



# **Comments on the draft recommendation of the Committee of Ministers to member states on the human rights impacts of algorithmic systems by Access Now and the Wikimedia Foundation**

August 2019



## Introduction

Access Now welcomes the opportunity to provide feedback on the draft recommendation of the Committee of Ministers to member states on the human rights impacts of algorithmic systems (hereinafter: “draft recommendation”). The Wikimedia Foundation joins Access Now in its comments on the draft recommendation.

**Access Now** (<https://www.accessnow.org>) defends and extends the digital rights of users at risk around the world. By combining direct technical support, comprehensive policy engagement, global advocacy, grassroots grantmaking, and convenings such as RightsCon, we fight for human rights in the digital age.

In recent years, Access Now has developed a position on AI governance, advocating for [human rights centric approach to AI](#).

Our work includes the launch of the [Toronto Declaration on Equality and Non-Discrimination in Machine Learning](#) – a statement on the role and accountability of states and the private sector where human rights harms arise. Spearheaded by Access Now and Amnesty International at RightsCon 2018, the Declaration is endorsed by Human Rights Watch and the Wikimedia Foundation, among others.

Access Now’s Europe Policy Manager, Fanny Hidvegi, got selected to join the European Union’s [High-Level Expert Group on Artificial Intelligence \(AI HLEG\)](#). We published our [preliminary recommendations](#) to improve the Ethics Guidelines on Trustworthy AI, and the [positives and negatives](#) of the Policy and Investment Recommendations for Trustworthy AI.

Finally, we published two reports on AI. One [that maps and analyses strategies and proposals for regulation on artificial intelligence in Europe](#). The report covers regional strategies from the European Union and the Council of Europe as well as national plans from several member states including France, Finland, Germany, and Italy. Access Now lays out a criteria to assess AI strategies to make sure that the development and deployment of AI is individual-centric and human rights-respecting. The second report, titled [Human rights in the age of artificial intelligence](#), provides a comprehensive analysis on the potential pitfalls of AI, and how to address AI-related human rights harms.

The basis for our recommendations can be found in these above listed papers, analyses and articles. In our response to the consultation we will first outline (1) our suggestions for further edits of the main principles stated in the Recommendation’s preamble. Second (2) we will propose concrete improvements for the Guidelines governing states’ obligation and private sector actors’ responsibilities to protect human rights of online users in the context of algorithmic systems and third



(3) we will emphasise the most essential elements of the Guidelines that Access Now is fully supporting and has been advocating for in our own work.

**The Wikimedia Foundation** (<https://wikimediafoundation.org>) is the non-profit organization that hosts and supports [Wikipedia](#), the world's largest online encyclopedia, and other websites for free knowledge. Through a collaborative process, Wikipedia has grown to include approximately 50 million articles in over 300 different languages. It is viewed more than 15 billion times each month. Many Wikimedia contributors are in Europe, and Europeans use Wikipedia and its sister websites on a daily basis.

We work towards a world in which every single human being can freely share in the sum of all knowledge. In accordance with this vision, we believe that everyone should have access to factual and accurate information as well as the ability to document the world they live in, in collaboration with others. We strongly believe in the fundamental right to freedom of expression and the right to freedom of assembly and association, which are enshrined in international law, including in the European Convention of Human Rights.

The Wikimedia Foundation supports the community of editors who contribute to Wikipedia by empowering these volunteers to improve, grow, and maintain the encyclopedia more efficiently. As part of these efforts, different teams of the Foundation take a [human-centered approach to developing machine learning systems](#) that aid volunteers in using [structured data to catalogue high-quality images](#), help them understand when a contribution to Wikipedia [may not comply with citation requirements](#), and allow them to [identify bad edits more quickly](#).

The Wikimedia Foundation is a signatory of the Toronto Declaration and supports policy that protects internet users from discrimination, which negatively affects their human rights. Collaborative online projects can be [significantly impacted by the mandatory use of algorithmic systems](#) to monitor or filter the information that is uploaded by the users of a platform. Algorithmic systems should aid and support humans in their ability to participate in culture and the digital economy. The Council of Europe's draft Recommendation on the Human Rights Impacts of Algorithmic Systems gives welcome guidance for the regulation, development, and use of such systems.

## Specific recommendations for further improvements

### Preamble

#### Paragraph 4

Proposed amendment: (deletions ~~striketrough~~, additions **bold**)

“Operating principally by detecting patterns in large datasets, algorithmic systems offer the potential to improve the performance of services (particularly through increased precision and targeting), provide new solutions, and **in some cases could deliver improvements in** ~~enormous~~ efficiency and effectiveness ~~gains~~ in task and system performance”.

