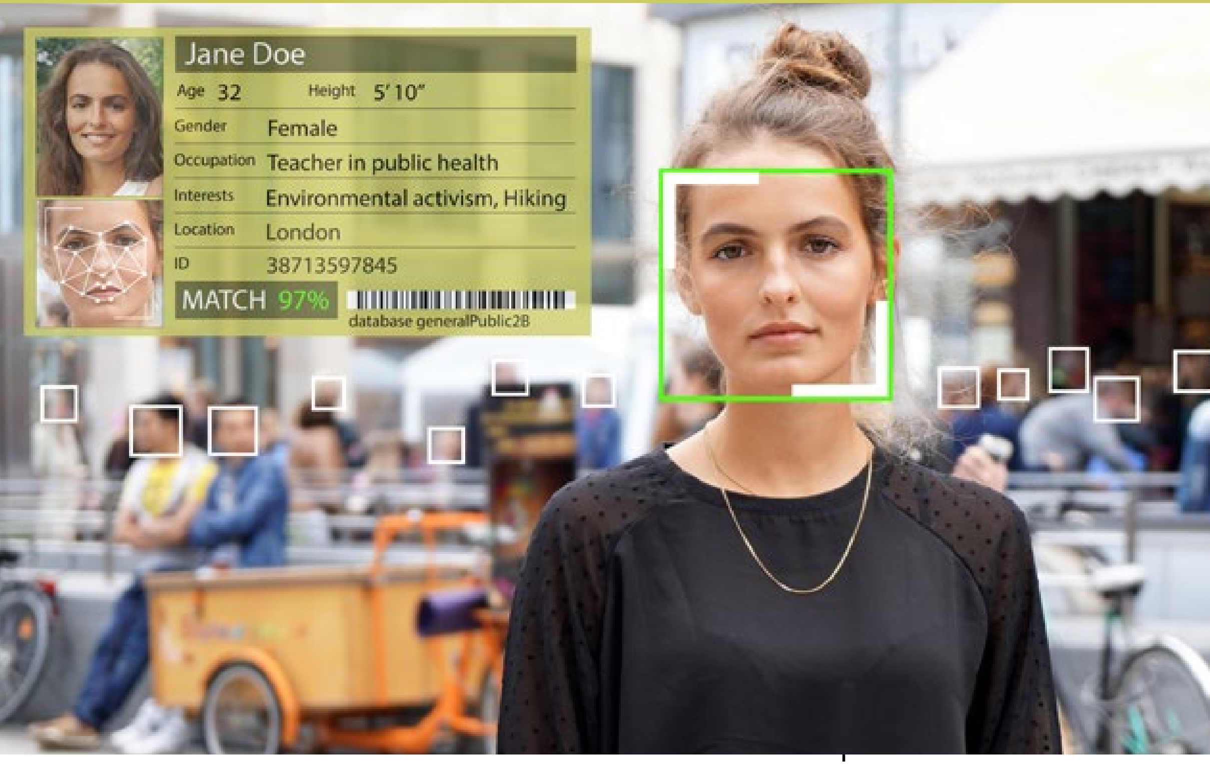
Human rights by design
future-proofing human rights protection in the era of AI



Follow-up Recommendation to "Unboxing AI" (2019)

COMMISSIONER FOR HUMAN RIGHTS
COMMISSAIRE AUX DROITS DE L'HOMME

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

Human rights by design

future-proofing human rights protection in the era of AI

Follow-up Recommendation
to "Unboxing AI" (2019)
by the Council of Europe
Commissioner for Human Rights

Council of Europe

All requests concerning the reproduction or translation of all or part of this document should be addressed to the Directorate of Communication (publishing@coe.int).

All other correspondence concerning this document should be addressed to the Office of the Commissioner for Human Rights.

Publications are available on the Commissioner's website:
www.commissioner.coe.int

Cover photo:
edited stock photo ©Shutterstock

Acknowledgements:

The Commissioner would like to express her gratitude to Nani Jansen Reventlow and Jonathan McCully for their research and assistance in preparing this Follow-up Recommendation.

Contents

INTRODUCTION 5

CHAPTER 1 - HUMAN RIGHTS IMPACT ASSESSMENTS 9

CHAPTER 2 - HUMAN RIGHTS STANDARDS IN THE PRIVATE SECTOR 13

CHAPTER 3 - INFORMATION AND TRANSPARENCY 17

CHAPTER 4 - PUBLIC CONSULTATIONS 19

CHAPTER 5 - PROMOTION OF AI LITERACY 21

CHAPTER 6 - INDEPENDENT OVERSIGHT 23

CHAPTER 7 - EFFECTIVE REMEDIES 27

CONCLUDING OBSERVATIONS AND RECOMMENDATIONS 31

Introduction

The pace of technological development is picking up speed and Artificial Intelligence (AI) is playing an ever-growing role in all aspects of our lives, including public administration. Member states of the Council of Europe must keep up with the increasing reliance on automated processes and machine-learning and ensure that they safeguard the human rights of everyone in society in this fast-evolving context.

This pressing need prompted the Council of Europe Commissioner for Human Rights, Dunja Mijatović, (Commissioner) in 2019 to publish the Recommendation "Unboxing AI: 10 steps to protect Human Rights" (2019 Recommendation), which provides guidance to member states on the main principles that should be followed to prevent or mitigate the negative impacts of AI systems on human rights. The 2019 Recommendation stresses the special risks stemming from AI to the rights to non-discrimination and equality, data protection and privacy, as well as freedom of expression, freedom of assembly and the right to work. In addition, it highlights seven key areas that require particular focus: the need to conduct human rights impact assessments (HRIAs) before an AI system is acquired, developed and/or deployed; the observance of human rights standards in the private sector; information and transparency; meaningful public consultations; the promotion of AI literacy; independent oversight; and effective remedies.

