Mona Diab, "Towards responsible natural language annotation for the varieties of Arabic", in *The 60th Annual Meeting of the Association for Computational Linguistics: Findings of ACL 2022* (Association for Computational Linguistics, 2022); and BigScience Workshop, "A 176B-parameter open-access multilingual language model" (ArXiv, 2022).

[72] See Regina Ta and Darrell M. West, "Should schools ban or integrate generative AI in the classroom?", Brookings Institution, 7 August 2023; and Robert Topinka, "The software says my student cheated using AI. They say they're innocent. Who do I believe?", *The Guardian*, 13 February 2024.

[73] InternetLab submission.

[74] Ibid.

[75] Ibid.

C. Emerging initiatives to regulate and manage artificial intelligence

51. States have begun to take promising steps to regulate and manage artificial intelligence. In the present section, the Special Rapporteur draws attention to some of these initiatives. Her non-exhaustive analysis is based on submissions from States and civil society groups, as well as on her country-based work and research conducted for the present report.

1. National initiatives

52. States have taken steps to regulate and manage artificial intelligence at the national level through both binding legal provisions and voluntary policy standards, as well as, in many cases, a mix of the two. For example, in Brazil, there are pending legislative provisions on the regulation of the technology space,[76] including artificial intelligence. The Government has also adopted a number of relevant policy documents, such as a document entitled "Racism on the Internet: evidence for the formulation of digital policies", which was reportedly developed by the Ministry of Racial Equality and contains measures to address algorithmic bias, including in relation to race.[77] The Special Rapporteur welcomes efforts to develop dedicated and binding regulatory provisions, complemented by relevant policy standards. However, she received concerning information about the reported lack of effective consultation with and participation of people of African descent in the development of legislative provisions on the regulation of artificial intelligence, as well as a lack of overall coherence and consistency among different standards and the current practices of the State.[78]

53. The United States, which has reportedly taken steps to develop a mix of binding and voluntary standards on the use of artificial intelligence, offers another example. Following her recent visit to the country, the Special Rapporteur welcomed the signing of Executive Order 14110 on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence and the references therein to the risks of bias and discrimination in the use of artificial intelligence. In the preparation of the present report, the Special Rapporteur received further information about artificial intelligence regulation in the United States, including about work on an artificial intelligence bill of rights, as a voluntary standard, efforts in states such as Virginia, California and New York to regulate artificial intelligence and initiatives to facilitate voluntary pledges from business entities to develop safe, secure and transparent artificial intelligence.[79] The Special Rapporteur welcomes these efforts, although she is concerned that, despite robust research on the profound algorithmic racial bias in digital commercial products in the United States context, there is a lack of explicit reference to racial discrimination and bias in Executive Order 14110.[80]

54. Canada is reported to be developing a mix of binding and voluntary standards. The Artificial Intelligence and Data Act is currently in a draft form and is reported to include binding legal controls for high-risk artificial intelligence systems. In addition, Canada has developed voluntary standards, including the Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems. It has also developed the Algorithmic Impact Assessment, a tool designed to help Government departments and agencies to assess and mitigate the risks of artificial intelligence, including those relating to discrimination and bias.[81]

[76] Brazil submission.

[77] Group of experts from Brazil submission.

[78] Ibid.

[79] NetMission.Asia submission. See also Kay Firth-Butterfield, Karen Silverman and Benjamin Larsen, "Understanding the US 'AI Bill of Rights' – and how it can help keep AI Accountable", World Economic Forum, 14 October 2022; United States, Office of Science and Technology Policy of the White House, "Blueprint for an AI bill of rights: making automated systems work for the American people", white paper, October 2022; and United States, White House, "Fact sheet: Biden-Harris Administration secures voluntary commitments from eight additional artificial intelligence companies to manage the risks posed by AI", 12 September 2023.

[80] A/HRC/56/68/Add.1, para. 54.

[81] Canada, Innovation, Science and Economic Development Canada, "The Artificial Intelligence and Data Act (AIDA) – companion document", 13 March 2023; and NetMission.Asia submission.