“Algorithmic systems can strengthen individual autonomy and self-determination and can enhance the exercise of human rights, for instance, by broadening access to information or by facilitating the enjoyment of the freedom of assembly and association, including by creating innovative ways of associating with others.”

Justification:

The language in the final clause of this sentence was hyperbolic (“enormous”), and should be modified as indicated to a more neutral and realistic tone, in line with the rest of the document.

In the second quoted sentence of Para 4 of the preamble we recommend that it should not state that algorithmic systems can strengthen individual autonomy and self-determination without acknowledging the negative impacts to autonomy as well.

#### Paragraph 5

Proposed amendment: (deletions ~~striketrough~~, additions **bold**)

“However, there are also significant human rights challenges attached to the increasing reliance on algorithmic systems in everyday life. Their functionality is frequently based on the systematic aggregation and analysis of data collected through the digital tracking of online and offline behaviour of individuals and groups at scale. In addition to personal data protection and privacy costs, tracking at scale can have an important chilling effect on freedom of expression, **freedom of assembly, the right to equality and non-discrimination**, and other human rights.”

Justification:

The specific functionality described in the second sentence of this paragraph (digital tracking) has a clear impact and chilling effect on freedom of assembly of individuals. If individuals are

tracked based on their online registration for or social media activity related to a protest (via an 'event' on Facebook, for example) and their actual physical attendance at the protest, this can lead to a chilling effect on freedom of assembly as individuals may be disinclined to exercise this legitimate right in the future. Therefore, we recommend that freedom of assembly be explicitly mentioned in this paragraph.

Second, we refer to the Toronto Declaration, which outlines the increased risk that machine learning systems, which can be opaque and include unexplainable processes, pose to this right.

### **Operative section**

#### **Obligation of states with respect to the protection and promotion of human rights and fundamental freedoms in the context of algorithmic systems**

##### Paragraph 1.3

Proposed addition: (deletions ~~striketrough~~, additions **bold**)

All relevant actors, including private sector actors, media, education establishments, academia and technical institutions, should promote, in a tailored and inclusive manner (taking account of diversity with respect to, for instance, age, gender, race, ethnicity or socio-economic background), appropriate levels of understanding of the functioning of algorithmic systems and of the human rights risks stemming from their use in everyday life, enhancing the ability of all users to be aware of their rights and freedoms, **how to act upon them**, and use these technologies for their benefit.

##### Justification:

The addition aims to emphasise that the listed actors should promote the appropriate levels of understanding of the functioning of algorithmic systems not only to enhance the ability of all users to be aware of their rights but also how to use them.

### Paragraph 3.3

Proposed addition: (deletions ~~striketrough~~, additions **bold**)

Depending on the potential impact of the algorithmic system on human rights and in order to avoid compromising other human rights, testing should, where possible, be performed without using real personal data of individuals, and should be informed through a diverse and representative stakeholder process, taking due account of the externalities of the proposed system on populations and their environments before and after deployment. We used underlying to indicate the wording that is not clear enough.

### Justification:

Paragraph 6 of the Preamble acknowledges that human rights of users may be negatively impacted if algorithm is being trained on non-observational and non-personal data. We appreciate the fact that potentially harmful outcome rising from the use of synthetic data is so well described and explained. Having said that, it is not quite clear from the wording of Paragraph 3.3 what is the exact meaning of the term “without using personal data of individuals.” The terminology needs to be clarified and bring in line with Para 6 of the Preamble.

### Paragraph 3.5

Proposed addition: (deletions ~~striketrough~~, additions **bold**)

Parallel modelling: As regards the use of algorithmic systems in the delivery of public services and in other high risk contexts in which States use such technologies, alternative and parallel modelling should be performed using other methods in order to ensure that the performance and output of the algorithmic model can be adequately tested in comparison to other options. **Parallel modelling or other form of measurements should take place during the decision-making process whether to procure or apply an algorithmic system in the delivery of public services and in other high risk contexts.**

Justification:

We welcome parallel modeling to measure the performance or output of an algorithmic system. It is necessary, however, to incorporate such factor in the decision-making process of States whether there is any justifiable reason and evidence that supports the procurement or the application of an algorithmic system instead of other policy or technology options or mechanisms.

Paragraph 5.1

Proposed addition: (deletions ~~striketrough~~, additions **bold**)

Indicators: States should cooperate with each other and with private sector actors and relevant rights groups to develop and implement appropriate indicators, criteria and methods for state of the art human rights impact assessment processes to be conducted with regard to all algorithmic systems with potentially significant human rights impacts, with a view to evaluating potential risks and tracking actual harms, especially when such mechanisms are applied for non-targeted, explorative purposes. **States may delineate the types of algorithmic systems that are subject to human rights impact assessments under the law, but such delineations must be comprehensive enough to cover all algorithmic systems that have the potential to interfere with an individual's human rights at any stage of the algorithmic system lifecycle.**

Justification:

We recommend adopting the language from the Council of Europe Human Rights Commissioner's recommendations ([Unboxing Artificial Intelligence: 10 steps to protect human rights](#)).

