

AMNESTY INTERNATIONAL SUBMISSION

INDEX NUMBER: TIGO IOR 10/2024.5404

11 April 2024

RECOMMENDATIONS TO THE PARLIAMENTARY ASSEMBLY AND COMMITTEE OF MINISTERS OF THE COUNCIL OF EUROPE ON THE DRAFT FRAMEWORK CONVENTION ON ARTIFICIAL INTELLIGENCE, HUMAN RIGHTS, DEMOCRACY AND THE RULE OF LAW

Amnesty International welcomes the initiative of the Council of Europe to draft a Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law as the first binding human rights instrument on Artificial Intelligence (AI).

In the context of a fast-moving technological landscape, where AI systems can exacerbate existing human rights violations at scale and introduce new ones, and in the absence of truly rights-respecting AI governance mechanisms, it is vital that the world's first binding rights-based framework on AI fully upholds international human rights law to set a high standard for regulating the development and deployment of AI technologies globally. Amnesty International emphasizes the need to ensure an ambitious and effective enforcement framework that prioritizes the protection and promotion of human rights above all.

In this regard, the text of the current draft Convention needs several crucial improvements detailed in this submission, and which should be considered by the Parliamentary Assembly of the Council of Europe during its Spring Plenary Session and subsequently by the Committee of Ministers, and introduced in a revised text in order to include the necessary human rights safeguards.¹

DRAW CLEAR RED LINES ON AI THAT IS INCOMPATIBLE WITH HUMAN RIGHTS

As the world's first binding rights-based instrument on AI, the Framework Convention must explicitly ban some of the most harmful AI-based practices, including systems used for public facial recognition, social scoring, predictive policing, biometric categorization, emotion recognition, risk assessment and profiling tools that violate rights of migrants, refugees and asylum seekers. Currently, as part of its "risk and impact management framework" the draft Convention leaves it up to each State Party to assess the need for a "moratorium or ban or other appropriate measures in respect of certain uses of artificial intelligence systems where it considers such uses are incompatible with the respect of human rights" (Article 16, paragraph 4), which is an inadequate approach to addressing the impact of AI technologies that are fundamentally incompatible with human rights.

Civil society organizations, including Amnesty International, have been calling for prohibitions on the above-mentioned AI practices based on extensive research,² including a ban on the development, production, sale, use, and export of remote biometric surveillance technologies by all public and private actors which lead to mass and discriminatory surveillance.³

¹ In addition to the specific recommendations on the draft text of the Convention, Annex I provides a non-exhaustive list of specific policy recommendations that should be reflected in any AI regulation as a bare minimum. Annex II provides an overview of Amnesty International's relevant research and advocacy outputs on AI and human rights.

² 'An EU Artificial Intelligence Act for Fundamental Rights', A Civil Society Statement, 30 November, 2021, <https://www.amnesty.eu/news/an-eu-artificial-intelligence-act-for-fundamental-rights/>

³ 'Ban dangerous facial recognition technology that amplifies racist policing', Amnesty International, 26 January, 2021, <https://www.amnesty.org/en/latest/press-release/2021/01/ban-dangerous-facial-recognition-technology-that-amplifies-racist-policing/>

Many of the concerns of civil society have been reflected in the 2021 annual report of the UN High Commissioner for Human Rights on *The Right to Privacy in the Digital Age*, who has called to 1) expressly ban AI applications that cannot be operated in compliance with international human rights law and impose moratoriums on the sale and use of AI systems that carry a high risk for the enjoyment of human rights, unless and until adequate safeguards to protect human rights are in place; and 2) Impose a moratorium on the use of remote biometric recognition technologies in public spaces, at least until the authorities responsible can demonstrate compliance with privacy and data protection standards and the absence of significant accuracy issues and discriminatory impacts, and until all the recommendations set out in A/HRC/44/24,⁴ paragraph 53 (j) (i–v), are implemented.⁵ In a 2023 open letter from the United Nations High Commissioner for Human Rights to European Union institutions on the European Union Artificial Intelligence Act (“AI Act”), OHCHR expressed support for a ban on the use of biometric recognition tools and other systems that process the biometric data of people to categorize them based on the colour of their skin, gender, or other protected characteristics, as well as bans on AI systems that seek to infer people’s emotions, individualized crime prediction tools, and untargeted scraping tools to build or expand facial recognition databases.⁶