55. The Special Rapporteur received information about other States, such as Australia, China, India and Japan, that have reportedly taken steps to manage and regulate artificial intelligence, including through policy measures and, in some cases, binding legislation.[82]

2. Regional initiatives

56. With regard to regional initiatives, the Special Rapporteur welcomes the information that she received from the European Union, as well as its member States, about the passing of the Artificial Intelligence Act.[83] She recognizes that the Act is a binding regulatory standard that will have a significant impact in the European Union region through the harmonization of national legal standards with its provisions. The Special Rapporteur welcomes that the text of the Artificial Intelligence Act incorporates race, has human rights safeguards for high-risk artificial intelligence uses, prohibits some uses of artificial intelligence and provides for remedy mechanisms for persons affected by the use of high-risk artificial intelligence systems. She also welcomes that the European Union anti-racism action plan for 2020–2025 reportedly addresses racial discrimination arising from the use of new technologies, such as artificial intelligence, suggesting a degree of policy coherence across different European Union standards.[84] However, the Special Rapporteur received deeply concerning information indicating that there are exceptions to the protections set out in the Act in the contexts of immigration and border management and law enforcement.[85] Such exceptions reportedly exist despite the significant historical racial discrimination in both of these domains and the inherent pitfalls of allowing parallel legal frameworks to develop.[86] Such an approach risks the entrenchment of existing racial hierarchies and the significant perpetuation of human rights violations in the contexts of immigration and border management and law enforcement across European Union member States.

3. International initiatives

57. The Special Rapporteur is aware of measures taken by the United Nations to contribute to the management of artificial intelligence. She welcomes the establishment by the Secretary-General of the High-level Advisory Body on Artificial Intelligence and the publication of its recent interim report. However, she laments that specific reference is not made in that report to the risk of racial bias and discrimination. The Special Rapporteur welcomes the work of the Office of the United Nations High Commissioner for Human Rights (OHCHR) in integrating human rights into international dialogues about emerging technologies, such as artificial intelligence, including through the B-Tech Project. In addition to the work of the United Nations, the Special Rapporteur is aware of other international initiatives to promote dialogue and management, including initiatives of the Organisation for Economic Co-operation and Development and the Group of Seven.[87]

58. International organizations are well placed to facilitate international cooperation, technical assistance and research aimed at ensuring that artificial intelligence is governed in a manner that does not exacerbate the already gross inequalities between countries, which exist in many cases as a legacy of colonialism and slavery. Significant differences in technological infrastructure may lead to different challenges for countries as they implement artificial intelligence tools. Much of the focus on artificial intelligence technology and the discrimination challenges that it poses has been focused on countries in the global North, which could lead to gaps in the understanding of how artificial intelligence will have an impact on cultural, religious and other minorities in the global South.[88] There is a risk that the most developed countries in the global North will be able to influence the debate and

[82] NetMission.Asia submission.

[83] European Union and Spain submissions.

[84] European Union submission. See also https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html.

[85] Privacy International submission; and Access Now, “The EU AI Act: a failure for human rights, a victory for industry and law enforcement”, 13 March 2024.

[86] See A/HRC/48/76.

[87] NetMission.Asia submission.

[88] Danni Yu, Hannah Rosenfeld and Abhishek Gupta, “The ‘AI divide’ between the Global North and the Global South”, World Economic Forum, 16 January 2023.

dialogue on artificial intelligence in a way that perpetuates global power imbalances and limits the ability of countries in the global South to reap the potential benefits.

D. International human rights law framework

59. Artificial intelligence technology should be grounded in international human rights law standards. The most comprehensive prohibition of racial discrimination can be found in the International Convention on the Elimination of All Forms of Racial Discrimination. As reflected in article 1 (1), States drafted the Convention to incorporate a broad definition of racial discrimination as any distinction, exclusion, restriction or preference based on race, colour, descent, or national or ethnic origin that 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.