Member states have acted on some of the key areas identified in the 2019 Recommendation, but the overall approach has not been consistent. Human rights-centred regulation of AI systems is still absent on many fronts and public authorities tend to get involved too late – and move forward at insufficient speed – for their engagement to be truly meaningful. While human rights norms and safeguards are technology-neutral and applicable to all contexts, including those involving AI systems, their enforcement is often lacking and oversight sporadic.

This Follow-up Recommendation reviews the challenges faced by member states in implementing the 2019 Recommendation, such as the adequacy of assessment of human rights risks and impacts, the establishment of stronger transparency guarantees, and the requirement of independent oversight. To what extent have member states been able to use AI as a tool for boosting human rights rather than harming them, and what broader trends are influencing the practices of member states? What role have national human rights structures (NHRSs), including national human rights institutions (NHRIs), equality bodies and ombudsman institutions, played and how can their role be strengthened to ensure that the multiple human rights harms stemming from AI are effectively addressed?

This Recommendation considers the negative impact that AI systems may have on the ability of people to enjoy a human right as a potential human rights harm. In understanding the potential harms of new technologies such as AI, it is key to consider the broader context, existing inequalities, and power asymmetries into which they are being deployed. Machine learning technologies learn from the patterns and assumptions that prevail in the data they use. Therefore, they further entrench and exacerbate already-existing and systemic biases and prejudice, for instance against women, young people, persons with disabilities, or persons with a minority background. Given the wide range of sectors in which AI applications are used, including the distribution of social welfare benefits, decisions on the creditworthiness of potential clients, staff recruitment and retention processes, criminal justice procedures, immigration and border control, policing and targeted advertising and newsfeeds, negative impacts translate not only into individual and possibly collective human rights violations. They also adversely impact social justice, alter the relationship and trust between citizens and government, and affect the integrity or even outcome of elections. Finally, the development of AI systems raises important questions about the working conditions of often informally employed digital platform workers, as well as about the exorbitant amounts of energy and water that it requires. Only by addressing the threats posed by AI to human rights in a holistic way, and by considering the nature of harms as multi-faceted, intersectional, and dynamic, can member states properly meet their international human rights obligations.

Since the Commissioner's 2019 Recommendation, NHRSs and their networks have taken several commendable initiatives to increase their capacity to tackle human rights issues arising from the design, development, and deployment of AI. In 2020, EQUINET, the network of Equality Bodies, published an extensive report on the impact of AI on equality and the role to be played by Equality Bodies in this regard. In March 2020, the International Ombudsman Institute convened a workshop dedicated to examining the challenges, roles and tools of Ombudsman institutions dealing with AI and human rights. In order to strengthen the capacity of their members to deal with this new area, the European Network of National Human Rights Institutions (ENNHRI) focused its 2022 NHRIs Academy on AI and human rights and, in October 2022, created an AI Working Group of NHRIs to be a platform for peer exchange and learning, while EQUINET also convened several trainings. ENNHRI and its members have also sought to inject human rights considerations into relevant legislative processes at regional level and presented submissions on the EU Commission's White Paper on Artificial Intelligence and the Council of Europe's draft Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law. Both networks, ENNHRI and EQUINET, have observer status at the Council of Europe Committee on AI (CAI) and have been attending meetings regularly, providing input on the draft Framework Convention. The Commissioner strongly welcomes these steps and encourages further efforts towards the increased involvement of NHRSs in defending and promoting human rights in the field of AI governance and regulation.

An exchange of views between the Commissioner and 30 heads and senior representatives of NHRSs of Council of Europe member states was organised in March 2023 as part of the preparation of this Follow-up Recommendation. Some of the points made by participants during that exchange are cited in Chapters I – VII, which provide an overview of some of the steps taken to implement the 2019 Recommendation. The Conclusions identify current shortcomings and broader trends, leading to recommendations on how to move towards the more effective enforcement of human rights standards in relation to AI design, development, and deployment.

Chapter 1
Human Rights Impact Assessments

Owing to the fact that the potential negative impact of AI systems on human rights can be significant, posing a threat to human life, the environment, democracy, and the rule of law, member states should conduct HRIAs before any AI system is designed, developed or deployed. For that process to be effective, it is essential to consider that human rights harms do not only stem from the use of AI systems but can be inflicted at any point during their lifecycle, including the data acquisition and modelling phases. In addition, it is important to pay attention to the dynamic conditions of the real-world environments in which the systems will be operating and interacting with other systems that are simultaneously at play.

Since 2019, risk and impact assessments have featured prominently in policy discussions around AI governance and regulation in Europe. However, many of these initiatives have failed to fully implement HRIAs in relation to all AI systems that are procured, developed and/or deployed by public authorities, applying instead a sectoral approach that limits the duty to undertake HRIAs to specific branches in industry or to specific rights only. This should be avoided, as it can lead to regulatory gaps and piecemeal implementation of HRIAs. While pilot projects, for instance in the health sector, are welcome, they should be extended to other contexts where AI systems are being used, including policing, migration and social welfare, and all should evaluate possible impacts on the comprehensive range of human rights.

Indeed, member states have usually either failed to recognise the full spectrum of human rights that are threatened or have discussed vague ethical frameworks rather than referring to specific and established human rights standards. Furthermore, most assessment frameworks are designed as voluntary exercises of the private sector, in the context of its general due diligence responsibilities, rather than as binding obligations of member states with a view to preventing AI-related human rights violations, including when induced by third parties.

HRIAs must have a broad enough scope to cover risks to all human rights and should not narrowly define the risks that are being assessed, as this is inconsistent with the principle that all rights are indivisible and interdependent. Content moderation algorithms, for instance, do not only affect freedom of expression. Erroneous takedowns can also feed into algorithms that are employed to screen potential job candidates and can affect the right to associate and disassociate oneself as it can lead to the temporary or permanent blocking of access. Underlining the need for a comprehensive approach to AI-related risk and impact assessments, the CAI has been preparing a methodology containing essential parameters of a risk and impact management process for AI systems from the point of view of human rights, democracy and the rule of law. The impacts on specifically protected groups, including children who make up one in three users of digital services world-wide, must also be considered and effectively mitigated.

