Privacy International's submission for the UNSR on racism's thematic report on artificial intelligence (AI) and racial discrimination

March 2024

1. Introduction

Privacy International (PI)[1] welcomes the opportunity to provide input to the forthcoming report the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related tolerance to the 56th session of Human Rights Council which will examine and analyse the relationship between artificial intelligence (AI) and non-discrimination and racial equality, as well as other international human rights standards.[2]

AI applications are becoming a part of everyday life: social media newsfeeds, mediating traffic flow in cities, connected consumer devices, automated cars, eligibility mechanisms for welfare services, access to medical diagnostics, location tracking, spam filters, voice recognition systems, and search engines. If implemented responsibly, AI applications have the potential to promote the enjoyment of human rights. However, there is a growing evidence that commercial and state use has a detrimental impact on human rights, and exacerbate existing inequalities and discriminatory practices.

Our submission focuses on the use of AI applications in specific sectors and how they can negatively affect the enjoyment of rights such as the right to privacy and freedom from discrimination in particular racial discrimination - in the areas of health, migration, welfare and social protection, and employment, in addition to specific discriminatory AI enabled technologies such as facial recognition technology by law enforcement and the private sector.

In particular PI suggests the following main aspects should be covered in the UNSR's report:

- Establish the need for a human rights-based approach to all AI applications and describe the necessary measures to achieve it including human rights by design and human rights impact assessments, as well as ensure the meaningful participation of affected communities in decision-making processes.

[1] PI is an international non-governmental organisation, which campaigns against companies and governments who exploit individuals' data and technologies. PI employs specialists in their fields, including technologists and lawyers, to understand the impact of existing and emerging technology upon data exploitation and our right to privacy. For more information: https://privacyinternational.org/

[2] Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related tolerance, Call for input: thematic report on artificial intelligence (AI) and racial discrimination, available at: https://www.ohchr.org/en/calls-for-input/2024/call-input-thematic-report-artificial-intelligence-ai-and-racial

- Identify the racial discrimination risks of specific AI applications, due to the technologies employed and/or the context of their use; and describe the circumstances when AI applications should be banned because of human rights concerns, including concerns of discrimination and in particular racial discrimination.
- Recommend states to adopt or review existing discrimination law, effective data protection legislation, and sectoral laws to address the negative human rights implications of AI applications – at individual, group and society level, including by effectively regulating the use of AI technologies by private companies.
- Explore the scope of state's obligation to ensure that public sector uses of AI technologies – particularly in health care, welfare, migration, policing, and surveillance, is used responsibly, and does not result or exacerbate racial discrimination .
- Define the scope of responsibility of non-state actors, including companies, for AI uses and the need for mechanisms to ensure that they are held accountable.

2. Key concerns regarding AI and the right to privacy

PI has long documented how AI applications are often used to process data and to identify individuals, predict and influence their behaviours. In particular AI technologies are being used:

- to infer and generate sensitive information about people;
- to profile people based upon population-scale data;
- to identify people who wish to remain anonymous; and
- to make decisions on the basis of the analysis of this data.

AI-driven consumer products and autonomous systems are frequently equipped with sensors that generate and collect vast amounts of data without the knowledge or consent of the users or those in their proximity.[3] On the internet, vast amounts of data about people's lives and behaviour is increasingly gathered through tracking technologies, including sensitive data, for example on mental health websites[4] or menstruation apps.[5] AI applications facilitate the further analysis of this data and the generation of inferences to create finely grained profiles.[6] Such profiles are then used to target people with advertising – both commercial and political – and ultimately feed into other consequential decisions which may negatively affect human rights, including access to credit and insurance. AI applications are also increasingly being used in digital identity systems for a range of purposes, including authentication and verification.[7]

[3] For further information on each of these, see PI and ARTICLE 19 publication on "Privacy and Freedom of Expression In the Age of Artificial Intelligence", April 2018, available at https://privacyinternational.org/sites/default/files/2018-04/Privacy%20and%20Freedom%20of%20Expression%20%20In%20the%20Age%20of%20Artificial%20Intelligence.pdf , pp. 6-7.

[4] PI, "Your mental health for sale", https://privacyinternational.org/campaigns/your-mental-health-sale.

[5] PI, "No Body's Business But Mine: How Menstruation Apps Are Sharing Your Data", https://privacyinternational.org/long-read/3196/no-bodys-business-mine-how-menstruations-apps-are-sharing-your-data.

[6] There is an entire ecosystem dedicated to these privacy invasive practices, including data brokers and ad tech companies. PI, "Challenge to Hidden Data Ecosystem", https://privacyinternational.org/legal-action/challenge-hidden-data-ecosystem

[7] For example, Yoti, developed Yoti Age Scan technology, that uses AI to estimate an individual's age based on their image. This is used, for example, within the Yoti app instead of providing a verified ID document that contains their age in order to be able to buy alcohol or to access adult content online. For further information see: PI, "The Identity Gatekeepers and the Future of Digital Identity", https://privacyinternational.org/long-read/3254/identity-gatekeepers-and-future-digital-identity and PI, "Demanding identity systems on our terms", https://privacyinternational.org/campaigns/demanding-identity-systems-our-terms

As such, AI applications can affect the whole range of human rights,[8] and have a negative effect on some of the most vulnerable groups in society, documenting how the use of AI has exacerbated, rather than addressed, existing discrimination and exclusion.[9]

Because of the central role data play in most AI applications, the right privacy is particularly affected. Some of the key concerns regarding AI applications and privacy are:

- **Data exploitation:** AI applications frequently rely on the generation, collection, processing, and sharing of large amounts of data, both about individual and collective behaviour. This data can be used to profile individuals and predict future behaviour. It is often difficult to fully understand what kinds and how much data devices, networks, and platforms generate, process, or share, indeed this is often opaque by design.
- **Opacity and secrecy of profiling and automated decision making:** Some AI applications can be opaque to individuals, regulators, or even the designers of the system themselves, making it difficult to challenge or interrogate outcomes. While there are technical solutions to improving the interpretability and/or the ability to audit of some systems for different stakeholders, a key challenge remains where this is not possible, and the outcome has significant impacts on people's lives.
- **Re-identification and de-anonymisation:** AI applications can be used to identify and thereby track individuals across different devices, in their homes, at work, and in public spaces. For example, while personal data is routinely (pseudo-) anonymised within datasets, AI can be employed to de-anonymise this data.[10]
- **Discrimination, unfairness, inaccuracies and bias:** AI-driven identification, profiling, and automated decision-making may also lead to unfair, discriminatory, or biased outcomes. People can be misclassified, misidentified, or judged negatively, and such errors or biases may disproportionately affect certain groups of people.

3. AI applications and contexts of particular concerns

The term 'Artificial Intelligence' or 'AI' is used to refer to a diverse range of applications and technologies, with different levels of complexity, autonomy and abstraction. This broad usage encompasses machine learning (which makes inferences, predictions and decisions about individuals), domain-specific AI algorithms, fully autonomous and connected objects and even the futuristic idea of an AI 'singularity'. This lack of definitional clarity is a challenge: different types of AI applications and the context into which they are deployed raise specific regulatory issues.[11]

---

[8] As noted by the UN General Assembly resolution on the right to privacy in the digital age, "artificial intelligence or machine-learning technologies […] may lead to decisions that have the potential to affect the enjoyment of human rights, including economic, social and cultural rights, and affect non-discrimination".(The right to privacy in the digital age, GA Res 75/176, 16 December 2020, https://undocs.org/A/RES/75/176.) See also PI, "Artificial Intelligence", Explainer, https://privacyinternational.org/learn/artificial-intelligence.

[9] Examples of abuse of AI applications can be found here https://privacyinternational.org/examples?field_type_of_abuse_target_id_2%5B%5D=264.

[10] Multiple studies have shown that potential de-anonymisation capabilities of AI technologies. Similarly, in a more recent study published in Nature, researchers were able to demonstrate that, despite the anonymisation techniques applied, "data can often be reverse engineered using machine learning to re-identify individuals." Luc Rocher, Julien M. Hendrickx & Yves-Alexandre de Montjoye, "Estimating the success of re-identifications in incomplete datasets using generative models", 23 July 2019, https://www.nature.com/articles/s41467-019-10933-3.

[11] On definitions of different AI applications and techniques, see PI and ARTICLE 19 publication on "Privacy and Freedom of Expression In the Age of Artificial Intelligence", April 2018, available at https://privacyinternational.org/sites/default/files/2018-04/Privacy%20and%20Freedom%20of%20Expression%20%20In%20the%20Age%20of%20Artificial%20Intelligence.pdf, pp. 6-7.

