# Amnesty International response to the EU's proposal for an Artificial Intelligence Act ("AIA")

In April 2021, the European Commission published a proposal for a new set of rules governing the development, marketing and use of artificial intelligence (AI). As the world's first binding legal framework on AI, the proposal represents a significant turning point and marks a growing recognition by States of the urgent need to address the negative impacts of these emerging technologies on people's human rights and society. However, as it stands the proposed regulation falls far short of the measures that will be required to meaningfully protect people from harmful AI systems in the EU and globally.

This position paper sets out Amnesty International's concerns around the key gaps and shortcomings in the current proposal and includes recommendations for strengthening human rights protections in the final regulation.

The European Union must seize this opportunity to lead the way and set a high bar for AI regulation that truly protects people's rights in the digital age.

# 1. Red lines/prohibited practices

Amnesty International is calling for red lines for certain AI-systems or uses that are too harmful to be left unregulated and which should not even be allowed.

#### (i) Facial recognition and other biometric technologies

Facial recognition and remote biometric technologies that enable mass surveillance and discriminatory targeted surveillance are incompatible with human rights and should be subject to an outright prohibition in the AIA. <sup>2</sup>

## Why is this a human rights issue?

Biometric technologies are one of the use cases of AI with the most clear-cut and inherent harms to human rights. The very nature of these technologies means they have a disproportionate impact on racialized and marginalized groups and further

Al Index: POL 30/4567/2021 1/13

<sup>&</sup>lt;sup>1</sup> EU Commission, *Proposal for a Regulation laying down harmonised rules on artificial intelligence*, April 2021

<sup>&</sup>lt;sup>2</sup> Amnesty International and more than 170 organisations call for a ban on biometric surveillance, 7 June 2021, <a href="https://www.amnesty.org/en/latest/news/2021/06/amnesty-international-and-more-than-170-organisations-call-for-a-ban-on-biometric-surveillance/">https://www.amnesty.org/en/latest/news/2021/06/amnesty-international-and-more-than-170-organisations-call-for-a-ban-on-biometric-surveillance/</a>

entrench discrimination.<sup>3</sup> As the UN Special Rapporteur on contemporary forms of racism has stated, these technologies must be understood as capable of "creating and sustaining racial and ethnic exclusion in systemic or structural terms".<sup>4</sup>

The term "biometric recognition" covers a wide range of Al-applications that can be used for: (i) identification of a person, (ii) authentication (face, fingerprint, etc.) to authorise access (iii) recognition of a person's biometric features such as gait, voice, heart rate, temperature, posture etc., and (iv) (supposed) recognition of a person's emotions, behaviour, intentions, gender, sexual orientation, religious beliefs and other personal attributes.

All these biometric recognition applications have significant technical flaws in their current forms, however, technical improvements to these systems will not eliminate the threat they pose to our human rights. Biometric recognition applications have the capacity to identify, follow, single out, and track people everywhere they go, undermining human rights and representing unacceptable risks of hampering, among others, the right to privacy and data protection, freedom of expression, freedom of assembly and association (leading to the criminalization of protest and causing chilling effects), the rights to equality and non-discrimination. Even when they "work", they fuel discrimination. These tools adversely affect our human dignity in general.

Moreover, biometric recognition systems which make inferences and predictions about people's gender, emotions, intentions, behaviour or other personal attributes such as the likelihood to perform well on a task or in a job, based on biometrics suffer from serious, fundamental flaws in their scientific underpinnings. This means that the inferences they make about us are often invalid, in some cases even operationalizing eugenicist theories of phrenology and physiognomy, thereby perpetuating discrimination and adding an additional layer of harm as we are both surveilled and mischaracterized.