NHRSs, given their specific knowledge of the human rights situation in the member states and the most affected risk groups, can provide important guidance on the types of human rights risks that may be encountered, both in terms of individual human rights violations and broader threats to specific groups, and on possible mitigating measures that may be applied. In this context, it is important that NHRSs themselves be familiar with the types of human rights harms that may arise from the use of AI in their specific national circumstances. In 2019, the Federal Anti-Discrimination Agency in Germany, for instance, funded the study "Discrimination risks concerning the use of algorithms", with the aim of bringing such risks to the attention of the authorities. In 2022, the Swedish Gender Equality Authority completed a study on how government agencies are using AI and to what extent they take discrimination risks into account. The study concluded that, although there was some awareness that AI systems affected large numbers of individuals and that discrimination risks existed, public authorities did not systematically take the discrimination perspective into account in risk analyses, nor in the implementation process, but rather considered ethics and questions of integrity.

Initiatives taken to assess the risks and impacts of AI systems often refer to ethics and ethical principles rather than established human rights norms, resulting in a de-centring of human rights protections and a watering down of legal accountability with respect to the design, development, and deployment of AI systems. International human rights law has been characterised as crystallisation of ethical principles into norms, and has emerged from decades of careful deliberations and weighing of human rights considerations and other interests, including ethics. While it may not ensure the protection of all ethical values of societal concern that are implicated by AI systems, it imposes a clear list of legal obligations on member states of the Council of Europe to which they are bound to adhere, including with respect to the design, development, and deployment of AI.

> *"Many algorithmic harms are hidden. If we try to lower the risk for gender discrimination, it may become higher for disability or ethnicity. These inherent features are still not sufficiently understood. We need more training to ask the right questions and understand the trade-offs within the systems."*

Established human rights standards further offer robust methods and processes to assess individual rights against competing rights and interests, embedding tests of legality, necessity and proportionality that are widely known and accepted, and can thus be applied across the range of organisations and actors involved in developing, building, and operating AI systems. International human rights standards should thus be the key reference in the impact and risk assessments carried out by public authorities in relation to AI systems, providing guidance to the authorities involved when considering the risks posed by the AI system to human rights and evaluating whether these risks (once sufficiently mitigated) are still necessary and proportionate for achieving a legitimate aim or interest. Only then can the risks posed by such AI systems be deemed to be legally justifiable.

In 2020, the Danish Institute for Human Rights created guidance and a toolbox on how to conduct HRIAs of digital activities, which has been used by a growing number of businesses and some public sector actors to assess the effects of their digital activities on rights-holders, including workers, local community members, consumers and others.

In 2022, the Netherlands became the first member state to introduce mandatory HRIAs for public institutions before using algorithms to make evaluations or decisions about people. Publication of the results of HRIAs has also been made obligatory, which is aligned with the Commissioner's 2019 Recommendation. The legal framework provides indicators for identifying whether an algorithm infringes on a fundamental right or freedom in the preparation, development or deployment phase, and whether there is an adequate legal basis under human rights law. It also sets out options for preventative and mitigating measures that can be adopted in response to the identified risks which, as noted above, supports the uptake of HRIAs over other, narrower assessment frameworks.

To be effective and meaningful, legal frameworks for mandatory and publicly accessible HRIAs in relation to AI systems must be strictly implemented and enforced in practice, regardless of the size or market share of the AI developer. Member states should endow public authorities with adequate resources, expertise, and access to information. Public authorities should not procure or use AI systems from third parties in circumstances where they cannot carry out an effective HRIA, including because the commercial AI developer refuses to share relevant information about the system, and member states must ensure that all barriers to the effective functioning of HRIAs are overcome in their legal frameworks.

Chapter 2
Human Rights Standards in the Private Sector

Given the rapid speed of AI development by the private sector, member states, as principal duty-bearers with respect to their human rights obligations, need to act swiftly to ensure that their laws stay apace with the threat to human rights posed by these advancements. Yet, since the 2019 Recommendation, member states have been slow to adopt legal frameworks that address, prevent and remedy human rights abuses by AI actors in the private sector. This might partly be due to the fact that efforts at Council of Europe and EU level to elaborate appropriate legal frameworks on AI design, development, and deployment are still ongoing, and member states intend to await their outcome. Negotiations at regional level, however, do not diminish the standing obligations of member states to protect individuals and systems against human rights abuses by third parties, including AI actors.

The assessment of the human rights impacts of business activities, including in the digital sphere, is considered a key component of corporate due diligence, as outlined in the UN Guiding Principles on Business and Human Rights. Much remains to be done in many member states to effectively implement the Guiding Principles, as well as the complementary Recommendation CM/Rec(2016)3 of the Committee of Ministers to member states on human rights and businesses. As has been recognised by the UN Working Group on the issue of human rights and transnational corporations and other business enterprises, business-related human rights abuses across the globe have often remained unaddressed and unregulated at national level. Information from member states on their implementation of these guidelines and recommendations is also often limited. These issues must be tackled as a matter of urgency.

Member states continue to rely almost exclusively on data protection frameworks to prevent, address, and remedy human rights abuses by AI actors in the private sector. While data protection laws are important means for holding businesses to account for violations of the right to privacy, enforceable through the imposition of fines, they will be insufficient when it comes to the violations of other human rights. They are also not suited to situations where AI systems do not involve the processing of personal data, but still pose risks to human rights without processing any identifying or identifiable information. These gaps must be addressed urgently by member states when implementing business and human rights standards in the AI context.