REJECT BLANKET EXEMPTIONS FOR NATIONAL SECURITY, DEFENCE, AND RESEARCH AND DEVELOPMENT

The current draft of the Convention leaves the door open to rights violations in critical areas of national security and military application of AI technologies. By excluding research and development from its scope, the Convention also risks allowing the development of AI technologies which are discriminatory and rights-violating by design.

NATIONAL SECURITY

Amnesty International calls on the Council of Europe Parliamentary Assembly and Committee of Ministers to reject a blanket exemption on national security (Article 3, paragraph 2) in order to allow for a case-by-case assessment by States of appropriate exceptions for national security reasons meeting the three-part test of legality, necessity/proportionality and legitimate aim.

Amnesty International’s research has documented the disturbing “Orwellian” trend sweeping across Europe⁷ and beyond,⁸ where states’ growing unchecked powers to tackle terrorism are trampling freedoms, undermining the principle of legality, as well as the right to privacy, freedom of expression, right to liberty, freedom of movement, right to a nationality, and violating the principle of non-refoulement.

In this context, a blanket exemption introduces a significant loophole that would effectively undermine the objective of the Convention to safeguard human rights. For this reason, civil society actors, including Amnesty International have strongly opposed a similar exemption in the EU’s Artificial Intelligence Regulation (AI Act).⁹ The UN Special Rapporteur on Human Rights and Counter Terrorism has also cautioned “strongly against a blank national security exemption in the AI act” and encouraged the European Union “to ensure that exemptions are proportionate and consistent with existing European Union law, including the Charter of Fundamental Rights.”¹⁰ This approach has been supported by established case law of the Court of Justice of the European Union (CJEU) noting that “although it is for the Member States to define their essential

⁴ ‘Impact of new technologies on the promotion and protection of human rights in the context of assemblies, including peaceful protests’, Annual Report, United Nations High Commissioner for Human Rights, 24 June, 2020, <https://www.undocs.org/Home/Mobile?FinalSymbol=A%2FHRC%2F44%2F24&Language=E&DeviceType=Desktop&LangRequested=False>

⁵ ‘*The right to privacy in the digital age*’, Annual report, United Nations High Commissioner for Human Rights, 13 September 2021, <https://undocs.org/Home/Mobile?FinalSymbol=A%2FHRC%2F48%2F31&Language=E&DeviceType=Desktop&LangRequested=False>

⁶ Open Letter from the United Nations High Commissioner for Human Rights to European Union institutions on the European Union Artificial Intelligence Act (“AI Act”), 08 November 2023, <https://www.ohchr.org/en/open-letters/2023/11/turk-open-letter-european-union-highlights-issues-ai-act>

⁷ ‘*Dangerously disproportionate: The ever-expanding national security state in Europe*’, Amnesty International, 17 January 2017, <https://www.amnesty.org/en/documents/eur01/5342/2017/en/>

⁸ ‘*Amnesty International Report 2022/23: The state of the world’s human rights*’, Amnesty International, 27 March, 2023, <https://www.amnesty.org/en/documents/pol10/5670/2023/en/>

⁹ ‘*Ensure fundamental rights protections in the Council position on the AI Act*’, Civil Society letter to Czech Presidency of the EU, 17 October 2022, <https://www.amnesty.eu/news/ensure-fundamental-rights-protections-in-the-council-position-on-the-ai-act/>

¹⁰ ‘Human rights implications of the development, use and transfer of new technologies in the context of counter-terrorism and countering and preventing violent extremism’, Report to the 52nd session of the UN Human Rights Council, Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism, Fionnuala Ní Aoláin, 1 March 2023 https://www.ohchr.org/sites/default/files/documents/hrbodies/hrcouncil/sessions-regular/session52/advance-version/A_HRC_52_39_AdvanceEditedVersion.docx

security interests and to adopt appropriate measures to ensure their internal and external security, the mere fact that a national measure has been taken for the purpose of protecting national security cannot render EU law inapplicable and exempt the Member States from their obligation to comply with that law.”¹¹ The draft Convention should follow this approach in order to avoid supporting developments where States exploit national security to clamp down on essential rights and freedoms enshrined in international law.

