AMNESTY INTERNATIONAL
INTERNATIONAL SECRETARIAT

www.amnesty.org

AMNESTY INTERNATIONAL'S OBSERVATIONS TO THE UNITED NATIONS COMMITTEE ON THE ELIMINATION OF RACIAL DISCRIMINATION'S DRAFT GENERAL RECOMMENDATION NO. 36 ON PREVENTING AND COMBATING RACIAL PROFILING

Amnesty International welcomes the call for contributions on the Draft General Recommendation No. 36 (20XX) on preventing and combating racial profiling. The comments below should not be seen as an exhaustive list of issues but rather include certain suggestions on how to strengthen the text and provisions of the draft General Recommendation. The structure of this document follows the outline of the draft General Recommendation.

For the purpose of this document, Amnesty International uses the term ethnic profiling in order to clarify that discriminatory profiling practices also go beyond the aspect of race. On this basis Amnesty International suggests the title of the Recommendation be changed to General Recommendation No. 36 on Preventing and Combating Racial and Ethnic Profiling.[1]

IV. DEFINING AND UNDERSTANDING ETHNIC PROFILING

Amnesty recommends that the definition of ethnic profiling in paragraph 16 is reformulated to clarify that it constitutes ethnic profiling if – in the absence of a suspect description - personal attributes such as presumed race, colour, descent, nationality or ethnic origin etc. are taken into account in law-enforcement decision-making, not only as a decisive factor but also in combination with other factors.[2] Further, the element of "individual behaviour" as currently included in the definition does not exclude the application of ethnic profiling practices, as even normal behaviour by individuals of certain ethnic appearance may be interpreted as suspicious (e.g. sitting on a park bench for extended amounts of time or walking quickly). It should thus be clarified that decision-making should be based on reasonable and objective criteria pointing to involvement in crime.

It is important to clarify that ethnic profiling is not limited to situations of open discrimination, nor does it require any intention to discriminate. Often, ethnic profiling is exclusively (mis-)understood as being the expression of an individual law enforcement officer's discriminatory attitude – hence why many law enforcement agencies deny having a problem of ethnic profiling. Therefore, Amnesty suggests that the General recommendation outlines the various different ways in which ethnic profiling can manifest itself, to clarify that a wide range of different measures are required to address all shapes and forms of ethnic profiling. On the level of the individual officer, ethnic profiling may be the result of clearly discriminatory attitudes or of unconscious bias. At an institutional level, there may be policies or approaches that are explicitly discriminating or encouraging ethnic profiling, while ethnic profiling practices may also be the result of seemingly neutral policies which set criteria that in practice disproportionately affect certain groups.[3] In some instances, discriminatory attitudes may be shared across the police agency, thus shaping police approaches and priorities. It is crucial to have these different manifestations of ethnic profiling in mind when adopting measures to address discrimination, from establishing policy frameworks to training and accountability.

V. CONSEQUENCES OF ETHNIC PROFILING

Regarding paragraph 18, we recommend including the risk that ethnic profiling practices are erroneously perceived as being an effective approach to counter crime. At the level of the individual officer, "confirmation bias"[4] may lead officers

[1] We recognise and support the Committee's practice of taking an intersectional approach in recognising that profiling can take place on the basis of any of the prohibited grounds of discrimination which may interact with each other.
[2] World Conference against Racism, Racial Discrimination, Xenophobia and Related Intolerance (Durban Declaration), 2001, para. 72, defines ethnic profiling as " [...] the practice of police and other law enforcement officers relying, to any degree, on race, colour, descent or national or ethnic origin as the basis for subjecting persons to investigatory activities or for determining whether an individual is engaged in criminal activity" [emphasis added].
[3] See for instance ECRI, General Policy Recommendation No. 11 on combating racism and racial discrimination in policing, adopted on 29 June 2007, para. 38.
[4] Confirmation bias can be understood as seeking or interpreting of evidence in ways that accord with the existing beliefs, or expectations. This can include both looking for information that affirms current beliefs, while not looking (even avoiding) information that disconfirms such beliefs. It can even persist after the information that shaped such beliefs has been discredited or withdrawn: Minhas and Walsh, Influence of racial stereotypes on investigative

to be falsely reassured that their decision-making is effective. For example, even if the large majority of bias-motivated stop-and-searches do not reveal any criminal involvement of the person stopped, each single case where something incriminating is found is likely to subconsciously confirm their bias towards the concerned group and that their way of selecting people for stop and search is effective, including when the "hit-rate" (i.e. the percentage of stops actually revealing criminal involvement) is lower for this group than for other segments of society.