60. States parties to the International Convention on the Elimination of All Forms of Racial Discrimination have committed to pursuing the realization of a domestic and international community free of all forms of racism. To facilitate the substantive realization of racial equality, article 2 of the Convention requires States parties to ensure that they neither take part in any act of racial discrimination nor further programmes that lead to racial inequality. Furthermore, where racism, racial inequality or racial discrimination exists, States parties have an obligation to take effective and immediate action. This obligation to act is absolute. States parties' obligations to prevent racial inequality and racial discrimination encompass both preventive and remedial actions. While the Convention provides the most comprehensive prohibition of racial discrimination, other treaties also provide protection from such forms of discrimination.

61. Obligations to achieve racial equality and ensure non-discrimination extend to all areas of government policy and influence, including the design and application of artificial intelligence technologies. Whether racial discrimination resulting from artificial intelligence is intentional or not is irrelevant to States parties' duty to act, given the scope of the prohibition of racial discrimination under the International Convention on the Elimination of All Forms of Racial Discrimination and other human rights treaties. The duties of States parties to the Convention to pursue the realization of a domestic and international community free of all forms of racial discrimination are also relevant to the way in which States prevent and address inequalities within and between countries in relation to the distribution of the benefits of artificial intelligence technologies.

62. States must also ensure that all racial and ethnic groups enjoy the full scope of their human rights, as encompassed in article 5 of the International Convention on the Elimination of All Forms of Racial Discrimination. Article 5 provides for equality before the law, including, inter alia, the rights to equal treatment before the tribunals and all other organs administering justice; to security of person and protection by the State against violence or bodily harm, whether inflicted by government officials or by any individual group or institution; to freedom of peaceful assembly and association; to public health, medical care, social security and social services; and to education and training. These rights, as well as provisions guaranteeing their non-discriminatory application, are also provided for in other human rights treaties, including the International Covenant on Civil and Political Rights and the International Covenant on Economic, Social and Cultural Rights.

63. There are other provisions of international human rights law that bestow upon States the responsibility to address the discriminatory impacts of artificial intelligence technologies, as described above. The collection and use of data without human rights safeguards raises significant privacy concerns, which can be amplified for those from marginalized racial and ethnic groups. Accordingly, the Special Rapporteur would like to remind States of the provisions of article 17 of the International Covenant on Civil and Political Rights that provide for freedom from arbitrary or unlawful interference in a person's privacy and bestow an obligation on States to ensure relevant legal protections. Other provisions of the Covenant also apply to manifestations of racial discrimination relating to artificial intelligence technologies. The use of artificial intelligence, including in contexts such as law enforcement,

can impact liberty and security of person and have life-and-death consequences for those from marginalized racial and ethnic groups. Article 6 of the Covenant outlines the inherent right to life and obligates States to provide legal protections in this regard. Article 7 provides that no one is to be subjected to torture or cruel, inhuman or degrading treatment or punishment. Article 9 provides that everyone has the right to liberty and security of person and that no one is to be subjected to arbitrary arrest or detention. Article 14 makes clear that all persons should be equal before the courts and tribunals. Article 26 provides for protection from discrimination for minority groups. Article 2 (1) of the Covenant establishes an obligation to ensure the non-discriminatory application of all the provisions of the Covenant. There are also provisions of the international human rights law framework relating to the use of artificial intelligence in immigration and border control and in the context of social media. These are explored in previous reports under the mandate.[89]

64. International human rights law provides that all people who may be subjected to racial discrimination have a right of access to remedies, which applies in cases in which discrimination occurs as a result of artificial intelligence. Article 6 of the International Convention on the Elimination of All Forms of Racial Discrimination provides for the right of access to effective protection and remedies, through competent national tribunals and other State institutions. In addition, the General Assembly has recognized five main elements of the right to a remedy and reparation for victims of gross human rights violations: restitution, compensation, rehabilitation, satisfaction and guarantees of non-repetition.[90]