NATIONAL DEFENCE

Similarly, **the draft Convention must reject a blanket exemption on technologies used and deployed for national defence purposes (Article 3, paragraph 4)**. As the UN Special Rapporteur on Human Rights and Counter Terrorism has also warned, to exclude the design, development and use of AI systems for national defence from the scope of the Convention “would effectively make the proposed convention irrelevant to the human rights concerns that are of greatest relevance in the region.”¹² Therefore, the Convention should establish:

- a prohibition on the development, production, use of, and trade in autonomous weapons systems (AWS) which by their nature cannot be used with meaningful human control over the use of force;
- a prohibition on “anti-personnel AWS” – that is, systems that use sensors to profile, target and apply force to humans, whether or not there is meaningful human control over the use of force;
- regulation of the use of all other autonomous weapons systems;
- and a positive obligation to maintain meaningful human control over the use of force.

RESEARCH AND DEVELOPMENT

The Convention must clearly apply to research and development to ensure that human rights are prioritised at the stage of conceptualisation, design, development, and testing of AI technologies. Not doing so risks allowing the creation of non-compliant technologies by default, and providing a loophole for deploying rights-violating AI systems under the guise of testing. This is especially vital due to the absence of adequate regulatory safeguards across the world, which would prevent the proliferation of rights-violating technologies within and across state borders.¹³

While the draft Convention does consider “testing and similar activities” with the “potential to interfere with human rights” to fall under its scope (Article 3, paragraph 3), lack of clarity on how that potential is assessed risks misinterpretation and leading to human rights violations by developers and deployers of AI technologies. Civil society research attests to the need to ensure clear enforceable obligations regarding research and development.¹⁴ This is supported by the 2020 Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance which examined the racially discriminatory impacts of new and emerging technologies on migrants, refugees and other non-citizens. The report demonstrates how “governments and non-state actors are developing and deploying emerging digital technologies in ways that are uniquely experimental, dangerous, and discriminatory in the border and immigration enforcement context. By so doing, they are subjecting refugees, migrants, stateless persons and others to human rights violations, and extracting large quantities of data from them on exploitative terms that strip these groups of fundamental human agency and dignity.”¹⁵

¹¹ ‘Judgment of the Court (Grand Chamber) in Case C-623/17,’ The Court of Justice of the European Union, 6 October 2020 <https://curia.europa.eu/juris/document/document.jsf?sessionId=DC1C6EE6335FCFE02B5D540D7C610EA2?text=&docid=232083&pageIndex=0&doclang=en&mode=Ist&dir=&occ=first&part=1&cid=2401090>.

¹² ‘Human rights implications of the development, use and transfer of new technologies in the context of counter-terrorism and countering and preventing violent extremism,’ Report to the 52nd session of the UN Human Rights Council, Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism, Fionnuala Ní Aoláin, 1 March 2023 https://www.ohchr.org/sites/default/files/documents/hrbodies/hrcouncil/sessions-regular/session52/advance-version/A_HRC_52_39_AdvanceEditedVersion.docx

¹³ For example, while the EU’s AI Act sets obligations for high-risk AI systems and prohibits the use of systems considered to pose unacceptable risk to human rights within the EU, it does not apply to EU-based companies when they develop systems for export purposes. See ‘Lawmakers reluctant to stop EU companies profiting from surveillance and abuse through the AI Act,’ Amnesty International, 5 December 2023, <https://www.amnesty.eu/news/lawmakers-reluctant-to-stop-eu-companies-profiting-from-surveillance-and-abuse-through-the-ai-act/>