On the other hand, there is a risk of "self-fulfilling prophecy".[5] If members of a certain group are stopped at a disproportionate rate, they are also likely to appear in crime statistics at a disproportionate rate, hence "confirming" the initial assumption of their heightened involvement in crime as compared to other groups. This in turn will lead to ineffective policing since crimes committed by the rest of the population remain undetected and ethnic profiling continues to be applied at the expense of more effective strategies. As an example, it was pointed out by the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism that "[...] profiles based on ethnicity, national origin and religion are [...] under-inclusive in that they will lead law-enforcement agents to miss a range of potential terrorists who do not fit the respective profile. [...P]rofiles based on ethnicity, national origin or religion are easy to evade. Terrorist groups have regularly proved their ability to adapt their strategies, with the use of female and child suicide bombers, to avoid the stereotype of the male terrorist as just one example. Thus, as law-enforcement specialists acknowledge, any kind of terrorist profile based on physical characteristics can easily become self-defeating."[6]

VI. DISCRIMINATORY BIASES ASSOCIATED WITH ARTIFICIAL INTELLIGENCE

Due to the lack of a clear and universally agreed definition of artificial intelligence, Amnesty considers that the term "artificial intelligence" in paragraph 21 would benefit from being defined in a way that ensures that the Recommendations apply to all systems that are based on automatic data processing and algorithmic decision making and that the term is not interpreted narrowly to only cover systems that mimic or supersede human intelligence or possess self-learning capabilities. We refer to the language of the definition of artificial intelligence in the Council of Europe Commissioner for Human Rights' 10-point Recommendation on Artificial Intelligence and Human Rights[7] as an example of good practice:

> "An artificial intelligence system is a machine-based system that makes recommendations, predictions or decisions for a given set of objectives. It does so by: (i) utilising machine and/or human-based inputs to perceive real and/or virtual environments; (ii) abstracting such perceptions into models manually or automatically; and (iii) deriving outcomes from these models, whether by human or automated means, in the form of recommendations, predictions or decisions."

As to paragraph 23, we note that analysis has shown the significant flaws in facial recognition technologies; of the alerts the system generates over two thirds are false positives.[8] Remotely-piloted aerial vehicles, or drones, are also being equipped with facial recognition technologies.[9] Irrespective of their accuracy, facial recognition technologies interfere with the rights to privacy, freedom of expression, freedom of association and peaceful assembly, and also have the potential to be used for discrimination or ethnic profiling. For example, in a study by the American Civil Liberties Union in July 2018, the facial recognition tool, called "Rekognition", incorrectly matched 28 members of Congress, identifying them as other

decision-making in criminal investigations: A qualitative comparative analysis, Cogent Social Sciences (2018), 4: 1538588, p. 2-3; available at: https://www.tandfonline.com/doi/full/10.1080/23311886.2018.1538588.

[5] See the circle of self-fulfilling prophecy described in: European Fundament Rights Agency, Preventing unlawful profiling today and in the future: a guide, 2018, p. 49.

[6] Report of the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism, Martin Scheinin, UN Doc. A/HRC/4/26 (2007), para. 52.