Without aiming to be comprehensive, in the following sections PI describes how specific AI applications and AI applications in specific sectors negatively affect the enjoyment of the right to privacy and other human rights and illustrates the nexus between digital technologies and contemporary forms of racism, racial discrimination, xenophobia and related intolerance.

3.1 AI and facial recognition technology

Facial Recognition Technology (FRT) involves the use of cameras to capture digital images of individuals' facial features, and the automated processing of these images to identify, authenticate or categorise people. The technology extracts biometric facial data, creates a digital signature of the identified face, stores it and searches records in a database or a watchlist to find a match.[12]

FRT can be live or retrospective. Live FRT captures and stores individuals' images and facial features and matches them in real time. This means that individuals' faces are processed, stored, and scanned against a database to identify someone on the spot, whereas retrospective FRT processes facial images by checking them against a database at a later time.

The risks to human rights, in particular the right to privacy, associated with the use of FRT been well-documented.[13] These concerns are further compounded when additional analytics features increasingly rely on AI applications[14] to carry out facial recognition as noted by the High Commissioner for Human Rights.[15] PI has highlighted how the deployment of FRT is not only happening in a regulatory void but it is not subject to public and democratic scrutiny.[16]

- **Racial discrimination in the use of FRT**

FRT for identification and categorisation purposes could lead to discrimination. FRT relies on probabilistic reasoning, and as such, inevitably produces varying levels of false positive and false negatives.

Among the specific concerns around racial discrimination resulting from the use of FRT are: non-representative training data with data sets used to train AI models and algorithms do not necessarily

[12] FRT may involve the use of cameras, which can capture individuals' facial images and process them in real time ("live FRT") or at a later point ("Static" or "Retrospective FRT"). The collection of facial images results in the creation of "digital signatures of identified faces", which are analysed against one or more databases ("Watchlists"), usually containing facial images obtained from other sources to determine if there is a match.

[12] See, for example, the report of the High Commissioner for Human Rights on the "Impact of new technologies on the promotion and protection of human rights in the context of assemblies, including peaceful protests", 24 June 2020, UN doc. A/HRC/44/24. The UN Special Rapporteur on freedom of opinion and expression has called for a moratorium of the sale and use of live facial recognition (LFR) technology (Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 28 May 2019. A/HRC/41/35, para 66 (f).

[13] Report of the United Nations High Commissioner for Human Rights on the right to privacy in the digital age, 13 September 2021, UN Doc. A/HRC/48/31; UNGA Resolution on the right to privacy in the digital age, 15 December 2022, UN Doc. A/RES/77/211, page 3

[14] Note: Even though most FRT is considered to be AI-fuelled, it seems that not all FRT necessarily amounts to AI. The key difference is whether the FRT algorithm was trained using a neural networks approach. Neural networks are a method in artificial intelligence that teaches computers to process data in a way that is inspired by the human brain. Essentially, after being fed thousands of training examples, neural networks help to spot patterns and classify images without human intervention. One key application of neural networks is computer vision, which allows computers to distinguish and recognize images similar to humans. FRT is a form of computer vision.

[15] Report of the United Nations High Commissioner for Human Rights on the right to privacy in the digital age, 13 September 2021, UN Doc. A/HRC/48/31, para 32

[16] PI, "UK MPs Asleep at the Wheel as Facial Recognition Technology Spells The End of Privacy in Public", 7 November 2023, https://privacyinternational.org/long-read/5155/uk-mps-asleep-wheel-facial-recognition-technology-spells-end-privacy-public

represent the communities on which the final system will be used,[17] and there are reported concerns of lower accuracy of facial recognition technologies with certain groups with skin colour being a key factor in the bias and lack of accuracy and profiling on the basis of race, ethnicity, national origin.[18]

In his 2019 Report, the UN Special Rapporteur on the right to freedom expression noted that FRT "seeks to capture and detect the facial characteristics of a person, potentially profiling individuals based on their ethnicity, race, national origin, gender and other characteristics, which are often the basis for unlawful discrimination".[19]

- **Use of FRT by law enforcement**

Across the world we have seen government deploy FRT in public spaces for law enforcement purposes.[20]

The concerns associated with racial discrimination outlined in the above section are further exacerbating racial inequalities given historical discriminatory law enforcement practices with studies showing for example that Black people in the USA are more likely to be arrested.[21] The UN High Commissioner for Human Rights has previously noted that "Surveillance operations tend to disproportionately target minorities and marginalized communities. The use of artificial intelligence risks perpetuating such patterns of discrimination, including the use of facial recognition technologies for racial and ethnic profiling"[22].

In the UK in 2020, in the case of Ed Bridges v South Wales Police, the Court of Appeal found that the police's use of FRT breached privacy rights, data protection laws and equality laws. The case was

[17] Joy Buolamwini, Unmasking the bias in facial recognition algorithms, 13 December 2023, Excerpted from the book "Unmasking AI: My Mission to Protect What Is Human in a World of Machines," by Joy Buolamwini (2023), Published by Random House, an imprint and division of Penguin Random House LLC, available at: https://mitsloan.mit.edu/ideas-made-to-matter/unmasking-bias-facial-recognition-algorithms

[18] Larry Hardesty, Study finds gender and skin-type bias in commercial artificial-intelligence systems, MIT News, 11 February 2018, https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212 and https://www.nist.gov/speech-testimony/facial-recognition-technology-frt-0

[19] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 28 May 2019, UN Doc. A/HRC/41/35, para 12

[20] PI, The End of Privacy In Public, https://privacyinternational.org/campaigns/end-privacy-public; World Economic Forum, UNICRI, INTERPOL and Netherlands Police, A Policy Framework for Responsible Limits on Facial Recognition Use Case: Law Enforcement Investigations, Insight Report, Revised November 2022, Available at: https://www3.weforum.org/docs/WEF_Facial_Recognition_for_Law_Enforcement_Investigations_2022.pdf; Rohit Talbot, Automating occupation: International humanitarian and human rights law implications of the deployment of facial recognition technologies in the occupied Palestinian territory, International Review of the Red Cross (2020), 102 (914), 823–849, Emerging Voice, available at: https://international-review.icrc.org/articles/ihl-hr-facial-recognition-technology-occupied-palestinian-territory-914; ADC, Tecnologías de Vigilancia en Argentina, December 2021, available at: https://adc.org.ar/wp-content/uploads/2021/12/ADC-Tecnologias-de-Vigilancia-en-Argentina.pdf; INCLO, In Focus: Facial Recognition Tech Stories and Rights Harms from Around the World, January 2021, available at: https://www.inclo.net/pdf/in-focus-facial-recognition-tech-stories.pdf; Maria Badillo, Navigating the complexities of facial recognition for public security in Latin America, *International Bar Association*, 9 May 2023, available at: https://www.ibanet.org/facial-recognition-security-latin-america

[21] Alex Najibi, Racial Discrimination in Face Recognition Technology, Special Edition: Science policy and social justice, Harvard GSAS Science Policy Group, 20 October 2020, Available at: https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/

[22] Report of the United Nations High Commissioner for Human Rights on the right to privacy in the digital age, 13 September 2021, UN Doc. A/HRC/48/31, para 33. Similar concerns were expressed by the UN High Commissioner for Human Rights report on the Promotion and Protection of the Human Rights and Fundamental Freedoms of Africans and of People of African Descent Against Excessive Use of Force and Other Human Rights Violations by Law Enforcement Officers, A/HRC/47/53, para 25

supported by Liberty[23] and brought by campaigner Ed Bridges, who had his biometric facial data scanned by the FRT on a Cardiff high street in December 2017, and again when he was at a protest in March 2018. The UK Court of Appeal ruled that these deployments were unlawful and noted that the force did not take reasonable steps to find out if the software had a racial or gender bias.[24]

In the UK, FRT is also reportedly being deployed by private companies in cooperation with the police.[25] PI joined anti-poverty, homelessness, human rights, criminal justice, data, tech and privacy experts to express concerns about the planned launch of Project Pegasus, a collaboration between retailers and the police that involves the use of facial recognition technology in response to a rise in shoplifting.[26] In a joint letter sent to the CEOs of shops involved in the scheme, we flagged concerns about such a system amplifying existing inequalities within the criminal justice system given that FRT has shown to misidentify people of colour, women and LGBTQ+ people, meaning that already marginalised groups are more likely to be subject to an invasive stop by police, or may be at increased risk of physical surveillance, monitoring and harassment by workers in those retail spaces.[27]