¹⁴ ‘Technological Testing Grounds: Migration Management Experiments and Reflections from the Ground Up,’ Petra Molnar, EDRI, and the Refugee Law Lab, November 2020, <https://edri.org/wp-content/uploads/2020/11/Technological-Testing-Grounds.pdf>

¹⁵ ‘Contemporary forms of racism, racial discrimination, xenophobia and related intolerance,’ Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance, E. Tendayi Achiume, 10 November 2020, <https://documents.un.org/doc/undoc/gen/n20/304/54/pdf/n2030454.pdf?token=uidiMrNDUKsnAQJiYh&fe=true>

INCLUDE IN SCOPE ALL PRIVATE AND PUBLIC ACTORS

PRIVATE ACTORS

The Framework Convention must directly apply to private actors. A major shortcoming of the draft Convention (Article 3, paragraph 1(b)) is that it currently leaves it to each State Party to determine whether to apply the Convention in relation to private actors or implement “other appropriate measures.” This risks jeopardizing the very aim of the Convention to protect and promote human rights during the development and deployment of AI technologies, given that private actors are key actors throughout the whole AI lifecycle and in fact leading the majority of AI development in the world. Disproportionate discretions to industry actors have already been criticized by civil society actors in other AI regulation efforts, such as the EU’s AI Act.¹⁶ Despite the acknowledgement in the explanatory report of several risks that are particularly relevant in relation to AI use by private actors,¹⁷ the draft Convention creates an even bigger risk of private sector exclusion, which must be corrected before adopting the final text of the treaty.

PUBLIC ACTORS

The Framework Convention must also apply to public actors at all levels of governance. The current draft text, referring to federal forms of government, notes the direct application of the Convention to central government, while obliging it to only inform the competent authorities of constituent states or other similar territorial entities “of the said provisions with its favourable opinion, encouraging them to take appropriate action to give them effect” (Article 33, paragraph 2). This loophole would mean that significant portions of states’ human rights obligations would be omitted from the Convention. As regards the International Covenant on Civil and Political Rights (ICCPR), for example – to which all Council of Europe member states are parties, the UN Human Rights Committee has noted that “The obligations of the Covenant in general and article 2 in particular are binding on every State Party as a whole. All branches of government (executive, legislative and judicial), and other public or governmental authorities, at whatever level - national, regional or local - are in a position to engage the responsibility of the State Party.”¹⁸ If not rectified, it will allow extensive areas of AI deployment to go unregulated, despite the impact of AI being greatly localised and context specific,¹⁹ therefore leaving the risk of human rights violations facilitated by public use of AI without public accountability and recourse for impacted people and communities.

ENSURE A COMPREHENSIVE RIGHTS-BASED FRAMEWORK

As the first ever human rights-based treaty on artificial intelligence, the Convention must designate comprehensive obligations to protect and promote rights rather than set “principles,” as Chapter III currently does (Articles 6 to 13). In addition to human dignity, individual autonomy, equality and non-discrimination, privacy and data protection already in the draft text, this chapter could benefit greatly from an expanded acknowledgement of impact of AI technologies on rights established in international law including the right to freedom of expression, freedom of peaceful assembly, of association, right to asylum, freedom of movement, the right to effective remedy, fair trial, right to liberty and security, right to a healthy environment, right to accessibility, right to social protection and other applicable rights. It should also address the relevant extra-territorial impact of AI technologies on rights beyond the geographic borders of State Parties to the Convention. These particularly should include prohibitions on exports of technologies considered incompatible with human rights by State Parties, and protections against labour exploitation and from environmental harm during the development of AI technologies. An effective due diligence process referred to in the next paragraphs is crucial in this regard.