65. Business entities play a significant role in the design and application of artificial intelligence. They are the main actors in its development and are often contracted by Governments to deploy it in public sector settings. The Guiding Principles on Business and Human Rights outline the relevant obligations of Governments and the relevant human rights responsibilities of both Governments and businesses. The Guiding Principles establish that States must protect against human rights abuses committed by third parties within their territory and/or jurisdiction, including business enterprises. States should provide such protection by ensuring effective policies, legislation, regulations and adjudication, among other actions. The Guiding Principles establish the responsibility of companies to prevent, mitigate and remedy human rights violations that they may cause or to which they may contribute and to conduct human rights due diligence with regard to relevant business activities.[91] In addition, the Guiding Principles establish government obligations and business responsibilities to ensure access to remedies for business-related human rights violations, complementing the right to remedy provided for in other standards, as outlined above. The OHCHR B-Tech project has involved work on guidance and resources for implementing the Guiding Principles in the technology space, including specific work on artificial intelligence.[92]

IV. Conclusions and recommendations

66. **The previous mandate holder issued a clear call to States and other stakeholders, including business entities, to reject a "colour-blind" approach to the governance and regulation of emerging technologies, including artificial intelligence. She urged States to regulate these technologies within an approach that recognizes structural racism and is based on key human rights standards. Nevertheless, the management and regulation of artificial intelligence largely remain insufficient, inadequately attentive to racial bias and not reflective of international human rights law standards. This reality persists**

[89] A/75/590 and A/78/538.

[90] 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, paras. 15–23.

[91] See also United Nations Office on Genocide Prevention and the Responsibility to Protect and Economic and Social Research Council Human Rights, Big Data and Technology Project, University of Essex, "Countering and addressing online hate speech: a guide for policy makers and practitioners", policy paper, July 2023; and A/74/486, paras. 44 and 45.

[92] See OHCHR, "B-Tech Project: OHCHR and business and human rights", available at https://www.ohchr.org/en/business-and-human-rights/b-tech-project.

despite the clarity and timeliness of prior calls for a "colour-blind" approach and the increase in awareness of systemic racism in the intervening four years. The assumption that technology is objective and neutral remains pervasive and drives a race to integrate artificial intelligence into society despite its racially discriminatory impacts and without due consideration of whether it is necessary. While artificial intelligence does have positive potential, including for equality and inclusion, it is not a panacea for all societal issues and must be effectively managed to balance its benefits and risks.

67. **The effective and comprehensive regulation of artificial intelligence is central to achieving this careful balance. While the effective regulation of artificial intelligence is vital, there are additional steps that States and others can take to effectively address the racially discriminatory impacts of these technologies. Developing human rights-based public education about emerging technologies and building artificial intelligence literacy are also very important. When individuals and groups understand artificial intelligence and are aware of their human rights in the digital space, they are empowered to use that knowledge responsibly and become a discerning and responsible audience that can improve accountability for artificial intelligence systems.**

68. **States should:**

(a) **Address the challenge of regulating artificial intelligence with a greater sense of urgency, bearing in mind the speed with which these technologies are being developed and the multitude of ways in which they are already perpetuating racial discrimination across societal domains;**

(b) **Develop artificial intelligence regulatory frameworks that are based on a comprehensive understanding of systemic racism and are grounded in international human rights law, including the prohibition of racial discrimination. Such frameworks should not be based on siloed approaches and should take into account different legal instruments, including dedicated artificial intelligence legislation, privacy laws, freedom of information provisions, anti-discrimination legislation and sectoral regulations, to achieve comprehensive and effective regulation that prevents and addresses the racially discriminatory impact of artificial intelligence;**

(c) **Consider the role that voluntary standards can play in artificial intelligence regulatory frameworks. Voluntary standards can provide actionable guidelines on the practical implementation of legal standards. However, artificial intelligence regulation should not rely solely on voluntary standards, due to the significance of the human rights implications of these technologies, including in relation to racial discrimination;**

(d) **Enshrine a legally binding obligation, within regulatory frameworks, to conduct comprehensive human rights due diligence assessments, including explicit criteria to assess racial and ethnic bias, in the development and deployment of all artificial intelligence technologies. Human rights due diligence should include data-testing protocols and thresholds that safeguard against algorithmic bias, including racial and ethnic bias. They should be completed before the deployment of new technologies, particularly in public settings, such as educational, law enforcement and health-care settings;**