- **Use in Schools**

Facial recognition is also being used to mediate children's access to education. This is despite the persistent evidence of discrimination within facial recognition systems, including systems being deployed by schools.[28] One student found her schools seemed to be "using a facial detection model that fails to recognize Black faces more than 50 percent of the time".[29] Some students found that - during the Covid-19 pandemic - in order to take vital exams they had to surround themselves with lights in order to ensure that the facial recognition systems recognised their face was there at all.[30] Many schools around the world have implemented these technologies without the appropriate oversight, transparency, or review.[31]

---

[23] Liberty, "Liberty Wins Ground-Breaking Victory Against Facial Recognition Tech", 11 August 2020, Available at: https://www.libertyhumanrights.org.uk/issue/liberty-wins-ground-breaking-victory-against-facial-recognition-tech/; Liberty, Legal Challenge: Ed Bridges v South Wales Police, Available at: https://www.libertyhumanrights.org.uk/issue/legal-challenge-ed-bridges-v-south-wales-police/

[24] R (on the application of Edward Bridges) v the Chief Constable of South Wales Police [2020] EWCA Civ 1058, para 201, available at: https://www.judiciary.uk/wp-content/uploads/2020/08/R-Bridges-v-CC-South-Wales-ors-Judgment.pdf; Also see: Evani Radiya-Dixit, A Sociotechnical Audit: Assessing Police Use Of Facial Recognition, October 2022, Minderoo Centre for Technology and Democracy, Available at: https://www.mctd.ac.uk/wp-content/uploads/2022/10/MCTD-FacialRecognition-Report-WEB-1.pdf

[25] Alex Hern, "MPs condemn Frasers Group's use of facial recognition cameras in stores", *The Guardian*, 23 April 2023, https://www.theguardian.com/business/2023/apr/23/mps-condemn-frasers-groups-use-of-facial-recognition-cameras-in-stores; PI, "Cooperating With Who?! Answers Needed as UK Retailer Southern Co-Op Tests Facewatch", 9 December 2020, https://privacyinternational.org/advocacy/4342/cooperating-who-answers-needed-uk-retailer-southern-co-op-tests-facewatch; BBW, "BBC – Big Brother Watch files legal challenge with the ICO against Southern Co-op's use of facial recognition systems", 26 July 2022, https://bigbrotherwatch.org.uk/2022/07/bbc-big-brother-watch-files-legal-challenge-with-the-ico-against-southern-co-ops-use-of-facial-recognition-systems/

[26] Liberty, "Rights Groups Urge Shops To Reject Facial Recognition", 29 October 2023, https://www.libertyhumanrights.org.uk/rights-groups-urge-shops-to-reject-facial-recognition/

[27] Letter available at: https://www.libertyhumanrights.org.uk/wp-content/uploads/2023/10/Liberty-Joint-letter-to-retail-CEOS-regarding-Project-Pegasus-October-2023.pdf

[28] Yoder-Himes DR, Asif A, Kinney K, Brandt TJ, Cecil RE, Himes PR, Cashon C, Hopp RMP and Ross E (2022) Racial, skin tone, and sex disparities in automated proctoring software. Frontier Education, 7:881449, Available at: https://www.frontiersin.org/articles/10.3389/feduc.2022.881449/full

[29] Todd Feathers, "Proctorio Is Using Racist Algorithms to Detect Faces", *Vice*, 8 April 2021, https://www.vice.com/en/article/g5gxg3/proctorio-is-using-racist-algorithms-to-detect-faces

[30] Morgan Meaker, "This Student Is Taking On 'Biased' Exam Software", *Wired*, 5 April 2023, https://www.wired.com/story/student-exam-software-bias-proctorio/

[31] InternetLab, "Surveillance Technologies And Education: mapping facial recognition policies in Brazilian public schools", Diagnosis and Recommendations nº 8, 2023, available at: https://internetlab.org.br/wp-content/uploads/2023/06/Educacao-na-mira-EN-03.pdf; Carolina Batista Israel,

Some data protection authorities have taken steps to prevent the technology from being used in classrooms,[32] and some other authorities - such as New York State - have banned the use of the technology in schools because of the "potentially higher rates of false positives for people of color".[33] However, many children live in countries either without an appropriate legal framework,[34] or where schools have been allowed to go ahead despite a seemingly protective legal framework.[35]

Facial recognition in schools is no more sophisticated and no less likely to discriminate than facial recognition deployed elsewhere. Its use threatens children's access to education, their right to privacy - and as is included under the UN Convention on the Rights of the Child their right to develop. Even where technology may not completely prevent access to these rights, the friction introduced into their experience of education - a friction that makes it clear the system was not designed for them - amounts to a freezing out of young black people from their right to education, and from academic spaces.

Further systems, intertwined with the technology found within facial recognition, intended to monitor children's emotions are also being deployed in schools.[36] These systems are fundamentally unsound, these technologies have been found to interpret the facial expressions of white and black people differently - attributing negative feelings, such as contempt and anger, more frequently to black people.[37] This data being recorded and used to assess children's engagement in lessons and their emotional state is deeply disturbing and dystopian. It is quite simply inappropriate for use on children in a classroom.

Based on our research and analysis, PI believes that live FRT in public places by state and non-state actors should be banned. The introduction of live FRT would result in the normalisation of surveillance across all societal levels and accordingly cast a "chilling effect" on the exercise of fundamental rights, such as our freedom of expression and freedom of assembly. Live FRT casts a chilling effect on societies and impose a sense of constant surveillance, self-restriction, and self-censoring, and normalises indiscriminate surveillance.

PI recognises that in limited circumstances and subject to strict safeguards, the deployment of static FRT by state actors such as law enforcement agencies could be justified. PI has highlighted the specific conditions in accordance with international human rights law on which any decision to use FRT technology should depend in its submission to the Scottish Parliament.[38] In summary, the minimum safeguards should include strict application of the principles of legality, necessity and proportionality,

---

Rodrigo Firmino, coordenadores; [autores] Carolina Batista Israel ... [et al.]; capa, Manoela M. Jazar - Curitiba (2023) Reconhecimento facial nas escolas públicas do Paran. Page 38, Available at: https://jararacalab.org/cms/wp-content/uploads/2023/12/RF_PR_2023.pdf

[32] Sofia Edvardsen, How to interpret Sweden's first GDPR fine on facial recognition in school, IAPP, 27 August 2019, https://iapp.org/news/a/how-to-interpret-swedens-first-gdpr-fine-on-facial-recognition-in-school/

[33] See: https://www.nysed.gov/sites/default/files/programs/data-privacy-security/biometric-determination-9-27-23.pdf

[34] PI, Stakeholder Report Universal Periodic Review 41st Session – India, April 2022, https://privacyinternational.org/advocacy/4981/right-privacy-indian-schools-universal-periodic-review

[35] Carolina Batista Israel,
Rodrigo Firmino, coordenadores; [autores] Carolina Batista Israel ... [et al.]; capa, Manoela M. Jazar - Curitiba (2023) Reconhecimento facial nas escolas públicas do Paran. Page 20-31, available at: https://jararacalab.org/cms/wp-content/uploads/2023/12/RF_PR_2023.pdf

[36] Ibid, page 37

[37] Lauren Rhue, Racial Influence on Automated Perceptions of Emotions, SSRN, 9 November 2018, available at: https://ssrn.com/abstract=3281765 or http://dx.doi.org/10.2139/ssrn.3281765

[38] For a complete analysis of facial recognition concerns, see PI, "Submission to the Scottish Parliament's Justice Sub-Committee on Policing's inquiry into facial recognition policing", November 2019, https://privacyinternational.org/advocacy/3274/submission-scottish-parliaments-justice-sub-committee-policing-inquiry-facial

prior judicial authorisation on the basis of reasonable suspicion of serious crime or serious threat to national security, strict rules on retention and destruction of personal data, prior judicial authorisation and independent monitoring and oversight, transparency in relation to the criteria used for the inclusion of individuals into watchlists, access to effective remedies, including the rights of individuals to be adequately notified of the processing of their biometric data and be given the opportunity to exercise their rights of rectification, access, erasure, as well as to challenge any processing operation before competent courts and regulators.