Paragraph 5.2

Proposed addition: (deletions ~~striketrough~~, additions **bold**)

States should ensure that they, ~~as well as any private actors engaged to work with them or on their behalf~~, regularly conduct such **mandatory** human rights impact assessments prior to public procurement, during development, at regular milestones, and throughout their

context-specific use to identify risks of rights-adverse outcomes. **Human rights impact assessments should be conducted for algorithmic systems already in place or being proposed.** For algorithmic systems with high risks to human rights, impact assessments should include an evaluation of the possible transformations that they may bring upon existing social, institutional or governance structures. **States have to ensure that the human rights impact assessment must also involve a meaningful external review of AI systems by independent oversight body with relevant expertise. Public authorities should include National Human Rights Structures in carrying out these external reviews.**

**[Here or in a following paragraph to detail the potential outcomes of a human rights impact assessment:]**

**In circumstances where the self-assessment or external review discloses that the algorithmic system poses a real risk of violating human rights, the human rights impact assessment must set out the measures, safeguards, and mechanisms envisaged for preventing or mitigating that risk. In circumstances where such a risk has been identified in relation to an algorithmic system that has already been deployed by a public authority, its use should be immediately suspended until the above mentioned measures, safeguards and mechanisms have been adopted. Where it is not possible to meaningfully mitigate the identified risks, the algorithmic system should not be deployed or otherwise used by any public authority. Where the self-assessment or external review discloses a violation of human rights, the public authority must act immediately to address and remedy the violation and adopt measures to prevent or mitigate the risk of such a violation occurring again. The human rights impact assessments, including research findings or conclusions from the external review process, must be made available to the public in an easily accessible and machine-readable format. Public authorities should not acquire algorithmic systems from third parties in circumstances where the third party is unwilling to waive restrictions on information (e.g. confidentiality or trade secrets) where such restrictions impede or frustrate the process of (i) carrying out HRIAs (including carrying out external research/review), and (ii) making HRIAs available to the public.**

Justification:

First, although States often rely on third parties to design and implement AI systems, ultimate duty to protect individuals against human rights violations must lie with States themselves and not private actors. A human rights due diligence or risk assessment for the private sector should be dealt with separately.



Second, it should be emphasised that HRIAs are mandatory and that they should be conducted for algorithmic systems that are already in place or being proposed (or being under procurement.)

Third, even though Paragraph 5.3 underlines the importance of a meaningful external review, the human rights impact assessment is an integral part of this mechanism. Thus, the need of an independent oversight body that provides human rights expertise should be mentioned in both paragraphs.

Finally, we recommend bringing this paragraph in line with the Council of Europe Human Rights Commissioner’s recommendations ([Unboxing Artificial Intelligence: 10 steps to protect human rights](#)) and include the details about the potential consequences of a human rights impact assessment and the HRIA’s transparency.

#### Paragraph 6.2

Proposed addition: (deletions ~~strikethrough~~, additions **bold**)

Advancement of public benefit: States should engage in and support independent research aimed at assessing, testing and advancing the potential of algorithmic systems for creating positive human rights effects and for advancing public benefit **including to ensure that the interests of marginalised and vulnerable individuals and groups are adequately taken into account and represented**. This may require the anticipation and possible discouragement of influences that may exclusively favour most commercially viable optimisation processes.

#### Justification:

We support the notion in the second sentence to emphasise the priority of public benefit over most commercially viable optimisation processes. The addition to the first sentence aims to ensure that the “calculation” of public benefit is fair and goes beyond the interest of the majority or better represented individuals or groups.

### **Responsibilities of private sector actors with respect to human rights and fundamental freedoms in the context of algorithmic systems**

Paragraph 1.4

Proposed addition: (deletions ~~striketrough~~, additions **bold**)

Discrimination: Private sector actors **that design, develop or implement algorithmic systems should follow an industry standard human rights due diligence framework to avoid fostering or entrenching discrimination and to respect human rights more broadly through all lifecycles of their systems.** ~~produce and provide their products and services without discrimination.~~ They should seek to ensure that the design, development or implementation of their algorithmic systems do not have direct or indirect discriminatory effects or harmful impacts on individuals or groups that are affected by these systems, including those who have special needs or disabilities or may face structural inequalities in their access to human rights.

Justification:

Following the recommendation from the “Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems” it is important to distinguish between human rights impact assessments as state obligations and human rights due diligence processes on the private sector’s side. In addition to the general requirement, the Toronto Declaration sets out the three core steps to the process of human rights due diligence: (i) Identify potential discriminatory outcomes ii. Take effective action to prevent and mitigate discrimination and track responses iii. Be transparent about efforts to identify, prevent and mitigate against discrimination in machine learning systems. You can find more details here: [https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration\\_ENG\\_08-2018.pdf](https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf).