¹⁶ ‘EU legislators must close dangerous loophole in AI Act,’ A Civil Society Statement, 7 September 2023, <https://www.amnesty.eu/news/eu-legislators-must-close-dangerous-loophole-in-ai-act/>

¹⁷ For instance, in para. 43-44 notes concern over artificial intelligence applications posing numerous threats to human rights and undermining civic participation. It states that AI tools could enable dissemination of disinformation and misinformation that could undermine the right of access to information, lead to prejudiced decisions about individuals, potentially resulting in discriminatory practices; influence court rulings, with potential implications for the integrity of the justice system; and undertake illegal or arbitrary surveillance, leading to restrictions on the freedom of assembly or freedom of expression, and privacy.”

¹⁸ ‘The nature of the general legal obligation imposed on States Parties to the Covenant : International Covenant on Civil and Political Rights,’ General comment no. 31 (80), adopted on 80th session of the Human Rights Committee (2187th meeting), 29 March 2004, <https://digitalibrary.un.org/record/533996?ln=en&v=pdf>

¹⁹ See for example the use of facial recognition by the New York Police Department (NYPD) to identify and locate Black Lives Matter activist Derrick “Dwreck” Ingram. ‘Ban the Scan, New York City’, Amnesty International, <https://banthescan.amnesty.org/nyc/>

ENSURE EFFECTIVE TRANSPARENCY AND ACCOUNTABILITY FOR AI DEVELOPERS AND DEPLOYERS

The Convention must establish an effective transparency and public accountability framework, including a designated public database for disclosing development and deployment of AI systems impacting human rights. The current proposed accountability mechanism (Article 14) is insufficient and lacks the necessary specificity for effectiveness. State Parties are required to ensure documentation of “relevant information regarding artificial intelligence systems which have the potential to significantly affect human rights,” provision of that information to bodies authorized with access, and “where appropriate and applicable” also to affected persons. (Article 14, paragraph 2(a)).

It is unclear what constitutes or who defines “significantly affecting human rights”, as well as what would be the content of the documentation. The draft Convention also leaves unclear whether this obligation falls on the developer, the deployer, or a mandated authority for this purpose, leaving room for differential interpretation among State Parties, undermining the norm-setting purpose and overall effectiveness of the Convention. The lack of requirement for ensuring public access to the relevant information undermines the accountability mechanism proposed by the draft Convention. To address the above concerns, Amnesty International recommends that the Convention mandates clearly the:

- Creation and maintenance of a publicly available and accessible databases by State Parties for reporting on development and deployment of AI technologies;
- Obligation on providers of AI systems to register themselves and the given AI system in the relevant public database, including during testing of AI systems;
- Obligation on all public and private deployers to register the use of AI systems in the relevant public database;
- Obligation on deployers to ensure meaningful transparency of AI systems and algorithmic decision-making to impacted people, including during testing of AI systems.

ENSURE EFFECTIVE HUMAN RIGHTS DUE DILIGENCE THROUGHOUT THE AI LIFECYCLE

The Convention must establish an effective and binding human rights due diligence requirement for developers and deployers of AI systems, including a detailed human rights impact assessment framework. Currently, the draft Convention fails to do this and instead designates State Parties to “adopt or maintain measures for the identification, assessment, prevention and mitigation of risks posed by artificial intelligence systems” (Article 16, paragraph 1) with a number of non-binding “requirements” (e.g., to “consider, where appropriate, the perspectives of relevant stakeholders in particular persons whose rights may be impacted” (Article 16, paragraph 2)). It also does not require State Parties to ensure that adverse impact of AI systems on human rights are addressed but rather mandates them to adopt and maintain measures that “seek to ensure” this (Article 16, paragraph 3). The Convention must set clear obligations for:

- Companies developing AI products to conduct adequate human rights due diligence to identify and address human rights harms that might appear at any stage of the supply chain or product lifecycle as outlined in the United Nations Guiding Principles on Business and Human Rights;
- Oblige deployers of AI technologies to conduct and publish human rights impact assessments prior to each deployment, including assessment of appropriateness of an AI-based solution in a specific scenario. As civil society has formerly recommended, such an assessment must at the very least detail specific information to the context of use of that system, including the intended purpose, geographic and temporal scope, assessment of the legality and human rights impacts of the system, compatibility with accessibility legislation, likely direct and indirect impact on human rights, any specific risk of harm likely to impact marginalized persons or those at risk of discrimination, the foreseeable impact of the use of the system on the environment, any other negative impact on the public interest; and clear steps as to how the harms identified will be mitigated, and how effective this mitigation is likely to be.²⁰