21 of 46 Council of Europe member states have prepared National Action Plans on Business and Human Rights (NAPs). However, few explicitly address business-related human rights risks stemming from AI and similar technologies. Some NAPs, including that of Norway, make reference to the human rights impacts of specific technologies, including military and surveillance technologies that require tighter export licensing regimes to prevent abuse. Others, including the Lithuanian, Polish, and Italian NAPs, reference the promotion in particular of renewable, environmentally friendly, and ecologically sound technologies. Switzerland's NAP mentions working with international institutions to establish "authoritative guidelines on application of UN Guiding Principles to fundamental issues in connection with the development, use and governance of digital technologies." While Luxembourg's NAP specifically addresses the human rights risks that might arise from the development and use of AI systems, it appears to rely on data protection laws as the relevant legal framework, even though the NAP refers also to other human rights implications.

> *" The voice of businesses is overamplified in all AI discussions. Because too little human rights expertise has been included in the development of AI strategies, they are usually industry-focused and refer, if at all, to vague notions of ethics and self-regulation rather than to binding human rights obligations. Regulation will not kill the industry, but it will stimulate innovation."*

As noted in the Global Alliance of National Human Rights Institutions (GAHNRI)'s Edinburgh Declaration, which pre-dates the UN Guiding Principles on Business and Human Rights, NHRIs have a key role to play in ensuring that business actors respect human rights, including through education, monitoring of human rights violations, complaints handling and mediation. NHRSs, more broadly, can provide guidance to member states on the regulatory gaps that persist under national law with respect to creating a holistic regulatory environment concerning human rights and equality violations by private sector actors.

Since 2019, AI actors themselves have come forward and called on legislatures across the globe to regulate their industry. These calls have downplayed the role of human rights in such regulation, however, and seem to have been at least in part motivated by the desire of those in the industry to shape ongoing regulatory efforts. It is important for member states to raise awareness, engage and consult with the industry when implementing business and human rights standards and taking legislative action. In doing so, member states must ensure, however, that they do not allow the industry to unduly influence ongoing regulatory efforts and that human rights protection remains central to any legal framework that results from such a process. Once regulatory frameworks are in place, the private sector should be required to be transparent about their compliance with them.

Chapter 3
Information and Transparency

To safeguard human rights effectively, it is essential first to know that they are potentially affected. The use of AI systems in any decision-making process that has a significant impact on a person's life should therefore be made public in clear and accessible terms. However, meaningful transparency requires more than providing information about the existence or use of an AI system. It should be made explicit why the decision to introduce an AI system was taken in the first place, what the advantages of automation are, the number of expected errors in the form of false positives and false negatives, and what possible human rights risks exist so that individuals are able to understand the workings of the AI system, the trade-offs that it contains, and the processes according to which decisions are reached and verified. Similarly, oversight processes over an AI system must be transparent and accompanied by publicly accessible information.

Member states have shown some commitment to strengthening transparency regarding the use of AI by public authorities, including with respect to procurement processes and the maintenance of public scrutiny over them:

In January 2021, the Dutch House of Representatives passed a motion calling upon the government to set up an algorithm register for AI used by public bodies, which should by 2023 "describe which algorithms the government uses, for what purpose and on which datasets they rely on, so that everyone can monitor whether the algorithms are discriminatory." In November 2021, the UK Government adopted the Algorithmic Transparency Standard, which obliges the public sector to provide more information on the role of algorithms in supporting decisions affecting individuals, especially in law enforcement.

New legislation in Greece imposes multiple transparency obligations on public bodies. This includes the obligation to establish a register of algorithmic decision-making systems, comply with the principle of transparency in the use of AI systems, and provide information to the public on the existence and methodology of AI systems. Compliance with these obligations will be monitored by the National Transparency Authority. Malta and Denmark have developed AI certification programmes. Malta's national AI certification framework, launched in 2019, is the first in the world and helps to recognise AI systems that have been developed in an "ethically aligned, transparent and socially responsible" manner. The D-Seal developed by the Danish government enables consumers to know which companies handle data and AI in a "trustworthy, ethical, and secure" way.

Local authorities have also developed creative tools for enhanced transparency on when and where an AI system is being deployed. For example, the cities of Helsinki in Finland and Amsterdam in the Netherlands jointly launched their public AI registers in September 2020. Other Dutch cities, such as Utrecht, The Hague and Rotterdam have since followed. These are positive developments and good practices for all member states.

> *"Algorithms are created in black box environments and without public participation. Intellectual property protections are valued higher than the right to information, this is a massive transparency problem. Access to information for NHRSs must be ensured through cooperation duties."*

It is also essential, however, to ensure that NHRSs are mandated to monitor and promote transparency, accountability, and public awareness throughout all processes and that related legal frameworks require private sector actors to co-operate with judicial and non-judicial bodies in a transparent manner, including NHRSs. The Dutch Ombudsman Institution, for instance, has issued a report offering guidance to the public authorities about the appropriate use of data and algorithms by the government, which included recommendations on clarity, accessibility and a solution-focused handling of difficulties that may arise. In her 2022 Annual Report, the Croatian Ombudswoman issued a recommendation to the Ministry of Economy and Sustainable Development to establish a register of AI systems that are used in the public sector.

Procedural rights should enhance effective transparency, such as the right to access information held by public authorities on publicly used AI systems, without a need to justify an access request. Particularly relevant pieces of information on AI systems should be required to be made public proactively and NHRSs should have the power to request and access any information that is necessary for the fulfilment of their mandate from public and private sector bodies, including documentation on HRIAs.

Chapter 4
Public Consultations

Public consultations about the design, development and deployment of AI systems in state administration play a key role in ensuring transparency, accountability and informed public participation. The 2019 Recommendation calls on member states to conduct public consultations at the procurement and HRIA stages as a minimum. When performed properly, public consultations provide an opportunity for all stakeholders, including state actors, private sector representatives, academia and civil society to provide input.

Public consultations related to AI have been held in many member states in recent years. For example, Ireland undertook an extensive consultation across the government, relevant stakeholders (industry, experts, academics, research organisations) and the general public when developing its national AI strategy. While no member state appears to have conducted public consultations specifically on AI and human rights or with regard to the procurement or deployment of a specific AI system, there are positive examples of human rights being addressed in some of the more general consultation processes. For example, the Malta.AI Taskforce, a group of experts in charge of developing Malta's AI strategy, launched a public consultation when drafting the Malta Ethical AI Framework, a set of guidelines on ethical and trustworthy AI. One of the framework's objectives is to achieve respect for all applicable laws and regulations, human rights and democratic values.