[7] Council of Europe, Commissioner for Human Rights, 'Unboxing Artificial Intelligence: 10 steps to protect Human Rights,' available at: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64

[8] For example, in the UK the London's Metropolitan Police Service (MPS) used automated facial recognition (AFR) at the annual Notting Hill Carnival in 2016 and 2017. The use of this technology by the MPS returned false positives in more than 98% of alerts generated according to a Freedom of Information request (see https://www.met.police.uk/SysSiteAssets/foi-media/metropolitanpolice/disclosure_2018/april_2018/information-rights-unit---mps-policies-on-automated-facial-recognition-afr-technology). South Wales Police has been using AFR since May 2017, including at sporting events and concerts. The AFR system used by South Wales Police has returned confirmed false positives at a rate of at least 50 % from is new algorithm. Its previous algorithm returned false positives at a rate of at least 72 % (See Universities' Police Science Institute Crime and Security Research Institute, Cardiff University, An Evaluation of South Wales Police's Use of Automated Facial Recognition, September 2018, http://afr.southwales.police.uk/cms-assets/resources/uploads/AFR-EVALUATION-REPORT-FINAL-SEPTEMBER-2018.pdf)

[9] http://www.gizmodo.co.uk/2018/06/dji-and-taser-maker-axon-are-teaming-up-to-make-police-drones/; https://www.haaretz.com/israel-news/israel-bolsonaro-discuss-drone-sale-senior-source-says-1.6801993

people who have been arrested for a crime.[10] The false matches were disproportionately of people of colour, including six members of the Congressional Black Caucus. The serious human rights risks combined with the lack of a legislative framework to regulate the use of facial recognition technologies, the lack of transparency regarding when and where these technologies can be and are being used (and to what end), the absence of public information and rights of review or appeal, all indicate that the use of facial recognition technologies, and the retention of related data, could be unlawful. We therefore suggest that the General Recommendation explicitly expresses the view that facial recognition technologies should not be deployed unless governments can demonstrate that they can be used in line with international law, including international human rights law and the principles of legality, proportionality and necessity.

We suggest the addition of a new paragraph between paragraphs 23 and 24 to address predictive policing. This paragraph would make clear that as predictive policing systems advance rapidly and are deployed across the law enforcement and security sphere, there is evidence that the use of artificial intelligence systems can perpetuate and exacerbate discrimination and identity bias. One research study from the Human Rights Data and Analysis Group (HRDAG)[11] developed a replica of a predictive policing algorithmic programme used by police forces in numerous US states and ran it as a simulation on crime data in Oakland. It concluded that the programme reinforced existing racial discrimination within the police, as the system was built using already biased data that recorded higher crime rates in parts of the city with a higher concentration of black residents. The algorithm therefore predicted more crime in those areas, dispatching more police officers, who unsurprisingly made more arrests. The new data was fed back into the algorithm, reinforcing its decision-making process and creating a pernicious feedback loop that would contribute to over-policing of black neighbourhoods in Oakland.

In 2014, the Metropolitan Police Service (MPS) announced it would introduce an automated system to assign risk scores to individual suspected of being 'gang members' in London. The pilot reportedly used data gathered from social media along with police crime reports to generate offending risk scores for all individuals associated with London gangs. Amnesty International's research into the MPS's Gangs Databases demonstrated that the current manual system used by the police to flag individuals as 'gang associated' is arbitrary, lacks adequate oversight and contributes to the overrepresentation of BAME young people in the criminal justice system. In this context, the introduction of automated risk-scoring raises significant human rights concerns which are compounded further by an already deeply flawed data collection policy with no effective oversight and safeguards in place.[12]

VII. RECOMMENDATIONS

A. Legislative measures

While legislation provides a general framework for policing powers, policies and procedures govern police officers' conduct in day-to-day policing. The risk of ethnic profiling is especially high in the absence of proper instructions, when the officer has full discretion to decide whom to stop. Amnesty thus recommends to more clearly express in this section or even in a new separate section the need for institutional policies for stop-and-search as well as all other areas where there is a potential risk of ethnic profiling. In addition to establishing a prohibition of ethnic profiling, law enforcement agencies should be urged to establish a solid policy framework and guidance for decision-making. This includes a clear definition of what constitutes reasonable and objective suspicion and an outline of legitimate criteria that may be considered in the decision of whether or not to stop a person.[13] It should also be ensured that policies which are neutral at first glance do not implicitly lead to or encourage ethnic profiling, or disproportionately affect certain groups.