ENSURE EFFECTIVE REDRESS AND REMEDY FOR IMPACTED PEOPLE AND COMMUNITIES

The Convention must go further and expand the rights to redress and remedy by impacted people and communities. While it is commendable that the draft Convention sets ground for rights and redress

²⁰ ‘Introduce obligations on users of high-risk AI systems,’ Civil Society Amendments, 3 May 2022, <https://edri.org/wp-content/uploads/2022/05/Obligations-on-users-AIA-Amendments-17022022.pdf>

mechanism for impacted people, these must be sufficiently strengthened to be effective. For example, conditional language (i.e., “where appropriate and applicable”) on providing the documented information on AI systems “which have the potential to significantly affect human rights” to affected persons undermines the accountability mechanism proposed by the draft Convention (Article 14, paragraph 2(a)), and puts at risk the right of individuals to effective remedy. Additionally, although a positive first step, the right to lodge a complaint with the competent authority (Article 14, paragraph 2 (c)) is not sufficient and must be complemented by other rights. In particular, the Convention must:

- Ensure the right to an effective remedy against state authorities and against a deployer for the infringement of rights;
- Ensure the right to information and explanation of AI-supported decision-making for impacted people, including about the use and functioning of AI in the system;
- Ensure impacted people have access to judicial and non-judicial pathways to remedy for violation of their rights by AI systems;
- Ensure public interest organisations the right to support impacted people seeking remedy, as well as to lodge cases on their own initiative.

ESTABLISH CLEAR OBLIGATIONS TO SUPPORT MEANINGFUL INVOLVEMENT OF IMPACTED COMMUNITIES, CIVIL SOCIETY ORGANIZATIONS AND HUMAN RIGHTS EXPERTS

The Convention must oblige State Parties to empower impacted communities, civil society organizations and human right experts so that they can meaningfully engage in the development and deployment of AI technologies, the implementation, monitoring and evaluation of the current Convention, and national or other policies stemming from its implementation. Currently, the draft text lacks clarity and strength to ensure this with few weakly formulated references to stakeholder engagement (Article 16, paragraph 2(c), Article 19, Article 23, paragraph 2(f), Article 25, paragraphs 2 and 3). For this purpose, the Convention must mandate State Parties to:

- Ensure binding rules for a clear, easy to access, transparent, and accountable implementation process that enables meaningful and equal participation of a wide range of rightsholders at national level;
- Ensure meaningful participation of impacted communities, by centring policy discussions around needs and priorities of those communities, enabling equal participation of representative advocates and organisations through resource-allocation, and creating level-field between all stakeholders and rightsholders, and valuing experiential expertise;
- Create equal conditions for engagement of community organisations and other civil society actors, by keeping in check lobbying power of industry players and state actors;
- Establish the above obligations in Conference of Parties’ rules of procedure.

ESTABLISH CLEAR OBLIGATIONS TO ENSURE CONSISTENT AND EFFECTIVE APPLICATION OF THE CONVENTION

The text of the Convention must avoid conditional and non-binding framings and set clear enforceable obligations on State Parties to ensure its effective implementation. While almost all articles start with “Parties shall”, and the explanatory text claims it is creating legally binding obligations, what follows are very vague formulations, without any criteria, and more often than not even weakened by conditions such as “as appropriate”, “differentiate”, “where possible”, and “seek to ensure”. The draft Convention does not provide any criteria or obligatory steps and processes for any of the “obligations” it establishes. In fact, the deference to domestic law allows and even encourages states to continue to rely on existing standards in their national law, which undermines the very aim of the Convention to set a strong level of protections through common standards applicable across State Parties. Amnesty International therefore recommends amending the language of the draft to ensure clear and binding obligations applicable to all State Parties, setting a high bar for the protection and promotion of human rights staying true to the spirit of the Convention.