3.2 AI in immigration enforcement and border management

New technologies have been deployed in immigration enforcement including AI and automated decision making. These have included lie detectors at the border,[39] tracking of social media accounts,[40] language analysis[41], automated decision making about visitor visa applications[42], to the identification refugees,[43] or as part of digital border monitoring systems.[44]

As recognised by the UNSR in their 2020 report on "Racial and Xenophobic discrimination and the use of digital technologies in border and immigration enforcement",[45] there is often no or inadequate legal framework regulating the deployment of these technologies by public authorities and private security companies and in most cases, there are not effective safeguards to protect refugee and migrants against undue interferences with their privacy. Because of their heightened vulnerability, refugee and migrants are very unlikely to be able to object to the application of these technologies or to seek remedy against abuses.
Building on our joint submission to the UNSR's 2020 report[46], below we outline some key recent concerning developments:

- **Automated decision-making (ADM) in immigration systems**

Governments around the world are using migrants as the testing ground for many technology "innovations" such as biometric schemes, invasive mobile phone extraction procedures, automated decision-making systems and more.[47] In Europe, this includes the use of technology which supposedly

[39] See: iborderCtrl website, https://www.iborderctrl.eu/The-project.
[40] PI, '#PrivacyWins: EU Border Guards Cancel Plans to Spy on Social Media (for now)', 19 November 2019, https://privacyinternational.org/advocacy/3289/privacywins-eu-border-guards-cancel-plans-spy-social-media-now
[41] PI, 'The UK's Privatised Migration Surveillance Regime: A rough guide for civil society', 2021, https://privacyinternational.org/sites/default/files/2021-01/PI-UK_Migration_Surveillance_Regime.pdf
[42] Foxglove, "Legal action to challenge Home Office use of secret algorithm to assess visa applications", https://www.foxglove.org.uk/news/legal-challenge-home-office-secret-algorithm-visas.
[43] Patrick Tucker, "Refugee or Terrorist? IBM thinks its software has the answer", *Defense One*, 27 January 2016, http://www.defenseone.com/technology/2016/01/refugee-or-terrorist-ibm-thinks-its-software-has-answer/125484/.
[44] Olivia Solon, "'Surveillance society': has technology at the US-Mexico border gone too far?", *The Guardian*, 13 July 2018, https://www.theguardian.com/technology/2018/jun/13/mexico-us-border-wall-surveillance-artificial-intelligence-technology.
[45] Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance: "Racial and Xenophobic discrimination and the use of digital technologies in border and immigration enforcement", 22 September 2021, UN doc. A/HRC/48/76
[46] PI, Fundaciòn Datos Protegidos, Red en Defensa de los Derechos Digitales (R3D) and Statewatch, Submission to the UN Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance, May 2020, https://privacyinternational.org/advocacy/3939/pis-joint-submission-un-special-rapporteur-contemporary-forms-racism-racial
[47] PI, '10 threats to migrants and refugees', 8 July 2020, https://privacyinternational.org/long-read/4000/10-threats-migrants-and-refugees

identifies if a person is lying based on their 'micro-gestures', a person's origin based on their voice, and their age based on their bones.[48]

The European Union's Horizon 2020 research and innovation programme has been funding a project called iBorderCtrl, defined as "an innovative project that aims to enable faster and thorough border control for third country nationals crossing the land borders of EU Member States".[49] In addition to other features, the system undertakes automated deception detection.[50]

The system was tested at the border in Hungary, Latvia and Greece.[51] In July 2019, The Intercept used the system at the Serbian-Hungarian border: reportedly, the system failed, and the results were not disclosed.[52]

In the UK, the Home Office uses an automated triage system to assess whether a prospective marriage warrants investigation as a 'sham,' aiming to circumvent immigration laws rather than reflecting a bona fide relationship. Public Law Project (PLP) highlighted their concerns over its implementation.[53] Firstly, there is a lack of transparency and accountability surrounding the algorithm's operation and decision-making process. This opacity undermines public trust and raises questions about its reliability. Secondly, there are significant privacy and data protection concerns associated with the algorithm's use, as it collects and analyses sensitive personal information. Thirdly, there is a risk of discrimination inherent in the algorithm's design, as it may disproportionately target certain groups or individuals based on factors like nationality or ethnicity - relying on racist stereotypes. Moreover, there is a risk of unlawfulness and unfairness in its outcomes, potentially leading to unjustified accusations and legal challenges. These issues underscore the need for greater oversight and scrutiny of such algorithms to ensure they operate fairly and lawfully.

PLP has filed a legal challenge the Home Office's use of this algorithm on the grounds that the triage tool's outcomes potentially discriminate based on nationality, that the Home Office may not have fulfilled its duty to prevent discrimination and promote equality, especially when using innovative digital systems, that the Home Office's secrecy about the system violates transparency regulations under the GDPR and that the failure to ensure human/manual review of 'fail' cases contradicts government policy and could constitute a breach of the Immigration Act 2014 by delegating decision-making to a machine-learning algorithm.[54]

Also in the UK, PI has sought information under the Freedom of Information Act (FOIA) regarding the "Identify and Prioritise Immigration Cases ("IPIC") Business Rules" used by the Home Office. This is a triage tool used to prioritise and recommend "interventions" to authorities regarding migrants, assessing "the removability and level of harm posed by immigration offenders".[55] There is a pervasive lack of transparency around the tool. In its response to PI's request, the Home Office redacted the definitions of the datasets that are relevant for the IPIC tool, categories of personal data that are processed through the tool, the explanation as to how the tool will be able to meet data subject rights,

[48] Melanie Ehrenkranz, "An AI Lie Detector Is Going to Start Questioning Travelers in the EU", Gizmodo, 31 October 2018, https://gizmodo.com/an-ai-lie-detector-is-going-to-start-questioning-travel-1830126881
[49] See: iborderCtrl website, https://www.iborderctrl.eu/The-project]
[50] See: iborderCtrl Participants, https://www.iborderctrl.eu/#Project-Participants
[51] See: iborderCtrl Pilot Results, https://www.iborderctrl.eu/Pilot-Results
[52] Ryan Gallagher and Ludovica Jona, "We Tested Europe's New Lie Detector for Travelers — and Immediately Triggered a False Positive", *The Intercept*, 26 July 2019, https://theintercept.com/2019/07/26/europe-border-control-ai-lie-detector
[53] Public Law Project, "Public Law Project (PLP) — Written evidence (NTL0046)", 29 September 2021, https://committees.parliament.uk/writtenevidence/39761/pdf/
[54] Public Law Project, 'Legal action launched over sham marriage screening algorithm', https://publiclawproject.org.uk/latest/legal-action-launched-over-sham-marriage-screening-algorithm/
[55] See: https://assets.publishing.service.gov.uk/media/5cd3e056e5274a3fd5871f36/Formal_response_ICIBI_FNO_ROM.PDF

the explanation as to the purpose of processing, and the explanation as to the legal and other significant affects that the profiling undertaken through the tool could have on the data subjects.[56]

PI is preparing to challenge the Home Office's redactions, as well as other failures to disclose necessary information.

- **The impact of the EU AI Act on people on the move**

The EU AI Act aims to regulate the use of AI within the European Union, setting prohibitions and accountability requirements for 'high-risk' AI applications.[57] However, the legislation falls short in addressing the potential harms of AI when used for border and immigration.[58] Most worryingly, prohibitions on AI systems do not extend to the migration context, allowing discriminatory risk assessments, emotion recognition or predictive analytics to persist. This exemption, vehemently fought for by certain EU Member States, is a blatant denial of human rights for migrants. The EU AI Act thereby overtly strips migrants of the very few fundamental rights protections it seeks to afford the rest of the population.

In addition, the AI Act fails to recognise the potential harms of many AI systems used in migration control, such as biometric identification systems (Annex III, point 1(a)) that have been shown to discriminate, exclude and serve as means of oppression if deployed without safeguards.[59] Such systems have been excluded from the list of "high-risk" systems that attract higher transparency and accountability requirements, and AI used in large-scale interoperable EU databases is exempted from regulation until 2030.

The AI Act also lacks provisions to address the export of harmful surveillance technology, enabling EU-based companies to export banned systems to third countries - which can then be used for border externalisation.[60]
Of significant concern is the creation of exemptions for law enforcement, migration, and national security authorities, allowing them to bypass crucial transparency and oversight safeguards. Transparency requirements are limited for AI systems used in migration control, such that details of these systems (as opposed to systems deployed in the general population) don't have to be published (Article 49(4)). And high-risk AI systems deployed for migration control purposes are exempted from a key human oversight safeguard of requiring independent human verification of any identification performed by an AI system (Article 14(5)). This exemption fosters secrecy around the deployment of AI systems, undermining accountability and perpetuating racialized and discriminatory surveillance practices.