Al systems that claim to be able to determine people's emotional states are also based on fundamentally flawed assumptions that have a highly questionable scientific basis. For example, in 2019 a major scientific meta-study found "insufficient evidence" for the view that emotions can be inferred from facial

Al Index: POL 30/4567/2021 2/13

\_

<sup>&</sup>lt;sup>3</sup> For example, see Amnesty International's Ban the Scan campaign, <a href="https://banthescan.amnesty.org/">https://banthescan.amnesty.org/</a> and the civil society Reclaim Your Face initiative <a href="https://reclaimyourface.eu/">https://reclaimyourface.eu/</a>

<sup>&</sup>lt;sup>4</sup> E. Tendayi Achiume, UN Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance, *Racial discrimination and emerging digital technologies: a human rights analysis*, A/HRC/44/57, June 2020

movements.<sup>5</sup> The study concluded that Al-driven emotion recognition could, at the most, recognise how a person subjectively interprets a certain biometric feature of another person. Far-fetched statements, such as that Al could determine whether someone will be successful in a job based on micro-expressions or tone of voice, are simply without scientific basis.<sup>6</sup>

Although some applications of biometric recognition claim to protect people's privacy by not linking to their legal identities, they can nevertheless be used to single out individuals in public spaces, or to make inferences about their characteristics and behaviour. The use of anonymised data is not sufficient to ensure compatibility of the surveillance conducted. Moreover, these technologies always have disparate racial impacts.

Biometric recognition is often used for authentication purposes (i.e. one-to-one matching) for example to grant access to buildings, airports or phones. While this type of biometric recognition is in general less problematic, in certain cases, these systems can be built and used in a manner that equally enables problematic forms of surveillance, such as by creating large, centralised biometric databases which can be reused for other purposes. There are worrying developments of private actors compiling and sharing databases of "suspicious" individuals.

While law enforcement and public use of these technologies has attracted attention and criticism, their use by private actors can pose the same threat to our human rights. Not only when private actors engage in surveillance on behalf of governments and public agencies, but also when private actors deploy biometric recognition technologies for themselves. We have seen a surge in biometric recognition applications in workplaces, in recruitment and human resources and in commercial settings, leading to corporate surveillance and the widespread use of techniques that constitute a threat to our human rights.<sup>8</sup>

# What does the AIA propose?

Art. 5.1 (d) of the AIA aims to ban real time remote biometric identification for law enforcement, except in a small number of defined situations. The Commission's proposed prohibition of only 'real-time' remote biometric identification systems by law enforcement in public spaces, with even this use being permitted in certain

Al Index: POL 30/4567/2021

<sup>&</sup>lt;sup>5</sup> Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. Psychological Science in the Public Interest, 20(1), 1–68.

<sup>&</sup>lt;sup>6</sup> Muller C.: "The Impact of AI on Human Rights, Democracy and the Rule of Law", 2020 for Council of Europe's CAHAI

 $<sup>^{7}</sup>$  Amnesty International and more than 170 organisations call for a ban on biometric surveillance, 7 June 2021

<sup>&</sup>lt;sup>8</sup> Biometric Update, *Union warns against biometric monitoring of employees amid increase in remote work,* October 2020 <a href="https://www.biometricupdate.com/202010/union-warns-against-biometric-monitoring-of-employees-amid-increase-in-remote-work">https://www.biometricupdate.com/202010/union-warns-against-biometric-monitoring-of-employees-amid-increase-in-remote-work</a>

circumstances, is too narrow and falls far short of an outright ban.

Furthermore, enabling law enforcement to use 'real-time' remote biometric technologies – even only in limited circumstances – means that the technology and associated infrastructure will still have to be adopted and put in place. This creates a serious risk of law enforcement misuse of technologies which are inherently discriminatory and incompatible with human rights. Amnesty International has extensively documented systemic human rights concerns in European states regarding institutional racism, discrimination in law enforcement and lack of accountability regarding allegations of unlawful use of force by law enforcement officials.<sup>9</sup>

Emotion recognition systems and biometric categorisation systems are not included in the list of prohibited AI practices under title II but are instead – partially – included in the category of 'high-risk' AI systems under Article 6(2). This is disappointing given the unacceptably high risks such systems pose to human rights in many contexts. Moreover, under Annex III, emotion recognition systems only qualify as 'high-risk' when used by law enforcement and immigration authorities. For all other uses, the only explicit obligation the AIA proposal includes is to inform people that they are exposed to an emotion recognition system. This is an inadequate safeguard.