> *"Every month new software is being introduced in transportation, health care, security, and other sectors, often bought from third countries. But there is no human rights discussion because the public does not understand the risks. In many countries, there are no legal safeguards at all."*

The Spanish Government conducted two rounds of public consultation during the elaboration of the Charter of Digital Rights, which contains an article dedicated to 'Rights regarding AI'. The consultation gathered citizens' opinions, directly or through representative organisations, on possible problems, solutions and objectives, and citizens could provide comments on the draft Charter and the rules that should be applicable to AI.

As AI systems are fed with data likely to contain historical biases, they can cement inequalities and societal divisions, and exacerbate the discrimination or marginalisation of certain groups. Member states should therefore pay particular attention to multi-stakeholder participation mechanisms and the proactive and timely consultation of individuals or groups who will be most affected by the AI systems in question, including children. NHRSs can act as a bridge between civil society and state authorities and can help to ensure that consultations are meaningful, including by identifying and facilitating the outreach to particularly impacted groups. To ensure that public consultations are easily accessible, the provision of input should be facilitated through other forms than in writing and participation in different languages should be accommodated.

Finally, to ensure that meaningful public consultations can take place, member states should increase their efforts to update their access to information and open government rules, including with respect to public procurement. The United Kingdom was the first Council of Europe member state to issue 'Guidelines for AI procurement', which among other things establish an obligation to conduct public consultations.

Chapter 5
Promotion of AI Literacy

Public awareness and understanding of AI systems are the foundations for creating systems of meaningful oversight and public engagement related to AI and human rights. The 2019 Recommendation emphasises that member states should promote knowledge of AI within government institutions, independent oversight bodies, national human rights structures, the judiciary, and law enforcement. AI literacy among the general public, including on the impact of AI systems on human rights, is another area in which member states should robustly invest through general and targeted awareness raising, training and education efforts, including in schools, and by engaging with marginalised groups. NHRSs, through their promotional mandate, can help facilitate collaboration in this respect, not just between state authorities and civil society, but also between national and regional levels.

> *"The business models are highly problematic. The public does not understand that they are the product of AI development. We must better listen to the younger generations, who are deeply affected at all levels, to understand how they are using AI systems. We must all cooperate to raise public awareness."*

AI literacy has been improved in member states through free national courses or study programmes in education facilities, and public service training. In the Netherlands, for example, a national course on 'AI and Ethics' was launched in October 2022 as a follow-up to a previous course on the basics of AI. The course is free and available to everyone with online access. Ireland's training programme for civil servants, launched in 2021, enables public officials to obtain a "Foundation Certificate" in AI, including a section on ethical practice in AI. Thus far, however, courses have mainly discussed general concepts and principles, rather than legal frameworks and human rights standards, let alone explicitly focus on human rights. A welcome initiative is the training course developed by the Council of Europe on AI and discrimination for NHRSs which has been offered in the UK and France so far, and is planned in Belgium in the fall of 2023.

When human rights are addressed in AI literacy campaigns, references are usually made to specific rights only, such as non-discrimination, the right to privacy or freedom of expression. A holistic approach that reflects on the impact – positive or negative – that AI can have on all human rights remains absent. This is urgently needed, however, as the use of AI systems in all aspects of our lives continues to increase and both public authorities and the general public require a better understanding to engage with the impacts at not only the individual but also the collective level. Given their specific knowledge of the human rights and equality situation in member states, NHRSs can play an important role in providing guidance in the design and roll-out of general education and awareness-raising campaigns as well as those that target specific groups, including but not limited to AI actors themselves, such as coders and engineers, and individuals most affected by possible AI harms.

Chapter 6
Independent Oversight

Independent and effective oversight of AI processes is crucial to ensure human rights compliance. While the legal framework and set-up of adequate oversight structures remains the subject of extensive debate in Europe, AI is developing at an exponential rate and the need for the proactive investigation and monitoring of the human rights implications of such developments is greater than ever.

The EU's proposed AI Act envisions national supervisory authorities that are responsible for the application and implementation of the law. This regulation is expected to give states some flexibility to decide whether to create a new supervisory authority or vest the new powers under the AI Act in an existing supervisory authority. As a result, there may be different approaches to oversight of its implementation across the EU once the AI Act comes into force. Whatever decision member states take in this regard, it is essential that existing NHRSs are closely consulted and involved in relevant decision-making processes and that ongoing cooperation between NHRSs and other relevant stakeholders, including regulatory authorities, is institutionalised into all national and regional oversight processes to ensure a truly multidisciplinary approach that is fully inclusive of human rights.

It is not yet clear whether the Council of Europe Framework Convention on Artificial Intelligence, Human Rights, Democracy and Rule of Law will contain an obligation on states parties to establish national supervisory authorities responsible for overseeing compliance with its provisions. As stressed in the 2019 Recommendation, independent and effective oversight over the human rights compliance of AI systems throughout their lifecycle is crucial to ensure that Council of Europe member states live up to their obligations. Supervisory authorities should have full formal and functional independence and be granted a strong mandate, including with respect to investigative powers, complaint-handling, reporting, and awareness-raising. They should have the power to suspend the deployment or use of an AI process in case of established human rights violations.

> *"We need far more holistic oversight and human control, including over the way HRIAs are conducted, and including also regarding the risks to rule of law and democracy."*

Some states have already taken very early steps in establishing supervisory authorities for AI. Spain is one of the first states in Europe to initiate the establishment of a separate agency. The Spanish Agency for the Supervision of Artificial Intelligence (AESIA) is expected to become functional in late 2023. AESIA will supervise the creation and use of AI systems, especially those that might affect fundamental rights, and take measures to reduce the risks to these rights. Close coordination with the Ombudsman and other relevant institutions will therefore be important.