B. Human rights education and training

Training should not be limited to theoretical human rights messages on discrimination. It should enable law enforcement officials to critically reflect on their own subconscious biases and how to overcome them in relevant situations of their

[10] ACLU, Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots, 26 July 2018, https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28: 11 of the 28 false matches misidentified people of color (roughly 39 percent), including civil-rights leader Rep. John Lewis (D-GA) and five other members of the Congressional Black Caucus. Only twenty percent of current members of Congress are people of color, which indicates that false-match rates affected members of color at a significantly higher rate.
[11] Lum, Kristian and Isaac, William Isaac, To predict and serve?, 7 October 2016, https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x
[12] Amnesty International, 'Trapped in the Matrix: Secrecy, stigma, and bias in the Met's Gangs Database,' May 2018, available at: https://www.amnesty.org.uk/files/reports/Trapped%20in%20the%20Matrix%20Amnesty%20report.pdf
[13] In this context, see also European Court of Human Rights, Gillian and Quinton v. The United Kingdom (Application no. 4158/05), 2010, concluding that stop and search powers which are not subject to any restrictions but allow police to stop individuals based solely on intuition are not "in accordance with the law" and hence a violation of Article 8, also due to the absence of adequate legal safeguards against abuse, as it will be difficult if not impossible to proof that powers were exercised improperly (paras. 83 – 87).

work. It should be practical in the sense that it conveys to police officers what is expected of them and provide them with the necessary skills to establish reasonable and objective suspicion in concrete situations.[14] Further, it should stress the consequences of ethnic profiling, both with regard to its ineffectiveness and counter-productivity, highlighting that it is not compatible with good policing. The potential consequences of engaging in discriminatory conduct, which include disciplinary action, must also be emphasised.[15]

C. Recruitment measures

In addition to promoting a diverse workforce, we recommend that paragraph 29 highlights that during the recruitment process, candidates should be assessed on their views, attitudes and abilities to comply with the principle of non-discrimination. Candidates displaying clearly biased or discriminatory opinions or attitudes should not be considered.

E. Disaggregated data

We consider that paragraph 31 would benefit from introducing stop and search forms. Such forms should state among other things the reason for the stop and the perceived ethnicity and gender of the person stopped, and may in themselves help to reduce ethnic profiling, as they require officers to justify their stop on legitimate grounds, both on paper and to the individual in front of them. A copy or receipt of the form should be handed to the person stopped, which should also specify how a complaint about the stop can be made. Such stop forms can further aid the collection of data and can give insights into any bias or disproportionality in regard to police stops of people in particular groups.[16] The copy of the form retained by the state agent implementing the stop should not contain any individual identifying data apart from ethnicity/gender.

F. Accountability

The General Recommendation should make clear that ethnic profiling is a widespread phenomenon and no law enforcement agency should consider itself free from it. It is thus crucial that accountability is ensured at all levels and at all stages where ethnic profiling may occur.

Amnesty International recommends strengthening paragraph 32 to clearly express that this includes accountability of the individual police officer for discriminatory conduct, which – in case of suspicion or indications to such behaviour or upon a complaint - should be thoroughly investigated and where appropriate met with corrective measures or disciplinary sanctions. It should further stress the responsibilities of the chain of command to enforce non-discriminatory policies and actively supervise the conduct of officers and to take corrective action whenever needed. Beyond the individual level, managers should ensure that their policies and practices are in line with the principle of non-discrimination and do not encourage or facilitate, explicitly or implicitly, ethnic profiling practices. Paragraph 32 should also highlight the duty of the authorities to take all necessary steps to investigate institutional racism.[17]

With regard to paragraph 34, recommendations should be made that States should ensure that there is an external oversight mechanism which is mandated to investigate suspected discriminatory police misconduct, including ethnic profiling. This oversight body should both accept individual complaints as well as have the powers to self-initiate investigations, review police policies and investigate suspected cases of institutional racism.[18]

Even where there is no such mechanism yet, a venue should be available where victims of ethnic profiling can complain.