3.3 AI in welfare and access to public services

[56] See: https://www.whatdotheyknow.com/request/identify_and_prioritise_immigrat_3
[57] See: https://www.europarl.europa.eu/doceo/document/A-9-2023-0188-AM-808-808_EN.pdf
[58] PI, "Joint statement – A dangerous precedent: how the EU AI Act fails migrants and people on the move", 13th March 2024, https://privacyinternational.org/advocacy/5264/joint-statement-dangerous-precedent-how-eu-ai-act-fails-migrants-and-people-move; For more details: https://protectnotsurveil.eu/
[59] Ben Hayes, Migration and displacement. Migration and data protection: Doing no harm in an age of mass displacement, mass surveillance and "big data", International Review of the Red Cross (2017), 99 (1), 179–209.A Available at: https://international-review.icrc.org/sites/default/files/irrc_99_12.pdf
[60] PI, "The EU, the Externalisation of Migration Control, and ID Systems: Here's What's Happening and What Needs to Change", 15 October 2021, https://privacyinternational.org/long-read/4651/eu-externalisation-migration-control-and-id-systems-heres-whats-happening-and-what

Current and emerging AI supported processes to access social welfare and public services are designed and managed in a way that comes at the cost of everyone's privacy, dignity and autonomy.

From the stage of eligibility and registration to access benefits, recipients need to turn over vast amounts of personal data – about their employment, their health conditions, their relationship status – which is processed by AI applications to make (or support the making of) decision related to access to social welfare benefits.[61]

Concerns about the negative impact of the use of AI applications in the welfare context have already been expressed by UN human rights experts[62] and national courts are beginning to rule against these systems on the grounds that they fail to comply with human rights law.[63]

Social protection systems around the world are increasingly 'conditional', meaning that aspects of state support, usually financial or practical, are dependent on claimants complying with a set of rules or conditions. These processes are increasingly tied to rigid digital identification systems and determined by algorithmic and automated decision making processes.[64] Those who fail to comply with the rules can find themselves automatically cut-off from welfare programs, have their assistance reduced or are subject to sanctions and fines. In some cases the most vulnerable groups of the population are subject to particularly intrusive level of control and surveillance via digital technologies.[65]

- **Discriminatory effects of AI applications in welfare systems**

At every stage of the decision-making process in the provision of social services, automation is being built into the system. From automated digital identity verification,[66] to eligibility assessments and so-called 'fraud' detection mechanisms[67]. Automating these processes while failing to build in sufficient safeguards which require human intervention and review has led to discrimination and unjust sanctions against people who are eligible for support.

- ***Discriminatory effects:*** It has been widely recognised that these practices have had discriminatory effects.[68] Using personal data points about individuals who are seeking to access social protection, such as their sex, age, place of residence, immigration status, ethnicity, history of employment, marriage status etc., to 'profile' them increases the risk of discrimination and exclusion against specific communities. This was recently recognised by a Dutch court after assessing the impact of a risk profiling method known as "System Risk

[61] PI, "When Big Brother Pays Your Benefits", https://privacyinternational.org/taxonomy/term/675
[62] Report of the UN Special rapporteur on extreme poverty and human rights, 11 October 2019, UN doc. A/74/48037
[63] PI, "The SyRI case: a landmark ruling for benefits claimants around the world", 20 February 2020, https://privacyinternational.org/news-analysis/3363/syri-case-landmark-ruling-benefits-claimants-around-world; and Tijmen Wisman, "The SyRI Victory: Holding Profiling Practices to Account", 23 April 2020, available at: https://digitalfreedomfund.org/the-syri-victory-holding-government-profiling-to-account/7/; https://gmcdp.com/gmcdp-foxglove-legal-challenge-department-work-and-pensions-dwp-fraud-algorithm
[64] See: PI, "Stage 1 - Applying for social benefits: facing exclusion", https://privacyinternational.org/news-analysis/3112/stage-1-applying-social-benefits-facing-exclusion
[65] See: PI, "What is an Aspen Card and why does it need reform?", 23 February 2021, https://privacyinternational.org/explainer/4425/what-aspen-card-and-why-does-it-need-reform
[66] PI, "Exclusion by design: how national ID systems make social protection inaccessible to vulnerable populations" 29 March 2021, accessed online: https://privacyinternational.org/long- read/4472/exclusion-design-how-national-id-systems-make-social-protection-inaccessible.
[67] PI, "Stage 3: The policing of social benefits: punishing poverty", 7 August 2019, https://privacyinternational.org/node/3114
[68] See: UN General Assembly, "Report of the Special Rapporteur on extreme poverty and human rights," A/74/493, 11 October 2019, and United Nations High Commissioner for Human Rights, The right to privacy in the digital age, 13 September 2021, UN Doc. A/HRC/48/31

Indicator" ("SyRI") which was being used by the Dutch government to detect individual risks of welfare fraud.[69] This profiling method "was primarily deployed in poor neighbourhoods" where "many residents are more likely to be immigrants and/or from racial and ethnic minority backgrounds."[70] Further, the risk models that were being relied on were secretive, and made it "impossible for citizens to 'defend themselves against the fact that a risk report had been submitted against them."[71] Using software which analyses data to profile welfare recipients without building-in safeguards that correct for system errors or unlawful discrimination can unfairly exclude entire groups of people from accessing social protection by making incorrect determinations about eligibility,[72] miscalculating welfare benefits, and incorrectly flagging individuals for "fraud".[73]

3.4 AI in health

The use of digital technologies including AI within healthcare has been expanding rapidly in recent years. While these technologies may offer new and efficient means to assist with medical diagnosis and to streamline services, the adoption of AI technologies in the delivery of healthcare should not be driven by 'technosolutionism' at a cost to fundamental rights, in particular the right to privacy and freedom from discrimination.[74] For many years, Privacy International has been raising the risks that digital health innovations and technologies, including AI, pose to human rights, both in our own work, and in our collective work as part of the Digital Health and Rights Consortium.[75]

AI technologies have been deployed widely within healthcare; from AI enabled medical devices, to sophisticated algorithms to interpret vast numbers of patients' electronic health records, to managing public health interventions. For example, AI technologies were used throughout the Covid-19 pandemic to study the virus to prevent spread and future outbreaks and to analyse its public health impacts.[76]

- **Examples of racial discrimination in healthcare**

There have already been well-documented cases of AI technologies and algorithm bias in healthcare leading to racial discrimination and interference with the enjoyment of human rights, and we have observed some systemic and structural concerns associated with racial discrimination in the enjoyment of the right to health. These highlight the need for the decision-making around the use of

[69] PI, "The SyRI case: a landmark ruling for benefits claimants around the world", 20 February 2020, available online at: https://privacyinternational.org/news-analysis/3363/syri-case-landmark-ruling-benefits-claimants-around-world; and Tijmen Wisman, "The SyRI Victory: Holding Profiling Practices to Account", 23 April 2020, accessed online: https://digitalfreedomfund.org/the-syri-victory-holding-government-profiling-to-account/7/

[70] Digital Freedom Fund, "NJCM, Platform Bescherming Burgerrechten and others v the Netherlands (the SyRI case): Case facts at a glance," accessed online: https://digitalfreedomfund.org/case-analyses/njcm-platform-bescherming-burgerrechten-and-others-v-the-netherlands/.

[71] Ibid, n11.

[72] Ibid, n3 at paras. 21 and 22, page 9

[73] See: PI, "Stage 3 – The policing of social benefits: punishing poverty", 7 August 2019, https://privacyinternational.org/node/3114

[74] See: PI, "Digital Health: what does it mean for your rights and freedoms", https://privacyinternational.org/long-read/4671/digital-health-what-does-it-mean-your-rights-and-freedoms

[75] For more information: https://warwick.ac.uk/fac/cross_fac/cim/research/digital-health-rights/; Also see: Digital Health and Rights Project, "Digital health and human rights of young adults in Ghana, Kenya, and Vietnam: final project report", Global Health Centre Report ; 2022, available at: https://repository.graduateinstitute.ch/record/300591?v=pdf

[76] See: PI, Submission for the UN High Commissioner for Human Rights' report on the right to privacy and artificial intelligence, June 2021, https://privacyinternational.org/advocacy/4538/privacy-internationals-submission-un-report-right-privacy-and-artificial-intelligence

AI in health sector to be informed by a current understanding of existing structural inequalities, such as race, gender, socioeconomic class, geography, language and disability.

Below we outline some of these[77]:

- ***Racial discrimination in accessibility:*** When racially discriminatory structures and policies are embedded within health-related policy domains, this impacts accessibility to healthcare. For example, all over the world governments make access to public services, including healthcare, conditional on the provision of a national digital identity.[78] This has severe ramifications for those already discriminated against who have obstacles in registering for such an identity document including ethnic minorities and migrant communities.[79] Furthermore, "structural and systemic factors, such as racism, gender inequality, socioeconomic inequalities and lack of the underlying social determinants of health" [80] have been reported to create digital divides in the access to digital health. Women and girls, migrants, those living in rural areas, people on low incomes and with less education, and disadvantaged ethnic minorities, all have less access to the internet[81] as well as lack of familiarity with online health platforms and tools[82].