Other member states have also taken steps to establish oversight mechanisms for AI. However, these have often favoured the reliance on ethical frameworks to monitor and assess compliance in specific sectors, without reference to established human rights standards. For example, in 2019, the Prime Minister of France established a pilot digital ethics committee, concerned with examining the use of AI for chatbots, autonomous vehicles, and medical diagnoses. The Finnish Government has established an AI Ethics Committee to enhance understanding of ethical principles and ensure that Finland's AI development is "human-orientated and based on trust." As noted above, ethical frameworks, while playing an important role in broader AI management and governance, should not be referred to at the expense of human rights standards that member states are legally obliged to safeguard.

There has also been a trend in Europe for Data Protection Authorities (DPAs) to be considered appropriate AI-related oversight mechanisms. While these bodies have strong mandates and are vested with the power to impose fines in case of non-compliance with data protection safeguards, they will rarely have the comprehensive human rights knowledge, technical expertise and mandate required to effectively monitor, investigate, and handle complaints regarding broader human rights violations stemming from AI systems.

AI supervisory authorities must be provided with appropriate resources and adequate interdisciplinary expertise and competences to deal with the variety and complex ways in which AI systems can interfere with human rights. They should further be granted the explicit mandate to investigate and monitor the actions of both public authorities and private sector actors to ensure that human rights risks are identified early and that human rights violations are prevented from occurring in the first place, thereby reducing the need for remedies and compensation. Civil society and NHRSs should be closely involved in the set-up and operation of oversight mechanisms to ensure full transparency and accountability. NHRSs can provide important guidance to member states to ensure that the oversight mechanisms deployed to oversee AI systems, be that through the creation of separate agencies, the integration into existing institutions, or the establishment of well-coordinated multi-institution mechanisms have an adequate mandate and powers to properly reflect the variety of ways in which human rights harms can be caused by AI systems. Whatever the oversight model chosen, it is essential to ensure that its proper functioning is adequately resourced.

Chapter 7
Effective Remedies

In the first place, member states should seek to avoid human rights violations by acting preventively, including through adequate legal frameworks and effective oversight mechanisms, rather than through 'test and remedy' approaches. Nevertheless, effective remedies must be available and accessible to those whose human rights have been violated in the design, development, or deployment of an AI system.

As noted above, legislative developments have been slow since 2019 compared to the rate of technological development. Nonetheless, cases have been brought before judicial and non-judicial bodies relating to the human rights impacts of AI systems. These cases demonstrate the need for redress and remedies for AI-related human rights harms, but they also showcase the fact that the legislation in place is inadequate to effectively remedy such harms. The few cases that have been brought, many relying on data protection regulations, also suggest a significant underreporting of human rights harms stemming from the use of AI systems, highlighting the need for greater access to information, knowledge, and expertise to effectively identify and respond to such harms.

In 2020, the District Court of The Hague overturned a law authorising government use of a risk model to identify individuals suspected of having committed welfare and other types of fraud. The District Court found the law to be in violation of the right to privacy, referring also to its discriminatory effects, as certain groups of beneficiaries were automatically profiled as constituting higher risks of fraud, resulting in them receiving mistaken demands for hefty paybacks that ruined families and drove many to depression and despair.

In the same year, the Court of Appeal of England and Wales found that the legal basis for police use of automated facial recognition technology was insufficient, as it left too much discretion to police officers. It further considered that the police had failed to carry out adequate Data Protection and Equality Impact Assessments in relation to the technology. Also in 2020, a court in Bologna Italy ruled that a food-delivery company's algorithmic system for setting conditions on delivery rider's access to work amounted to indirect discrimination. In 2021, the Polish Supreme Administrative Court found that the algorithm for the System of Random Allocation of Cases, which automatically assigned cases to judges, was disclosable under freedom of information laws. In the same year, the Italian Supreme Court called for a higher degree of informed consent from data subjects to be given in the context of AI-driven reputation rating systems. In 2022, the District Court of Amsterdam and the City of London Magistrates Court ordered that drivers affected by algorithmic "robo-firing" applied by a ride-hailing company should be reinstated and paid compensation.

Similar developments can be seen before non-judicial bodies. In one of the first important cases in 2017, the Finnish Non-Discrimination Ombudsman took a case to the National Non-Discrimination and Equality Tribunal against a bank concerning use of automated decision-making in granting loans. The latter concluded that the practice was discriminatory on multiple grounds and imposed a substantial fine on the party found guilty. In May 2022 the UK Data Protection Authority imposed a fine of £7.5 million against a facial recognition company for using images of people in the UK, ordering it to delete all data belonging to UK residents from its systems, and the Dutch Data Protection Authority imposed its highest fine ever (€3.7 million) on the Dutch Tax Administration for unlawfully processing personal data over a period of six years through its algorithm-based 'fraud identification facility'. This fine came in addition to a €2.75 million fine imposed in 2021 on the Tax Administration for its use of a discriminatory algorithm during the Dutch childcare benefits scandal.

The right to an effective remedy entails the right to prompt and adequate reparation and redress. In the aforementioned cases, however, it took several years before a first instance decision was delivered. The data rights NGO noyb reported that its data protection complaints against a number of powerful online platforms that use AI as part of their business models had been considered by national DPAs for over three years without a final decision was made. More must thus be done to ensure that mechanisms for redress are prompt and efficient.

> " *We have no comprehensive framework to ensure effective remedies but only patchwork responses. In cases where algorithmic discrimination is alleged, the burden of proof should be on the user of the algorithm, not the claimant."*

To fully understand the human rights harms presented by an AI system, judicial bodies, non-judicial bodies, and individuals bringing claims must have access to the information needed to properly assess claims concerning such technologies. As outlined above, AI actors have been shown to overly protect and withhold this information, frustrating the mechanisms for accessing remedies. In the case concerning the food-delivery company's algorithm, the Bologna court criticised the fact that it did not have access to more information on the workings of the statistical model and held that the company failed to disclose the operating rules of the algorithm or the specific calculation criteria adopted to determine the statistics of each rider, precluding a more in-depth examination of the case.