(e) **Consider prohibiting the use of artificial intelligence systems that have been shown to have unacceptable human rights risks, including those that violate the prohibition of racial discrimination;**

(f) **Ensure that there are provisions within regulatory frameworks to guarantee full transparency about automated decision-making processes, including rights of access to information, in cases in which artificial intelligence use is deemed permissible, based on comprehensive human rights due diligence;**

(g) **Put in place clear and accessible appeals processes, which have a mandate to assess and address the racially discriminatory impacts of artificial intelligence and involve human review. Equitable access to such appeals processes should be ensured;**

(h) **Establish mechanisms to enable affected individuals and groups to gain access to remedies that ensure restitution, compensation, rehabilitation, satisfaction and guarantees of non-repetition in cases in which artificial intelligence technologies have led to human rights violations, including racial discrimination;**

(i) **Avoid any exceptions within regulatory standards that could lead to violations of the prohibition of racial discrimination under international human rights law;**

(j) **Ensure that stakeholders from all marginalized racial and ethnic groups, as well as professionals in relevant sectors, are consulted in a meaningful and effective way in the development and implementation of artificial intelligence regulations, as well as the development and use of artificial intelligence technologies;**

(k) **Invest in disaggregated data collection across all relevant sectors to obtain the information necessary to base artificial intelligence regulation on an understanding of systemic racism, to address data problems in artificial intelligence systems and to better monitor and evaluate the impact of artificial intelligence technology on those from marginalized racial and ethnic groups;**

(l) **Adopt an approach to data that is grounded in human rights standards, by ensuring disaggregation, self-identification, transparency, privacy, participation and accountability in the collection and storage of data;**[93]

(m) **Set up robust mechanisms for the oversight and continuous monitoring of artificial intelligence tools, including regular audits of their impact, to ensure compliance with regulations and to address any concerns raised by affected individuals or communities, as well as potential biases generated by artificial intelligence models over time;**

(n) **Engage in international cooperation, capacity-building and research to ensure that the benefits of artificial intelligence are distributed more equitably among countries, to avoid the deepening of inequalities that exist as a legacy of colonialism and slavery;**

(o) **Develop human rights-centred public education about the acceptable and responsible use of artificial intelligence technology to increase artificial intelligence literacy, including components that specifically raise awareness about the racially discriminatory impacts of artificial intelligence.**

69. **Business entities should:**

(a) **Undertake human rights due diligence assessments at all stages of artificial intelligence product design, development and deployment;**

(b) **Ensure meaningful and effective consultation with those from marginalized racial and ethnic groups, professionals from relevant societal domains and those with expertise in systemic racism in the design, development and deployment of artificial intelligence products;**

(c) **Develop protocols for ensuring full transparency and the sharing of information about algorithmic decision-making for products that have human rights implications;**

(d) **Ensure the continuous monitoring of artificial intelligence products for racial bias;**

(e) **Develop training on racial discrimination, including implicit bias and systemic racism, for all those involved in the design, development and deployment of artificial intelligence. The development of such training should draw upon both international human rights law standards and research undertaken on the racially discriminatory impact of artificial intelligence technologies;**

[93] See A/HRC/42/59 and A/HRC/44/57.

(f) **Contribute to efforts to develop human rights-centred public education to enhance the accessibility of artificial intelligence literacy.**

70. **The United Nations and its independent human rights mechanisms should:**

(a) **Facilitate effective dialogue and debate about artificial intelligence technologies and their regulation among stakeholders;**

(b) **Integrate an explicit focus on the racially discriminatory impact of artificial intelligence technologies into the work of the High-level Advisory Body on Artificial Intelligence;**

(c) **Ensure that publications and guidance on artificial intelligence technologies are grounded in international human rights law, including the prohibition of racial discrimination, and explicitly recognize racial bias in the design and deployment of artificial intelligence technology as a serious global issue;**

(d) **Play a role in monitoring the human rights implications of artificial intelligence technologies, including those relating to racial discrimination;**

(e) **Support international cooperation, capacity-building and research to try to achieve a more equitable distribution of the benefits of artificial intelligence among countries.**