Emotion recognition and biometric categorisation systems are highly intrusive practices that purport to infer sensitive characteristics about people with high risks of discriminatory outcomes. The potential use cases of such systems are very broad, ranging from commercial use to selection procedures to law enforcement. For example, Spotify has already patented technology that would listen to people's private conversations and recognize their emotions in order to help recommend its users content such as podcasts and songs. The Chinese government is testing emotion recognition systems on Uyghurs in police stations in Xinjiang purportedly to assess their state of mind, including negative or anxious ones, reportedly intended to reach conclusions "without any credible evidence".

In response to the EU's proposal, the EU's European Data Protection Supervisor has also expressed regret that the regulation does not go further and calls for a ban

Al Index: POL 30/4567/2021 4/13

\_

<sup>&</sup>lt;sup>9</sup> Amnesty International, *Europe: Policing The Pandemic: Human Rights Violations In The Enforcement Of Covid- 19 Measures In Europe*, 24 June 2020, <a href="https://www.amnesty.org/en/documents/eur01/2511/2020/en/">https://www.amnesty.org/en/documents/eur01/2511/2020/en/</a>

<sup>&</sup>lt;sup>10</sup> Access Now, Spotify, don't spy: global coalition of 180+ musicians and human rights groups take a stand against speech-recognition technology, 19 May 2021, <a href="https://www.accessnow.org/spotify-spy-tech-coalition/">https://www.accessnow.org/spotify-spy-tech-coalition/</a>

<sup>&</sup>lt;sup>11</sup> Jane Wakefield, BBC News, *AI emotion-detection software tested on Uyghurs*, 26 May 2021, https://www.bbc.com/news/technology-57101248

on remote biometric identification in public space "whether these are used in a commercial or administrative context, or for law enforcement purposes". 12

In conclusion, this prohibition is too narrowly defined and opens the door to many forms of biometric recognition by public and private actors.

# What is Amnesty International calling for?

Given the above, Amnesty International calls for:

- an outright ban on uses of facial recognition and remote biometric recognition technologies that enable mass surveillance and discriminatory targeted surveillance.
- a ban on biometric categorisation and emotion recognition systems by both public and private actors, when they are used for surveillance in publicly accessible spaces and in spaces which people cannot avoid.
- strict conditions prohibiting the storage, sharing and (re-)use, of biometric data gathered for the purpose of biometric authentication.

# (ii) Al driven harmful manipulation

#### Why is this a human rights issue?

Human dignity is the very basis of our human rights. It ensures that every human being possesses an "intrinsic worth", which should never be diminished, compromised or repressed by others — nor by new technologies like AI. This means that all people are to be treated with respect rather than merely as objects to be monitored, sifted, sorted, scored, herded, conditioned or manipulated. Many AI systems however do exactly that, making our human dignity vulnerable.

Al has shown an enormous capability to condition people and even manipulate people into certain behaviour, leading to adverse personal, societal or democratic effects.<sup>13</sup> Political or other motives might lead to Al systems being optimised to select or prioritise particular content in an effort to coerce and influence individuals

Al Index: POL 30/4567/2021 5/13

<sup>&</sup>lt;sup>12</sup> European Data Protection Supervisor, *Artificial Intelligence Act: a welcomed initiative, but ban on remote biometric identification in public space is necessary*, 23 April 2021 <a href="https://edps.europa.eu/press-publications/press-news/press-releases/2021/artificial-intelligence-act-welcomed-initiative\_en">https://edps.europa.eu/press-publications/press-news/press-releases/2021/artificial-intelligence-act-welcomed-initiative\_en</a>

<sup>&</sup>lt;sup>13</sup> Council of Europe's Committee of Ministers, *Declaration on the Manipulative Capabilities of Algorithmic Processes*, February 2019, https://search.coe.int/cm/pages/result\_details.aspx?ObjectId=090000168092dd4b

towards certain points of view, for example during election processes. These practices can violate our rights including freedom of opinion and freedom of thought.<sup>14</sup>

Al used in media and news curation, bringing ever more 'personalised 'online content and news to individuals, raises concerns. Search engines, algorithmic recommender systems and news aggregators often are opaque, both where it comes to the data they use to select or prioritise the content, but also where it comes to the purpose of the specific selection or prioritisation.<sup>15</sup> The business models of many companies in the tech sector are based on online advertising; in order to drive ad revenues, they select and prioritise content that will keep people on their platform.