Paragraph 4.2

Proposed addition: (deletions ~~striketrough~~, additions **bold**)

Contestability: Private sector actors should make public information about the number and type of contests made by affected individuals or groups regarding the products and services they offer, **and the outcomes of the contests**, with a view to ensuring that the results do not only lead to remedial action in the specific case but are also fed into the systems themselves to draw lessons from complaints and correct errors before harm occurs at massive scale.

Justification:

The publication of how the complaints are addressed and to what results they lead to strengthen not only the transparency of the process but also supports the desired outcome that these complaints would feed into the systems as well and creates additional accountability mechanism to ensure that the remedy is effective.

## **Support for essential elements of the draft recommendation**

Below we list the paragraphs of the draft recommendation that Access Now is particularly supportive of. We follow the order of the text in this section too.

### **Preamble**

Paragraph 11

Comment and justification:

We commend paragraph 11 for pointing to the specific risk associated with public sector use of algorithmic systems in the provision of public services, and in particular the following sentence:

“In some cases, the application of an algorithmic system may prompt a particular, higher risk to human rights, for instance because it is used by states for their public service or public policy delivery and the individual does not have a possibility to opt out”

This is an extremely important point to make, especially given that groups such as the European Commission’s High Level Expert Group on Artificial Intelligence are specifically calling for increased uptake of Artificial Intelligence in the public sector. We believe that public sector use of AI, or algorithmic systems, poses a particularly high risk, and we therefore commend paragraph 11 for highlighting this fact.

Paragraph 12

Comment and justification:

We strongly support the wording of Paragraph 12, which underlines the importance of

transparent and open public procurements. The text rightly points out the possible danger present in outsourcing a part of public services to private contractors. When governments decide to acquire AI system or element, there should always be a public procurement mechanism in place that is open and transparent. The governments' use of AI should be governed by a high standard, including open procurement standards, mandatory human rights impact assessments, full transparency, explainability and accountability processes.

#### Paragraph 18

##### Comment and justification:

Paragraph 18 recognises human rights responsibilities held by private sector actors, as stipulated in the UN Guiding Principles on Business and Human Rights. We welcome the emphasis on the importance of risk management exercised by private actors. It is fundamental that private actors are able to identify human rights risks, which will enable them to effectively mitigate, prevent and track any potential harms inherent in AI systems over time.

### **Operative section**

#### **Obligation of states with respect to the protection and promotion of human rights and fundamental freedoms in the context of algorithmic systems**

#### Paragraph 2.2

##### Comment and justification:

The quality and accuracy of datasets as well as representative nature of data used for development of machine learning systems should be included in the assessment of any AI tools. For machine learning systems, measurement errors are directly connected to the features that are included in the training data. The issue of data quality needs to be seriously taken into consideration and therefore, it is very appreciated that the draft recommendation dedicate a separate paragraph to this outstanding problem.

#### Paragraph 4.1

##### Comment and justification:

The paragraph highlights States' obligation to establish minimum levels of transparency about the use, design and basic processing criteria and methods of algorithmic systems deployed

by the private sector. Maximum possible transparency is necessary for an AI system in relation to its purpose, how it is used and how it works. The meaningful transparency must continue throughout a system's life cycle. The non-disclosure agreements and other contracts with third parties under the guise of protecting intellectual property are often used as justification for preventing public oversight and accountability. We especially appreciate that paragraph 4.1 rejects this argument and warns about its common exploitation.

### **Responsibilities of private sector actors with respect to human rights and fundamental freedoms in the context of algorithmic systems**

#### Paragraph 5.2

##### Comment and justification:

We strongly support the claim that private sector actors developing and deploying AI systems have to ensure proper training of their staff members. Only a proper training throughout the AI system's life cycle will guarantee the highest standards for human rights impact assessment as well as the systems' reviews. Continuous education and training are a precondition for good accountability practices.

## **Conclusion**

We welcome the Council of Europe draft recommendation on the human rights impact of algorithmic systems with great enthusiasm and appreciation. We especially appreciate the high level of expertise reflected in the Guidelines for States and private sector actors addressing the impact of algorithmic systems on individuals' human rights.

We look forward to working with the Committee on Media and Information Society (CDMSI) to ensure that the design, development and deployment of AI assisted technologies is individual centric and respect human rights. We hope to be informed about the results of September review by MSI-AU and to be involved in the draft finalisation process.

For more information, please contact:

**Fanny Hidvegi** | Europe Policy Manager | [fanny@accessnow.org](mailto:fanny@accessnow.org) | +32489825097