The right to an effective remedy is an indispensable aspect of human rights law, and member states must ensure that avenues of effective redress are available and accessible to everyone claiming to be a victim of a human rights violation arising from an AI system, paying particular attention to especially vulnerable groups, including children. As the cases documented in this section show, remedies themselves can vary in nature but they should all be capable of directly remedying the impugned situation arising from the AI system both in theory and in practice. As AI systems usually interfere with the rights of large numbers of individuals, member states should also consider adopting legal frameworks for collective redress in cases concerning AI-related human rights harms.

Concluding Observations and Recommendations

Technology is part of the world we live in and part of nearly every aspect of our lives. This phenomenon will further increase in the foreseeable future and AI will only become more complex and sophisticated. As the design, development and deployment of AI systems have the potential to impact significantly on our human rights and living environments, member states must pay heightened attention to ensuring that people's human rights are safeguarded throughout these processes. Given their essential role in the supervision and enforcement of the rights enshrined in the European Convention on Human Rights and in other human rights instruments, NHRSs play a key role in ensuring that this responsibility is met.

As societies, we are continuously working to protect and safeguard human rights and interests in areas that are complex for many of us. We are not all engineers or car designers, yet we all understand and adhere to production requirements and traffic rules to safeguard human life and individual rights. We are not all economists, yet we can agree on the regulation of our financial markets. The same principle applies to AI. We do not all have to understand how it works in detail to be able to put in place the systems and processes that effectively protect human rights when AI is developed and used. The fact that an administrative decision may be formed or taken through an opaque algorithmic process does not alleviate the state from its legal obligation to ensure that the decisions are based on lawful criteria and open to review.

The continued narrative of AI as so highly technical and inscrutable that it escapes the grasp of human control and effective regulation in a human rights-compliant manner, at least not without risking economic growth and prosperity, dominates the debate and is often furthered by the private sector itself, which tends to prioritise profit maximisation over public concerns. Unfortunately, however, this misconception has led to a remarkable reluctance at senior policy level to engage comprehensively and in-depth with the potential human rights harms caused by the increasing development and deployment of AI, hindering the effective enforcement of existing legal standards and the creation of adequate mechanisms to mitigate threats. As exemplified by the continued use of highly intrusive spyware, member states are overall still moving too slowly in their responses to human rights risks posed by AI, particularly when developed and deployed by powerful private sector actors.

In the 2019 Recommendation, the Commissioner set out practical guidance on how member states could protect and promote human rights in relation to the design, development and deployment of AI systems. Four years have since passed and Europe has been facing multiple and interdependent human rights crises, posing significant challenges to governments and populations alike. Throughout these crises, the reliance on digital technologies and data-enhanced automation has only increased and the technology industry has kept growing.

Member states are exploring the topic of AI regulation at policy level and taking steps to adopt legislative frameworks or establish oversight mechanisms for AI systems. Judicial and non-judicial bodies have been considering how existing legal frameworks apply to cases concerning AI systems and have reviewed complaints brought before them. Amid ongoing negotiations at regional level towards the adoption of new legal frameworks related to AI systems and their impacts on human rights, it is vital to focus on the extent to which existing standards and human rights safeguards, including those contained in the 2019 Recommendation, have been implemented so far. Based on her findings and in consultation with NHRSs, the Commissioner has observed three interdependent trends that constitute obstacles to the full implementation of international human rights standards related to AI in Europe. In response, she formulates the below recommendations to member states of the Council of Europe to overcome the remaining shortcomings and strengthen the enforcement of human rights obligations with respect to the design, development, and deployment of AI systems.

Lack of Comprehensive and Human Rights-Based Approaches

Member states have overall adopted sector-specific approaches to their implementation of the 2019 Recommendation and other international human rights standards related to the use of AI systems, rather than ensuring their consistent and holistic application across all relevant sectors and with respect to all actors. Legal frameworks, where existing, have often not been effectively and promptly enforced, as infrastructure dependence on large platforms may hinder implementation and oversight remains fragmented. When assessing the impact of AI systems, member states tend to focus on a subset of rights, such as data protection rights or non-discrimination principles, or have made vague references to ethical frameworks that do not sufficiently integrate all human rights standards and obligations. DPAs have been relied upon to provide independent oversight of AI systems and their impact on human rights, resulting in a skewing towards data protection concerns among policymakers and an overall still limited understanding of the broader human rights risks and impacts that must be considered. As a result of sectoral approaches to regulation, judicial and non-judicial actors examining complaints related to human rights harms stemming from AI systems have also applied sectoral legal frameworks to attempt to remedy such harms.

» Member states should reassess their legal frameworks, including those related to remedies and oversight, to ensure that they comprehensively encompass the scope and circumstances of human rights complaints concerning AI systems. Newly-developing legal frameworks related to AI should be brought in line with existing human rights safeguards, such as the reversal of the burden of proof in discrimination complaints. To be effective and meaningful, all legal frameworks must be strictly implemented and enforced in all cases.

» Legal frameworks should respond to all AI systems that have the potential to interfere with human rights, regardless of the intended sector in which they will be used and in full awareness of the fact that human rights must be protected at all stages of an AI system lifecycle and throughout its interaction with other systems.

» Member states should, in close consultation with NHRSs, rights-holders and civil society, develop a specific, stand-alone process for public authorities to follow when carrying out HRIAs in relation to AI systems. These HRIAs should be mandatory, performed ex ante and ex post, and their processes and outcome made publicly accessible. Public authorities should not procure or use AI systems from third parties in circumstances where they cannot carry out an effective HRIA due to a lack of information about the system. All barriers to the effective functioning of HRIAs should be proactively removed.