If personal AI predictions become very powerful and effective, they may even threaten to undermine human agency and autonomy.<sup>16</sup>

#### What does the AIA propose?

Art. 5 (1) (a) and (b) are very unclear as to what it is that they aim to prohibit and contain conditions that reduce the actual applicability of the prohibitions to very unusual and rare practices, making them virtually 'empty'. Subliminal techniques beyond a person's consciousness in order to materially distort a person's behaviour in a manner that causes or is likely to cause that person or other persons physical or psychological harm are prohibited under the AIA. The same goes for AI-driven exploitation of people's vulnerabilities. However, Article 5(1)(b) only prohibits the exploitation of specific groups of people that are vulnerable due to their age, physical or mental disability, while failing to capture vulnerabilities that go beyond these categories.

The EU and Member States should take this opportunity to have these prohibitions cover the use of AI systems that harmfully condition, manipulate or exploit people. The feasibility of such a prohibition depends on where the line is drawn between acceptable and non-acceptable conditioning and manipulation. Currently, only those practices are prohibited that cause physical or psychological harm. AI-driven manipulation and exploitation can however also cause material harm as well as

Al Index: POL 30/4567/2021 6/13

<sup>&</sup>lt;sup>14</sup> One of the components of the right to freedom of opinion under Article 19(1) ICCPR is the right not to be manipulated. See Evelyn Aswad, *Losing the Freedom to Be Human*, Columbia Human Rights Law Review, Vol. 52, 2020

 $<sup>^{\</sup>rm 15}$  Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, 3(1), 2053951715622512.

<sup>&</sup>lt;sup>16</sup> Taddeo, M., & Floridi, L. (2018b). How Al can be a force for good. Science, 361(6404), 751–752.

harm to multiple human rights and even democracy and the rule of law.

#### What is Amnesty International calling for?

The use of AI systems that condition and manipulate people into certain behaviour without their knowing and having adverse effects such as material and/or immaterial damage or a violation of fundamental rights, is unacceptable and should be banned.

#### 2. Loopholes

The proposed AIA does not go far enough in protecting people's human rights and should be more ambitious to effectively protect them. In particular, the AIA contains several 'loopholes' or 'missed opportunities'.

#### (i) Self-assessment and standardisation

A fundamental concern with the AIA proposal is the over-reliance on the providers of AI systems to self-assess their compliance with the regulation. Under Chapter 5 of Title III, the vast majority of AI systems designated to be 'high-risk' are only subject to a limited conformity assessment procedure based on internal control, without any third-party conformity assessment. This amounts to a very weak safeguard against human rights abuses, especially given the recognised high-risk nature of such systems. This ranges from AI used for evaluating creditworthiness, work performance or the eligibility for public benefits, to crime prediction AI and AI used to examine visa applications. Such uses of AI pose real dangers to human rights yet are left to self-certification.

The only high-risk systems that do require a more stringent conformity assessment including the involvement of an independent third party are AI systems intended to be used as safety components, and AI systems for remote biometric identification. In the latter case, given that as set out above such use should properly be subject to an outright prohibition, the more stringent conformity assessment measures are insufficient.

Moreover, the requirements for high-risk AI against which the assessment must be done, will be standardised through opaque processes of standardisation, which is an industry dominated process. There is a risk this will obstruct the participation of representatives of affected groups and other civil society organisations in rule-making procedures that have highly consequential implications for human rights.<sup>17</sup>

AI Index: POL 30/4567/2021 7/13

<sup>&</sup>lt;sup>17</sup> Michael Veale and Frederik Zuiderveen Borgesius, *Demystifying the Draft EU Artificial Intelligence Act,* July 2021

The EU and member states must ensure that competent administrative and judicial authorities have the mandate to enforce compliance with the regulation, and put in place third-party verification for all high-risk AI systems.