G. Artificial intelligence

We consider that the General Recommendation would benefit from inclusion of language about the need for transparency in the development and use of artificial intelligence systems. We suggest the addition of a new paragraph at the start of this section, which would emphasise that States must ensure and require accountability and maximum possible transparency around public sector use of machine learning systems. This must include explainability and intelligibility in

[14] ECRI, General Policy Recommendation No. 11 on combating racism and racial discrimination in policing, adopted on 29 June 2007, para. 46.
[15] UN High Commissioner for Human Rights, Preventing and countering racial profiling of people of African descent: Good Practices and Challenges, 2019, para. 47.
[16] Ibid., paras. 35 and 41; However, considering that the collection of ethnic data can also be used to facilitate ethnic profiling, safeguards against possible misuse of the data must be established. See also Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance, Mutuma Ruteere, A/HRC/29/46, 2015, para 68.
[17] See for instance European Court of Human Rights, Lingurar v. Romania (Application No. 48474/14), 2019, paras. 75 - 77.
[18] See ECRI, General Policy Recommendation No. 11 on combating racism and racial discrimination in policing, adopted on 29 June 2007, paras. 58 and 59; Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance, Mutuma Ruteere, A/HRC/29/46, 2015, para. 70.

the use of these technologies so that the impact on affected individuals and groups can be effectively scrutinised by independent entities, responsibilities established, and actors held to account. States should publicly disclose where machine learning systems are used in the public sphere, provide information that explains in clear and accessible terms how automated and machine learning decision-making processes are reached, and document actions taken to identify, document and mitigate against discriminatory or other rights-harming impacts. They should enable independent analysis and oversight by using systems that are auditable. We suggest that it is recommended that States must avoid using 'black box systems' that cannot be subjected to meaningful standards of accountability and transparency, and refrain from using these systems at all in high-risk contexts.[19]

As regards paragraph 35, in the first sentence we suggest framing this paragraph so that it includes all government agencies rather than just law enforcement agencies. The risks of discrimination and ethnic profiling posed by the use of artificial intelligence are not limited to law enforcement. After the sentence "States should also ensure the human rights compliance and the ethical governance of artificial intelligence," we suggest that the paragraph notes that the legal framework governing the use of artificial intelligence should also clearly define the purpose for which artificial intelligence systems are deployed. States should specify areas of application in their legislation and policies which reflect legitimate objectives as well as include safeguards to ensure that systems cannot be used, neither upon their introduction nor at a later stage, for unintended purposes.

As regards paragraph 36, we suggest amendments be made so that the community impact assessment also involves a human rights impact assessment. Such assessments should focus on not only the algorithm but also the context in which the data was collected, as well as the potentially adverse impact on human rights of the actions taken in response to artificial intelligence outputs. Assessments should be conducted regularly and continuously to ensure their continued human rights compliance.

We suggest adding two new paragraphs between paragraph 36 and 37 to address data protection and privacy in the development of artificial intelligence systems, and that public officials are properly trained and provided with clear guidance on how to act on the output of artificial intelligence systems, and that they are aware of and sensitive to the risks of discrimination and other human rights harms.

At the end of paragraph 37, we suggest an additional sentence is added to address the need to ensure that the law enforcement agency remains responsible and accountable for their actions, even if acting on artificial intelligence. Artificial intelligence may be used to support, but not to replace human decision making.

As to paragraph 38, we suggest an addition is made after the first sentence, requiring States to effectively implement the UN Guiding Principles on Business and Human Rights. We also suggest that the paragraph address the human rights responsibilities of private sector actors in further detail. Specifically, private sector actors that develop and implement artificial intelligence systems should disclose the process of identifying risks, the risks that have been identified, and the concrete steps taken to prevent and mitigate identified human rights risks. This may include: disclosing information about the risks and specific instances of discrimination the company has identified, for example risks associated with the way a particular system is designed, or with the use of artificial intelligence systems in particular contexts; in instances where there is a risk of discrimination, publishing technical specification with details of the artificial intelligence system and its functions, including samples of the training data used and details of the source of data; and establishing mechanisms to ensure that where discrimination has occurred through the use of an artificial intelligence system, relevant parties, including affected individuals, are informed of the harms and how they can challenge a decision or outcome.

At the end of paragraph 40 we suggest adding a requirement that States should further ensure that individuals affected by artificial intelligence have an effective opportunity to challenge decisions taken as a result of the output of those systems.

[19] The AI Now Institute at New York University, AI Now 2017 Report, https://ainowinstitute.org/AI_Now_2017_Report.pdf