ANNEX I – AMNESTY INTERNATIONAL RECOMMENDATIONS ON AI REGULATION

When developing local, national, regional or global governance frameworks for the development and deployment of AI, policymakers should at the very least ensure the following:

Draw clear red lines on the development and deployment of AI that is incompatible with human rights

- Ban the development, production, sale, and use of biometric technologies by all public and private actors that enable mass surveillance and discriminatory targeted surveillance;
- Ban the development, production, sale, and use of AI systems that create or expand facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage;
- Ban automated risk assessment and profiling systems in the context of migration, when used to determine whether people on the move present a ‘risk’ of unlawful activity or security threats;
- Ban predictive analytic systems used to interdict, curtail and prevent migration;
- Ban the use of AI in predictive policing (prediction of crimes by individual person(s) and/or in given spaces and/or times);
- Ban the use of AI systems for social scoring;
- Ban the export of AI systems that are incompatible with human rights.

Reject loopholes and exemptions which risk violation of human rights

- Reject blanket or disproportionate exemptions based on policing, national security or military grounds;
- Reject granting discretion to companies developing AI systems the possibility of opting out of established rules;
- Ensure that technical and human rights safeguards apply to exported AI technologies.

Ensure human rights due diligence throughout the AI lifecycle

- Require companies developing AI products to conduct adequate human rights due diligence to identify and address human rights harms that might appear at any stage of the supply chain or product lifecycle as outlined in the United Nations Guiding Principles on Business and Human Rights;
- Oblige deployers of AI technologies to conduct and publish human rights impact assessments prior to each deployment, including assessment of appropriateness of an AI-based solution in a specific scenario.

Ensure public accountability and transparency measures when developing and deploying AI technologies

- Create and maintain publicly available and accessible databases for reporting development and deployment of AI technologies;
- Oblige providers of AI systems to register themselves and the given AI system in relevant public databases, including during testing of AI systems in real world conditions;
- Oblige all public and private deployers to register the use of AI systems in relevant public databases;
- Oblige deployers to ensure meaningful transparency of AI systems and algorithmic decision-making to impacted people, including during testing of AI systems in real world conditions.

Empower people and communities impacted by AI

- Ensure the right to an effective remedy against state authorities and against a deployer for the infringement of rights.
- Ensure the right to information and explanation of AI-supported decision-making for impacted people, including about the use and functioning of AI in the system;
- Ensure impacted people have access to judicial and non-judicial pathways to remedy for violation of their rights by AI systems;
- Ensure public interest organisations the right to support impacted people seeking remedy, as well as to lodge cases on their initiative.

Address specific challenges posed by Generative AI

- Require companies developing new AI models, platforms, products, services and tools to institute effective and on-going human rights due diligence processes,²¹ to both identify and mitigate risks as early as possible in the product development, release and update cycle as per UN Guiding Principles on Business and Human Rights²²;
- Require companies to proactively engage with academics, civil society actors, and community organisations especially those representing traditionally marginalised communities during the development of said models, platforms, products, services, and tools.²³

Address specific challenges posed by algorithmic recommender systems

- Require technology companies to ensure that content-shaping algorithms used by online platforms are not based on profiling by default and must require an opt-in instead of an opt-out, with consent for opting in being freely given, specific, informed and unambiguous. The collection and use of inferred data (for example, recommendations based on watch time and likes) to personalize ads and content recommendations should be banned.

²¹ Others have suggested to assess impact of AI on inclusiveness and sustainability as part of risk and impact assessments, to protect the wider interests of society. See for example, Louis Au Yeung, UC Berkeley Center for Long-Term Cybersecurity, July 2021, "Guidance for the Development of AI Risk and Impact Assessments", https://cltc.berkeley.edu/wp-content/uploads/2021/08/AI_Risk_Impact_Assessments.pdf; or Katell, M., Young, M., Dailey, D., Herman, B., Guetler, V., Tam, A., Bintz, C., Raz, D. and Krafft, P. M. (2020). "Toward situated interventions for algorithmic equity: lessons from the field," Conference on Fairness, Accountability, and Transparency pp.44-45 [online] ACM: Barcelona. Available at: <https://dl.acm.org/doi/abs/10.1145/3351095.3372874>