- Member states should ensure that AI design, development and deployment is supervised through one or more authorities that are fully independent, adequately resourced, and mandated to monitor, investigate and handle complaints in relation to actions taken by public authorities and private sector actors. In their oversight function, supervisory authorities should be required to cooperate closely with other independent institutions, notably NHRSs.
- Judicial and non-judicial bodies, including regulatory authorities and NHRSs, should be adequately resourced and have the capacity to intervene in circumstances where compliance with any human right has been raised in the context of an AI system in an effective and timely manner, and from a holistic and intersectional human rights perspective.

Insufficient Transparency and Information Sharing

Although there have been notable initiatives to improve transparency around AI systems in certain member states, clear and updated information around the use of AI systems and their potential impact on human rights remains insufficiently accessible across Europe. Intellectual property protections related to trade secrets still constitute obstacles to the enforcement of information rights, including for the judiciary, NHRSs and regulatory authorities, who require full and effective access to all relevant information, including datasets and source-codes, to perform their independent oversight mandate and monitor and review possible human rights violations. Public consultations, where they have taken place, do not usually address the broader human rights impacts of AI systems, and there have been few proactive efforts to ensure that the perspectives of vulnerable groups, such as marginalised communities or children, are being included. Member states have yet to deliver robust, structural education programmes that have an explicit focus on the human rights implications of AI systems and are effectively accessible to all strata of society.

- Member states should boost initiatives, such as public registers, that promote transparency about the use of AI systems in daily life and the harmful impacts they may have on human rights, including about their basic functioning and the aims behind them. Newly established or appointed supervisory authorities should regularly report on their activities to the public, in easily accessible and understandable formats.
- Judicial and non-judicial bodies, including regulatory authorities and NHRSs, should be granted the powers to oblige AI actors to disclose detailed information about the AI systems under examination and to take measures against non-compliant public or private entities.

- When preparing, adopting and implementing legal frameworks that address the human rights impacts of AI systems, member states should hold timely and regular public consultations, targeting not only experts, industry, researchers, NHRSs, civil society and academia, but, importantly, the wider public and representatives of those groups that are most affected, including children, and their representatives.
- To make these stakeholders aware of the existence of public consultations and the opportunity to engage, member states should conduct targeted outreach through multi-stakeholder participation methods and facilitate the provision of input in multiple formats.
- Member states should take immediate steps to design, resource and deliver comprehensive AI literacy and awareness-raising campaigns that ensure that those involved in the use and oversight of AI systems and those affected by it adequately understand the systems and their multiple impacts on human rights. Member states should closely involve NHRSs in these processes and provide them with the relevant resources to effectively develop their capacities related to AI.

Initiatives to use AI for Strengthening Human Rights

Given that the AI market, with all its potential and risks, lies firmly in the hands of private sector actors, and most AI development is driven by commercial incentives, public authorities have overall adopted a reactive rather than proactive approach. This has allowed the private sector to take the lead in the strategic direction of AI development based on business models that regard individuals as consumers and prospective targets of development. By delaying regulation that would prompt alternative innovation, member states risk missing the opportunities that AI capacities offer towards the implementation and strengthening of human rights protections and the fundamental principles of democracy and the rule of law, equality promotion and good governance. While encouraging initiatives exist, such as those related to equality promotion and data protection by design, there is still an untapped potential for AI design, development and deployment that is incentivised by value-based objectives, such as exposing and dismantling existing prejudice and stereotypes, boosting public participation, amplifying the voices of those usually unheard, and addressing inequalities by helping prioritise those most in need. Finally, to ensure that AI development can deliver on its promise for all, existing digital divides must be overcome, and safe and reliable Internet access made available throughout and for all strata of society.

- Member states, public authorities and other AI actors should refer explicitly to established human rights law and legal obligations when preparing, adopting, or implementing regulatory measures related to AI systems and incentivise the private sector to direct innovation towards human rights-compliant systems. Where regulatory sandboxes are adopted, NHRSs should be consulted and provided with an opportunity to provide human rights guidance to AI developers and promote the integration of human rights safeguards throughout the AI lifecycle.

» AI systems must be subjected to full human rights scrutiny, control and product liability frameworks, through institutionalised cooperation between the various non-judicial bodies involved in its supervision, including notably the NHRSs as principal guardians of human rights.

» Judicial and non-judicial bodies, including regulatory authorities and NHRSs, should be granted the power to prevent any further human rights harms by AI systems, including through the imposition of moratoriums on further use, the deletion of datasets, or the destruction of systems that are the product of unlawful data processing.

» Member states should proactively engage with AI developers, civil society and independent experts to explore new opportunities offered by AI technologies to enhance human rights protections and promote effective equality. NHRSs should be closely involved in these processes. Targeted funding should be made available for research exploring the human rights and equality promotion potential of AI systems.

» Member states should take urgent action to implement the UN Guiding Principles on Business and Human Rights and the Recommendation CM/Rec(2016)3 of the Committee of Ministers to member states on human rights and businesses.

The Commissioner for Human Rights is an independent and impartial non-judicial institution established in 1999 by the Council of Europe to promote awareness of and respect for human rights in the member states.

The activities of this institution focus on three major, closely related areas :

- country visits and dialogue with national authorities and civil society,
- thematic studies and advice on systematic human rights work, and
- awareness-raising activities.

The current Commissioner, Dunja Mijatović, took up her funtions in April 2018. She succeeded Nils Muižnieks (2012-2018), Thomas Hammarberg (2006-2012) and Álvaro Gil-Robles (1999-2006).

www.commissioner.coe.int

ENG

www.coe.int

The Council of Europe is the continent's leading human rights organisation. It comprises 46 member states, including all members of the European Union. All Council of Europe member states have signed up to the European Convention on Human Rights, a treaty designed to protect human rights, democracy and the rule of law. The European Court of Human Rights oversees the implementation of the Convention in the member states.

COMMISSIONER
FOR HUMAN RIGHTS

COMMISSAIRE AUX
DROITS DE L'HOMME

COUNCIL OF EUROPE

CONSEIL DE L'EUROPE