#### (ii) Sandboxes

The term "sandbox" has become a buzzword in the sphere of regulating disruptive technologies which pose novel challenges to regulators and for which impact little knowledge exists. The AIA proposal defines the objective of the AI sandboxes as "to foster AI innovation (...) with a view to ensuring compliance of the innovative AI systems with the regulation and relevant Union and Member State legislation" and to "enhance legal certainty for the innovators and the competent authorities' oversight and understanding of the opportunities, emerging risks and the impacts of AI use, and to accelerate access to markets, including by removing barriers for small and medium enterprises (SMEs) and start-ups". 18

According to the Sherpa report on Regulatory Option for AI and Big Data,<sup>19</sup> and an analysis by Katerina Yordanova,<sup>20</sup> the most important values and principles in the implementation of sandboxes for AI are transparency in design, operation and outcomes, and close communication and cooperation with stakeholders. The AIA does not reflect this.

While the AIA specifies certain conditions for AI systems in the sandboxes, the provision provides a carte blanche for the use of personal data for other purposes, generating a legal basis to override the GDPR principle of purpose limitation (art. 5.1 (b) GDPR and 6.4 GDPR) for certain AI developments.

The AIA allows the use of personal data beyond its original purpose limitation for the development of AI systems for "prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties". These types of AI uses should be excluded from the sandbox regime.

Furthermore, the sandbox approach should not be seen as a silver bullet for dealing with problematic AI systems which push the boundaries of the AIA. Instead, they should be used as a precautionary step for the regulators to keep up with new

Al Index: POL 30/4567/2021 8/13

<sup>&</sup>lt;sup>18</sup> See recital 72 of the AIA proposal.

<sup>19</sup> https://www.project-sherpa.eu/regulatory-options-for-smart-information-systems-sis-ai-and-big-data-analytics/

<sup>&</sup>lt;sup>20</sup> https://www.law.kuleuven.be/citip/blog/the-shifting-sands-of-regulatory-sandboxes-for-ai/

<sup>&</sup>lt;sup>21</sup> See Article 54 of the AIA proposal.

regulatory challenges that future AI systems could present.

#### (iii) Categorisation choices of Annex III for high-risk AI

The 'list-based' approach for high and medium risk AI (in ANNEX III and art. 52 AIA) can lead to the legitimisation, normalisation and mainstreaming of several AI practices that have been widely criticised and will likely lead to human rights violations. Moreover, this approach assumes that these uses bring enough social benefits to justify their acceptance and that the requirements set out in the AIA for high and medium risk AI will sufficiently mitigate all possible risks of harm to 'health, safety and fundamental rights'.

# (iv) Legacy high-risk AI and AI for EU's centralised information systems for borders and security

Article 83 of the AIA contains two major loopholes. It excludes high-risk AI systems already placed on the market or put into service before application of the AIA (unless after that date the AI system undergoes significant changes). Amnesty International stresses that these 'legacy high-risk AI systems' must also be covered by the AIA, to prevent states and private sector actors from fast tracking the deployment of high-risk AI systems to avoid compliance requirements.

It also excludes AI systems which are components of large-scale European IT systems in the realm of "freedom, security and justice" already put into service before the application of the AIA. These systems are the Schengen Information System (SIS), the Visa Information System (VIS), Eurodac, the Entry/Exit System (EES), the European Travel Information and Authorisation System and the European Criminal Records Information System on third-country nationals and stateless persons (ECRIS-TCN). These systems are also known as the "EU's centralised information systems for borders and security". According to a recent paper by the EPRS39 "These systems are increasingly incorporating biometric technologies for the purpose of identity verification or identification." The EPRS paper identifies (a.o.) the use of automated fingerprint identification by SIS, Eurodac VIS, EES and ECRIS-TCN. Facial recognition is expected to be used by all systems except the European Travel Information Authorisation System in the near future. A number of EU- funded projects and initiatives have explored and piloted emotion recognition technologies at the EU border.

Excluding these AI uses from the scope of the AIA (as long as they are put into place before the application of the AIA) enables EU's centralised information systems for borders and security to implement prohibited and high-risk AI systems in the near future, only having to apply the AIA when the entire system is

Al Index: POL 30/4567/2021 9/13

evaluated. These systems should not be excluded from the scope of the AIA.