- ***Racial discrimination in acceptability and sensitivity:*** Racialised surveillance policies and practices by governments and companies have created an environment of mistrust because of increased tracking, monitoring and surveillance of communities at risk and subject to racial discrimination.[83] Access to digital health data by law enforcement and other security agencies have contributed to disproportionately targeting racial and ethnic minorities for surveillance and law enforcement purposes.[84]

- ***Racial discrimination in quality and adaptability:*** Digital health technologies may ingrain racial bias. For example, the UK government commissioned an independent review into the equity of medical devices and the role they play in perpetuating discrimination of minority ethnic people, women and people from deprived communities and the risks of poorer

[77] For more detailed analysis see: PI and Stopaids, Joint response to the call for input on the Draft General Recommendation n°37 on Racial discrimination in the enjoyment of the right to health, August 2023, available here: https://www.ohchr.org/en/calls-for-input/2023/call-contributions-draft-general-recommendation-ndeg37-racial-discrimination

[78] See: PI, Identity Crisis, https://privacyinternational.org/campaigns/identity-crisis

[79] Haki Na Sheria, "Biometric Purgatory: How the double registration of vulnerable Kenyan Citizens in UNHCR Database left them at risk of statelessness" (2021).Double Registration Report - 3.0.pdf; PI, "Exclusion and Identity: Life without an ID", 14 December 2018, https://privacyinternational.org/long-read/2544/exclusion-and-identity-life-without-id

[80] Digital Health and Rights Project Consortium and Sara L M Davis, Towards digital justice: participatory action research in global digital health, BMJ Glob Health. 2022; 7(5). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9114965/

[81] ITU, Measuring digital development: Facts and Figures 2022 https://www.itu.int/hub/publication/d-ind-ict_mdd-2022/; Laura Silver, "Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally", Pew Research Center (5 February 20219), https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/#fn-43448-1; Bhaskar Chakravorti, "How to Close the Digital Divide in the U.S.", Harvard Business Review (20 July 2021) https://hbr.org/2021/07/how-to-close-the-digital-divide-in-the-u-s

[82] Digital Health and Rights Project Consortium, "Digital health and human rights of young adults in Ghana, Kenya, and Vietnam: Final project report", p. 14. (2022), https://stopaids.org.uk/wp-content/uploads/2022/11/2022_11_DHRP_research_report_final-3.pdf; Asociación por los Derechos Civiles, Privacy is Health: A preliminary review of the legal framework and technological developments on electronic health records and telemedicine in Argentina, March 2021, pp. 16-17, https://adc.org.ar/wp-content/uploads/2021/06/ADC-Privacy-is-health.pdf

[83] Report of the UN Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance on: Racial discrimination and emerging digital technologies: a human rights analysis, 18 June 2020, UN Doc A/HRC/44/57, para 38-43

[84] Report of the UN Special Rapporteur on the right of everyone to the enjoyment of the highest attainable standard of physical and mental health on: Digital innovation, technologies and the right to health, 21 April 2023, UN Doc. A/HRC/53/65, para 62

healthcare outcomes.[85] The report evidenced biases at every stage of the lifecycle of medical tools and devices which are then magnified in algorithm development and machine learning. In particular, the report confirmed a link between pulse oximetry devices, racial bias and Covid-19. These were widely used devices during the Covid-19 pandemic to measure low oxygen levels in the blood, which were found to be overestimating the amount of oxygen in the blood of people with dark skin and Black and minority ethnic people.[86] Other AI enabled devices, such as dermoscopes used in dermatology, which are used to capture and help interpret images of skin lesions have been attributed to the under-diagnosis of skin cancer, as they do not cater as well for non-White skin. The consequences could include increased false negative error rates for skin cancer detection and delayed treatment for patients from some ethnic groups.[87]

- ***Discrimination in the right to control one's health and body:*** A project carried out by an Argentinean local government in partnership with Microsoft, used AI applications to predict teen pregnancy. It was built on a database that captured the data of 200,000 female residents of Salta, including highly sensitive data ranging from nationality, ethnicity and disability status to access to hot water. The aim was to predict which girls from low-income areas would become pregnant in the next five years, although it was never made clear how the information would be used. Despite this, the project has now been expanded to other provinces in Argentina. Concerns were also expressed that this was yet another tool to control the bodily autonomy of economically disadvantaged communities by preventing and avoiding abortions.[88]

- ***Lack of representation of affected communities:*** Medical researchers have called for the design of AI models to be done in collaboration with healthcare workers and patients to understand how these could be applied in practice and with what implications.[89] In particular, this requires ensuring the meaningful participation of the communities where these tools are intended to be deployed, in particular those who have experienced systematic discrimination such as women, young people, indigenous populations, LGBTQ+ people, and displaced persons, among others. Understanding AI technologies and their impact on the right to non-discrimination and the right to health requires strengthening the knowledge, skills, and competences of these groups to enable them to comprehend and exercise their rights while critically analysing the uses of AI-driven tools.

- ***Abuse, misuse, and extraction of health data:*** There are well-documented concerns with data sharing in the health sector, and data being used for other purposes for which they were intended.[90] Research conducted by PI and others CSOs into period-tracking apps have shown how extensive data collection and data-sharing practices by companies directly impact on

[85] Nicola Davis, "UK report reveals bias within medical tools and devices", *The Guardian*, 11 March 2024, https://www.theguardian.com/society/2024/mar/11/medical-tools-devices-healthcare-bias-uk
[86] See: https://assets.publishing.service.gov.uk/media/65e89e9e62ff48001a87b2d8/equity-in-medical-devices-independent-review-report-web-accessible.pdf
[87] Ibid
[88] Peña, P. and Varon, J., Teenager pregnancy addressed through data colonialism in a system patriarchal by design, 3 May 2021 (updated 26 April 2022), https://notmy.ai/news/case-study-plataforma-tecnologica-de-intervencion-social-argentina-and-brazil/
[89] The Lancet, Artificial intelligence for COVID-19: saviour or saboteur?, January 2021, https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30295-8/fulltext
[90] Centre for Internet & Society, "Big Data and Reproductive Health in India: A Case Study of the Mother and Child Tracking System", 17 October 2019, https://cis-india.org/raw/big-data-reproductive-health-india-mcts; PI, "No Body's Business But Mine: How Menstruation Apps are Sharing Your Data", 9 September 2019, https://privacyinternational.org/long-read/3196/no-bodys-business-mine-how-menstruations-apps-are-sharing-your-data

users' privacy.[91] There also concerns about AI technologies may embed or perpetuate racial discrimination with data being extracted from racialised people and communities in LMICs and is then used for fuelling AI technologies design and deploy in high-income countries, amplified past inequalities and creates new ones.[92] Furthermore, research has exposed that AI software used in hospitals was equating 'health care spending with health' consequently the software 'routinely let healthier white patients into the programs ahead of black patients who were sicker and needed them more'.[93]

International organisations such as the World Health Organisation[94] and UNDP[95], International human rights mechanisms and bodies have been sought to address these concerns as exemplified with the report of the UNSR on the right to health on digital innovation and technologies, and the Committee on the Elimination of Racial Discrimination (CERD)'s the Draft General Recommendation n°37 on Racial discrimination in the enjoyment of the right to health.[96]

With regard to the use of AI, the UN SR on the right to health called for the coding of artificial intelligence tools used in medical diagnostics, treatment, monitoring, reporting and information to adhere to principles of non-discrimination, in part to ensure quality.[97] Furthermore, that a global effort must be undertaken to encourage and invest in the creation of digital public goods: open-source software, open data, open artificial intelligence models, open standards and open content.