²² "Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework", United Nations, 2011, https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf

²³ Campbell, E., Kleinman, M., Al Jazeera, 13 June 2023, "AI must not become a driver of human rights abuses" <https://www.aljazeera.com/opinions/2023/6/13/ai-must-not-become-a-driver-of-human-rights-abuses>

ANNEX II – RELEVANT RESEARCH AND OTHER OUTPUTS

- Dangerously disproportionate: The ever-expanding national security state in Europe, Amnesty International, 17 January 2017, <https://www.amnesty.org/en/documents/eur01/5342/2017/en/>
- Trapped in the Matrix: Secrecy, stigma, and bias in the Met's Gangs Database, Amnesty International, May 2018, https://www.amnesty.org.uk/files/2018-05/Trapped%20in%20the%20Matrix%20Amnesty%20report.pdf?VersionId=IJSxllcKfkZgr4gHZsz0vW8JZ0W3V_PD;
- Out of Control: Failing EU Laws for Digital Surveillance Export, Amnesty International September 21, 2020, <https://www.amnesty.org/en/documents/EUR01/2556/2020/en/>;
- We sense trouble: Automated discrimination and mass surveillance in predictive policing in the Netherlands, Amnesty International, September 29, 2020, <https://www.amnesty.org/en/documents/eur35/2971/2020/en/>;
- Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal, Amnesty International, October 25, 2021, <https://www.amnesty.org/en/documents/eur35/4686/2021/en/>;
- Automated Apartheid: How facial recognition fragments, segregates and controls Palestinians in the OPT, Amnesty International, May 2, 2023, <https://www.amnesty.org/en/documents/mde15/6701/2023/en/>;
- Digitally Divided: Technology, Inequality, and Human Rights, Amnesty International, October 2, 2023, <https://www.amnestyusa.org/reports/digitally-divided-technology-inequality-and-human-rights/>;
- Driven into Darkness: How TikTok's 'For You' Feed Encourages Self-Harm and Suicidal Ideation, Amnesty International, November 7, 2023, <https://www.amnesty.org/en/documents/POL40/7350/2023/en/>;
- "I Feel Exposed": Caught in TikTok's Surveillance Web, Amnesty International, November 7, 2023, <https://www.amnesty.org/en/documents/POL40/7349/2023/en/>
- Trapped by Automation: Poverty and discrimination in Serbia's welfare state, Amnesty International, December 4, 2023, <https://www.amnesty.org/en/latest/research/2023/12/trapped-by-automation-poverty-and-discrimination-in-serbias-welfare-state/>;
- Ban the Scan Campaign, Amnesty International, <https://banthescan.amnesty.org/>;
- Primer: Defending the rights of refugees and migrants in the digital age, Amnesty International, February 5, 2024, [https://www.amnesty.org/en/documents/pol40/7654/2024/en/#:~:text=the%20European%20Union.-,Defending%20the%20rights%20of%20refugees%20and%20migrants%20in%20the%20digital.issues%20arising%20from%20their%20use](https://www.amnesty.org/en/documents/pol40/7654/2024/en/#:~:text=the%20European%20Union.-,Defending%20the%20rights%20of%20refugees%20and%20migrants%20in%20the%20digital.issues%20arising%20from%20their%20use;);
- Briefing: Social protection in the Digital Age, Amnesty International, March 6, 2024, <https://www.amnesty.org/en/documents/pol40/7771/2024/en/>
- An EU Artificial Intelligence Act for Fundamental Rights, A Civil Society Statement, November 30, 2021, <https://www.amnesty.eu/news/an-eu-artificial-intelligence-act-for-fundamental-rights/>
- The EU's Artificial Intelligence Act: Civil society amendments, May 3, 2022, <https://edri.org/our-work/the-eus-artificial-intelligence-act-civil-society-amendments/>