# 3. Human rights safeguards

States are the primary duty bearers for the realisation of human rights, and have the obligation to respect, protect and fulfil human rights. This obligation means that states must take positive action to facilitate the enjoyment of human rights, including implementing the following human rights safeguards in laws and policies that regulate the use and placing on the market of artificial intelligence systems.

#### (a) Transparency

Transparency is a core principle of good governance and an important human rights safeguard.<sup>22</sup> The opacity of some AI systems and the inability to scrutinize the way these systems work (the 'black box phenomenon') means individuals may be unaware of how (semi)automated decision-making affects their rights and freedoms and whether this process was discriminatory. The asymmetry of information between those negatively impacted by AI systems and those developing and using them, stresses the need to reinforce mechanisms of transparency.<sup>23</sup> The AIA lacks transparency measures that ensure the provision of necessary information to the subjects of AI systems and the wider public.

The right to meaningful information about the logic involved in automated decisions as laid down in the General Data Protection Regulation<sup>24</sup> is not sufficient in the context of AI systems as it only applies when the decision was taken solely by an automated system, excluding instances with even a limited and trivial degree of human involvement – and for example subjected to automation bias, severe time pressure or lack of information when reviewing the decision – rendering the right inapplicable. It is therefore important to include such transparency measures in the AIA.

Amnesty International recommends the EU and Member States to create

Al Index: POL 30/4567/2021 10/13

<sup>&</sup>lt;sup>22</sup> https://www.ohchr.org/EN/Issues/Development/GoodGovernance/Pages/AboutGoodGovernance.aspx

<sup>&</sup>lt;sup>23</sup> CAHAI, Feasability Study, para. 85.

 $<sup>^{24}</sup>$  See article 22 in conjunction with articles 12-17 GDPR. According to article 14 (2) (g) GDPR, the controller of personal data has the obligation to provide the data subject with information regarding the existence of automated decision-making, including profiling, as referred to in article 22 GDPR and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

maximum possible transparency around the use of artificial intelligence systems and to promote the explainability and intelligibility of artificial intelligence systems, so that 1) the impact on the rights and freedoms of individuals and groups can be effectively scrutinised by the individual, the wider public, the supervisory authorities and independent entities, 2) responsibilities can be established, and 3) actors can be held accountable. In this regard, Amnesty International endorses the proposal by Algorithm Watch to widen the scope of the proposed EU Database of AI systems.<sup>25</sup>

#### (b) Human rights due diligence and impact assessments

States and the European Union must take positive action to protect human rights in the context of AI systems. This includes laying down the obligation for designers, developers and deployers of algorithmic systems to identify, prevent, and mitigate potential and actual adverse human rights impacts. Such requirements are missing in the AIA. The rapidly developing and iterative nature of the technology means that new risks are likely to emerge in a variety of unforeseen contexts. As such, it is vital that all AI systems are subjected to an ongoing process of human rights due diligence. This includes identifying the actual and potential impacts on human rights, such as through a human rights impact assessment (HRIA), and demonstrating appropriate action has been taken to address these risks, on an ongoing, transparent and dynamic basis.<sup>26</sup>

Amnesty International recommends the EU and Member States to implement mandatory human rights due diligence applicable to the private and public sector, including to law enforcement authorities, to identify, prevent, and mitigate potential and actual adverse human rights impacts in the design, development and deployment of AI systems. The HRIA must be carried out in the design, execution, and evaluation phases of artificial intelligence systems. Public sector bodies should at a minimum conduct a HRIA for all AI applications.

#### (c) Oversight

Independent oversight mechanisms are central to ensuring that states and businesses comply with their respective human rights obligations and responsibilities. Oversight plays an important role in the context of AI systems that

Al Index: POL 30/4567/2021 11/13

<sup>&</sup>lt;sup>25</sup> Algorithm Watch, Draft AI Act: EU needs to live up to its own ambitions in terms of governance and enforcement, 4 August 2021, <a href="https://algorithmwatch.org/en/eu-ai-act-consultation-submission-2021/">https://algorithmwatch.org/en/eu-ai-act-consultation-submission-2021/</a>

 $<sup>^{26}</sup>$  Amnesty Int Submission to the European Commission's Consultation White Paper on AI (June 2020), Rec 3

have potential adverse impact on human rights due to their opaque nature, with harms that are difficult to identify without extensive expertise. <sup>27</sup>

Human rights oversight is insufficient in the AIA. The use of algorithms is often hidden or unknown, making it difficult or impossible to know whether there is an impact on human rights. Significant expertise on all relevant human rights aspects, such as data protection compliance, data science, algorithmic systems and (semi-)automated decision-making is required to understand the possible dangers of AI systems and to effectively keep up with the introduction and ongoing development of such systems.