The medical and research community has also challenged the lack of transparency and regulatory void in which these AI technologies are being deployed. Organisations such as Ada Lovelace Institute, Connected by Data and Just Treatment have conducted research into the inequalities in data-driven health systems and digital health services in the UK.[98]

3.5 AI in employment

Workers around the world are increasingly submitted to intrusive data collection by their employers, often deployed through invasive surveillance technologies. From warehouses to gig economy

[91] PI, "No Body's Business But Mine: How Menstruation Apps are Sharing Your Data", 9 September 2019. Available at:https://privacyinternational.org/long-read/3196/no-bodys-business-mine-how-menstruations-apps-are-sharing-your-data; PI, "We asked five menstruation apps for our data and here is what we found", 4 December 2020, https://privacyinternational.org/long-read/4316/we-asked-five-menstruation-apps-our-data-and-here-what-we-found; Felizi N. and Varon, J., "MENSTRUAPPS – How to turn your period into money (for others), Coding Rights, https://chupadados.codingrights.org/en/menstruapps-como-transformar-sua-menstruacao-em-dinheiro-para-os-outros/

[92] Karen Hao, Artificial intelligence is creating a new colonial world order, 19 April 2022, MIT Technology Review, https://www.technologyreview.com/2022/04/19/1049592/artificial-intelligence-colonialism/

[93] Sharon, Begley, Discovery of racial bias in health care AI wins Stat Madness 'editors' pick', *STAT*, 6 April 2022, https://www.statnews.com/2020/04/06/stat-madness-editors-pick-racial-bias-in-health-care-ai/; See also: Report of the special rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance on: ecological crisis climate justice and racial justice, 25 October 2022, UN Doc. A/77/2990

[94] WHO, Ethics and governance of artificial intelligence for health, Guidance, June 2021, available at: https://www.who.int/publications/i/item/9789240029200

[95] UNDP, Guidance on the rights-based and ethical use of digital technologies in HIV and health programmes, July 2021, available at: https://www.undp.org/publications/guidance-rights-based-and-ethical-use-digital-technologies-hiv-and-health-programmes

[96] UN Committee on the on the Elimination of All Forms of Racial Discrimination, First draft General recommendation No. 37 (2023) on Racial discrimination in the enjoyment of the right to health, 12 May 2023, UN Doc CERD/C/CG/37

[97] Report of the UN Special Rapporteur on the right of everyone to the enjoyment of the highest attainable standard of physical and mental health on: Digital innovation, technologies and the right to health, 21 April 2023, UN Doc. A/HRC/53/65

[98] Anna Studman and Mavis Machirori, Access denied? Inequalities in data-driven health systems and digital health services, Ada Lovelace Institute, 18 September 2023, https://www.adalovelaceinstitute.org/policy-briefing/healthcare-access-denied/; See: Connected by Data and Just Treatment, Our Data Stories: Health, 22 November 2023, https://docs.google.com/document/d/12Jxf0NoWFpq2aedfHlF2PC0Ti9cnXYJh-krwuagMzqA/edit#heading=h.hz2xizfuf49s

platforms, the data collected can be used to make automated decisions that can have major consequences for those impacted by them. This reliance on algorithmic management can determine how much individuals are paid and even whether their employment or accounts are suspended or terminated.

AI technologies often supercharge these practices, with "black-box" algorithms, greatly limiting transparency and making it harder for workers to challenge decisions. Managing individuals through data-intensive surveillance can affect people's physical and mental health, put them in precarious financial positions, and result in unfair discrimination.[99] This takes away from their dignity, agency, and autonomy.

Gig economy workers in particular are at the forefront of automated decision making systems powered by AI technologies, affecting the entire employment process from recruitment to job allocation to account termination.[100] Facial Recognition systems to verify workers' identities are prone to biases and can lead to discrimination of workers of colour, preventing them from accessing jobs and livelihood.[101] Moreover, gig economy platforms often implement account termination algorithms which, when not safeguarded by meaningful human interventions, can automatically close workers accounts leaving them out of a job with little means to understand or challenge the decisions. Concerns about the platform economy reliance of AI-powered automated decision-making systems and its negative impact on workers have been echoed by the ILO[102] and is at the core of discussions towards new international labour standards.[103]

The employment and recruitment sector has also been a fertile ground for the adoption of AI technology, exposing job-seekers to the biases, malfunction and bugs of imperfect models.[104] In the process of automating data-driven mass recruitment, these tools further enhance the risk of discriminatory and biased recruitment decisions, of negative impacts on workers' autonomy and control, and of a lack of platform transparency, explainability, and accountability.[105]

4. Assessing the national legal frameworks for the use of AI

Because of the capacity of certain AI technologies to deliver discriminatory outcomes, including discriminating on the basis of race, modern anti-discriminatory laws should be adopted and implemented. These laws should prohibit certain AI applications, such as FRT in public places, FRT in schools, emotion recognition, predicting policing and social scoring. For the application of other AI technologies, national law should include clear prohibition of discrimination including on the basis of race and safeguards (as further outlined in the sections below.)

---

[99] PI, examples of algorithmic management abuses, https://privacyinternational.org/examples/examples-algorithmic-management-abuses

[100] PI, Managed By Bots, https://privacyinternational.org/campaigns/managed-by-bots

[101] PI, how a facial recognition system potentially failed to recognise a driver of colour and may have cost him his job, 13 December 2021, https://privacyinternational.org/video/4710/pas-story-how-facial-recognition-system-potentially-failed-recognise-driver-colour-and

[102] ILO, Negotiating the algorithm": Automation, artificial intelligence and labour protection, https://www.ilo.org/wcmsp5/groups/public/---ed_emp/---emp_policy/documents/publication/wcms_634157.pdf

[103] LO, New report on platform economy marks first step towards considering a new international labour standard, https://www.ilo.org/global/about-the-ilo/newsroom/news/WCMS_909150/lang--en/index.htm

[104] PI, "AI-powered employment practices: PI's response to the ICO's draft recruitment and selection guidance", 22 March 2024, https://privacyinternational.org/advocacy/5287/ai-powered-employment-practices-pis-response-icos-draft-recruitment-and-selection

[105] Grimshaw, D. (2020). International organisations and the future of work: How new technologies and inequality shaped the narratives in 2019. Journal of Industrial Relations, 62(3), 477-507. https://doi.org/10.1177/0022185620913129

As AI technologies rely significantly on the processing of personal data, thereby interfering with the right to privacy, the overarching principles of legality, necessity and proportionality should apply to any use of such technology. The data protection legal framework – requiring *inter alia* an appropriate legal basis for any data processing, fairness and transparency, ensuring purpose limitation and data minimisation, accuracy, storage limitation, integrity and security, and accountability[106] - should apply to any application of AI technology that process personal data, whether used by governments or private actors.

Modern data protection principles offer useful protection against racial discrimination. In particular, it is common for certain categories of personal data to be distinguished on the grounds that they are 'sensitive', or a special category, which, when processed, requires additional levels of protection. While there is no exhaustive list of what constitutes sensitive personal data, data pertaining to the racial or ethnic origin of individuals has become widely regarded as constituting sensitive personal data. This category of data attracts higher safeguards, including limitations on the permitted grounds for processing it.

While special category of personal data offers some protection against discriminatory outcomes of AI technologies, it is necessary to assess the effectiveness of national data protection legislation against the specific challenges posed by AI technologies. Existing data protection laws tend to provide safeguards only in relation to the processing of personal data, i.e. data from which an individual can be identified either directly or indirectly. AI technologies often blur this distinction between personal and non-personal data. Machine learning and big data analytics, for example, are fundamentally based around the idea of extracting information from data and these technologies develop ways to identify individuals from data that would historically be considered non-personal data, and therefore outside the purview of data protection law. AI applications may also blur the distinction between sensitive and non-sensitive personal data. Certain categories of personal data, similar to protected characteristics, are usually considered more sensitive, and are thus subject to higher protections. Through advanced data analytics, highly sensitive details revealing or predicting an individual's sexual life, health status, religious or political views, can be gained from seemingly mundane data.

Further, AI applications may rely on non-personal data to make or inform decisions that still negatively impact the human rights of individuals and groups affected. In these circumstances, data protection law offers little in ways of protection.

In assessing the adequacy of the national legal framework to protect human rights, it is therefore necessary to consider the wider range of laws relevant to AI technologies, including equality and anti-discrimination, consumer protection, electronic safety, product liability, competition, redress and administrative law, to name a few, together with sectoral legislation governing the deployment of AI applications in specific sectors, such as health care, criminal justice, immigration control, financial and insurance sector, etc.

- **Public procurement of AI applications**

Because of the increasing reliance by governments on AI applications for the delivery of a wide array of public services, PI believes that specific attention should be paid on the legislative framework governing public procurement of AI technologies and the safeguards to be put in place in contracting public services to private companies employing AI technologies. In our research on the public-private surveillance partnership, PI has identified some common concerns related to: lack of transparency and accountability in the procurement processes; failure to conduct due diligence assessments; growing

[106] See for further information PI, "Data Protection Guide", available at https://privacyinternational.org/data-protection-guide.

dependency on technology designed and/or managed by private companies, with loss of control over the AI applications themselves (to modify, update, fix vulnerabilities, etc.), over-reliance on the technical expertise of the private company and there are also risk of vendor lock-in. In many cases, the private company supplies, builds, operates and maintains the AI system they deployed, with public authorities not having sufficient knowledge or effective oversight. Lack of adequate legal framework is often compounded by limited enforcement safeguards provided for in contracts, resulting in limited or no venues for redress.[107]

5. Safeguards for the use of AI

There are certain specific safeguards that are key to mitigate the risks that design and use of AI technologies may result in racial discrimination or other human rights violations.