Amnesty International recommends establishing independent supervisory authorities that advise on, monitor and enforce human rights obligations and responsibilities in artificial intelligence systems and enforces the AIA. Those supervisory authorities must have the legal mandate, capacity and resources to investigate artificial intelligence systems, issue binding decisions and guidance, as well as penalize entities that design, develop or deploy artificial intelligence, algorithmic systems and (semi-)automated decision-making systems that result in human rights abuse or harm or a high and unmitigated risk thereto. The supervisory authorities must have access to the documentation, HRIA, training data, data categories and algorithms to examine the system and its outcomes in terms of respecting, promoting and fulfilling human rights. The EU and member states must ensure that the supervisory authorities will have capacity and expertise on all relevant human rights aspects, such as anti-discrimination expertise, data protection compliance, data science, and artificial intelligence in order to effectively keep up with the introduction and ongoing development of in artificial intelligence systems in society.

#### (d) Redress

A complaints and redress mechanism for individuals that have suffered human rights harm from the use of any Al-system that falls within the scope of the AlA is missing.

When human rights violations and abuses occur, international law requires that the perpetrator is held accountable, and the victim receives an effective remedy.<sup>28</sup>

Al Index: POL 30/4567/2021 12/13

<sup>&</sup>lt;sup>27</sup> McGregor, Murray and Ng, International Human Rights Law as a Framework for Algorithmic Accountability, p. 331.

<sup>&</sup>lt;sup>28</sup> Article 8 of the Universal Declaration of Human Rights; Article 2(3) of the International Covenant on Civil and Political Rights; Article 2 of the International Covenant on Economic, Social and Cultural Rights; Article 6 of the International Convention on the Elimination of All Forms of Racial Discrimination; Article 13 of the European

Securing justice and redress both for individuals directly affected and in order to protect the rights of society as a whole are vital elements of human rights.<sup>29</sup> The UN Special Rapporteur on Freedom of Expression has highlighted how AI systems often interfere with the right to remedy.<sup>30</sup> There is an inherent challenge around informing affected individuals, as "individuals are not aware of the scope, extent or even existence of algorithmic systems that are affecting their rights". This opacity is exacerbated because algorithms are constantly adapting and changing, such that even the designers of the system may not or poorly be able to explain how they reached their outcomes.<sup>31</sup> The onus is on governments to make algorithmic systems visible, allow outputs or impacts to be queried and appealed, and create accessible and practical routes for remedy and redress when human rights are negatively impacted.<sup>32</sup>

Amnesty International recommends including in the AIA a complaints and redress mechanism for individuals that have suffered human rights harm. The mechanism must 1) facilitate equal and effective access to justice; 2) adequate, effective, and prompt reparation for harm suffered and 3) access to relevant information concerning violations and reparation mechanisms.<sup>33</sup>

Convention for the Protection of Human Rights and Freedoms; Article 47 of the Charter of Fundamental Rights of the European Union.

Al Index: POL 30/4567/2021 13/13

<sup>&</sup>lt;sup>29</sup> Right to remedy is also discussed in Amnesty International, Injustice Incorporated: Corporate Abuses and the Human Right to Remedy, 2014.

<sup>30</sup> David Kaye, 2018, para 40.

<sup>&</sup>lt;sup>31</sup> Al Now Institute, Annual Report 2017, p. 30, https://ainowinstitute.org/Al\_Now\_2017\_Report.pdf

<sup>32</sup> Amnesty International Submission to the European Commission's Consultation on Artificial Intelligence.

<sup>&</sup>lt;sup>33</sup> Principal 11 of the UN Basic Principles and Guidelines on the Right to a Remedy and Reparation for Victims of Gross Violations of International Human Rights Law and Serious Violations of International Humanitarian Law.