5.1 Ensuring transparency, interpretability and explainability

The opacity of complex AI applications poses significant challenges to accountability and ultimately to access to effective remedies. However, not all sources of opacity are of a technical nature and many can be addressed by adopting a human rights centred approach. This is particularly the case when opacity is due to proprietary software and trade secrets; deliberate opacity by design; or lack of technical expertise that is required to properly understand advanced processing using AI.[108]

Data protection standards, such as the right to information, articulate some transparency requirements. Information shall include the category, purpose and sources of the data processed; the existence of profiling, of automated decision-making; and the logic involved and the significance and envisaged consequences of the processing. This may be elaborated further to include for example "factors taken into account for the decision-making process, and their respective 'weight' on an aggregate level" and how a profile was built "including any statistics used in the analysis".[109] Such an obligation should apply even where the task is burdensome.[110] The domestic legal system, including intellectual property and trade secrecy, should not preclude transparency of AI applications.

5.2 Respecting human rights by design

[107] See: PI, Public-Private surveillance partnerships, https://privacyinternational.org/learn/public-private-surveillance-partnerships

[108] As noted in the Recommendation CM/Rec (2020) 1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems: "the legislative frameworks for intellectual property or trade secrets should not preclude such transparency, nor should States or private parties seek to exploit them for this purpose. Transparency levels should be as high as possible and proportionate to the severity of adverse human rights impacts, including ethics labels or seals for algorithmic systems to enable users to navigate between systems. The use of algorithmic systems in decision-making processes that carry high risks to human rights should be subject to particularly high standards as regards the explainability of processes and outputs."
https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154

[109] Article 29 Data Protection Working Party, Guidelines on Automated Decision-Making and Profiling for the Purposes of Regulation 2016/679, 17/EN. WP 251rev.01, 6 February 2018, p 27.

[110] The Article 29 Working Party Guidance on Transparency (adopted by the European Data Protection Board) has underlined that "[...] the mere fact that a database comprising the personal data of multiple data subjects has been compiled by a data controller using more than one source is not enough to lift this requirement if it is possible (although time consuming or burdensome) to identify the source from which the personal data of individual data subjects derived. Given the requirements of data protection by design and by default, transparency mechanisms should be built into processing systems from the ground up so that all sources of personal data received into an organisation can be tracked and traced back to their source at any point in the data processing life cycle."
https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227

Decisions made in the design stage of AI application have a significant impact on whether the technology is human rights compliant. Relevant factors that would affect the design of an AI application include: deciding which processes will be automated; setting the values the AI application is designed to optimise; assessing the training data used; deciding in which circumstances the AI application shall be used.[111]

Data protection legislation often includes obligations of privacy by design, requiring inter alia to ensure that the design of AI applications which process personal data limit data collection, restrict further data processing, prevent unnecessary and unauthorised access, amongst other privacy enhancing measures. These measures should all be part of the design of AI applications, but they should be complemented by considering other measures aimed at addressing other human rights risk factors. For example, testing and evaluation of AI application should consider the specific context in which they are intended to be deployed; the data to be used in testing should allow to mitigate risks of bias and discriminatory outcomes. These requirements and safeguards should be built in laws that regulate AI technologies in the relevant sectors, for example in healthcare.

5.3 Human Rights Impact Assessment

Human rights impact assessments of AI applications should be conducted at all stages of the AI applications: prior to the design, during the development, the testing, the deployment and regularly thereafter in order to identify the emerging human rights risks. These assessments not only enable the identification of the risks and corresponding mitigation strategies required to respond to them, but they also provide a framework for deciding whether to go ahead with a particular initiative. The outcomes of the assessment should result in redesign or cancellation if the risks outweigh the benefit.

While certain AI applications which carry significant risks for human rights (due to the technology used and/or the sector in which they are used, see above) require additional scrutiny, PI believes that at a minimum, an impact assessment should include privacy and data protection impact assessments as well as an assessment of other human rights likely affected by the AI application as well as potential discriminatory effects. Such assessments should consider the necessity and proportionality of any interference with privacy or other human rights, the risks to individuals and groups, and how these risks are to be addressed and mitigated.

The assessments should be conducted with the participation of affected individuals and groups, civil society actors and independent experts. The outcome of the assessment should be made public and should detailed the mitigation and oversight measures envisaged. As noted by the Committee of Ministers of the Council of Europe "confidentiality considerations or trade secrets should not inhibit the implementation of effective human rights impact assessments."[112]

5.4 Security of AI

[111] For some examples of the factors to consider see comments by Privacy Researchers on the proposals of the Office of the Privacy Commissioner of Canada (OPC) to amend the Personal Information Protection and Electronic Documents Act (PIPEDA) for ensuring appropriate regulation of artificial intelligence, https://tlpc.colorado.edu/wp-content/uploads/2020/03/2020.03.13-Academic-Researchers-Comment-on-ensuring-appropriate-regulation-of-artificial-intelligence-final-1.pdf.

[112] Council of Europe, Addressing the impacts of Algorithms on Human Rights, Recommendation of the Committee of Ministers, https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154

The security of the data, at rest and in transit, as well as the infrastructure relied upon for processing, should be protected by security safeguards against risks such as unlawful or unauthorised access, use and disclosure, as well as loss, destruction, or damage of data.[113]

When assessing the level of security for AI applications, organizations should consider central processing and data storage sites, as well as the security of remote devices where data also may be collected or received. Security measures should include appropriate mechanisms for addressing actual and suspected security breaches. PI research has shown how cheap smart phones are often marketed with pre-installed apps which not only collect personal data without users' ability to control, but are also riddled with vulnerabilities which can be easily exploited, particularly because of lack of security updates.[114] As PI's correspondence with Google outlines, big tech companies have an important role to play to ensure the security and privacy of devices, including by prohibiting certain practices which put privacy and security of users' data at risk.[115]

5.5 Independent oversight

Any deployment of AI technology should be subject to independent, effective, adequately resourced and impartial oversight. Oversight should cover all parts of the design, use and throughout the deployment of AI application.

Oversight depending on the type of technology and the sector in which it is deployed should include judicial, administrative and/or parliamentary domestic oversight mechanisms capable of verifying the legality of the use of AI, ensuring transparency and accountability. Oversight mechanisms should be able to verify the fairness and accuracy of AI application.

Oversight mechanisms must have the power and capacity to conduct regular auditing of AI applications to ensure their compliance with human rights and other standards. As noted by the UN Special Rapporteur on freedom of expression, protection of intellectual property and trade secrets cannot justify refusal of such oversight, particularly when the AI application is used by the public sector. Further, there are technical and policy options to address legitimate concerns related to proprietary technology, including allowing regulators and independent researchers access to AI applications on a confidential basis.[116]

5.6 Ensuring access to remedies – both individual and collective

Individuals should have access to an effective remedy against applications of AI technologies that affect them. As access to a remedy is dependent on the ability to know if and how one has been affected by AI applications, transparency and explainability noted above are necessary preconditions to exercise the right to seek remedy.

Individuals should have access to accessible, affordable, independent and effective judicial and non-judicial authorities with the power to receive complaints from individuals, investigate them, and take enforcement action - or refer the case to a court. As noted by the UN Special Rapporteur on freedom of expression, there are concerns whether AI applications, such as automatic response processes, to

[113] PI, "A Guide for Policy Engagement on Data Protection; The Keys to Data Protection", https://privacyinternational.org/sites/default/files/2018-09/Data%20Protection%20COMPLETE.pdf.
[114] See: PI, "Buying a smart phone on the cheap? Privacy might be the price you have to pay", 20 September 2020, https://privacyinternational.org/long-read/3226/buying-smart-phone-cheap-privacy-might-be-price-you-have-pay
[115] See: PI, "Our response to Google: Privacy isn't a Luxury", 18 August 2020 (updated 7 October 2020), https://privacyinternational.org/news-analysis/4118/our-response-google-privacy-isnt-luxury
[116] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 29 August 2018, UN doc. A/73/348

respond to complaints constitute an effective remedy, "given the lack of discretion, contextual analysis and independent determination built into such processes."[117]

Beyond individual redress, mechanisms of collective redress are an important and effective tool for accountability of AI applications. As noted above, challenges in transparency and explainability and the fact that AI systems often affect groups and communities, as well as the society more broadly, make collective complaints appropriate procedure to complement individual redress.[118]

[117] Ibid, para 41.

[118] See for some examples of PI's complaints: https://privacyinternational.org/legal-action/complaints