AVAAZ
THE WORLD IN ACTION

27 Union Square West, Suite 500
New York, NY 10003

Avaaz Feedback on the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence and Amending Certain Union Legislative Acts

Introduction

Avaaz is the world's largest online civic movement. Our 69 million members, including 22 million in Europe, campaign for urgent action on the key issues of our time - the climate crisis, ecological collapse, and the erosion of democracy. We have been at the forefront of research on online disinformation, and this has led us to a deep understanding of the role artificial intelligence ("AI") can have in exacerbating risks of harm to fundamental human rights.

AI can improve healthcare and even assist in the modelling needed to reduce carbon emissions but just as with any human tool, AI has the potential to exacerbate the harms and prejudices of any system of control and can be used as a tool for repression, surveillance, violation of privacy, and institutionalised bias.

It is with this in mind that whilst we applaud the Commission's multi-sectoral approach, and think that its assessment system is a major step forward, we are concerned about its **tiered approach to risk assessment that leaves mainstream AI development without any standardised ethical framework or structures.** The approach that limits the assessment of AI to the relatively small section of perceived high risk categories in Annex III misses the fundamental point that AI, without human oversight, has the potential to usher in unparalleled risks to our human rights through myriad apparently low risk systems.

We believe that every AI system provider whose AI system poses a risk to health and safety, or a risk of adverse impact on fundamental rights should conduct an assessment or audit of their proposed AI against the categories currently set out only for "high-risk" AI. That assessment should be openly available to national public authorities who will be established to oversee high risk activities. Our comments below provide the rationale for suggested amendments on the following topics:

Section 1: The right framework for AI risk assessment

Our key suggestions in this section include:
Measures to extend obligations of risk assessment, transparency and accountability across all AI systems that could pose risks of harm to health and safety or a risk of adverse impact on fundamental rights (Articles 7, 13, 14, 52, and 45);

- **Risk assessment**:

    - Article 7: compulsory assessment process for all new AI systems that could pose risks of harm to health and safety or a risk of adverse impact on fundamental rights;
    - Article 7: risk-based approach for providers of any AI system that distributes content online through any form of AI assisted content distribution algorithm and meets the test of posing a risk of harm to health and safety or a risk of adverse impact on fundamental rights.
- **Transparency:**
    - Article 13: transparency of operation on all AI systems that pose risks of harm to health and safety or a risk of adverse impact on fundamental rights.
    - Article 52: transparency in relation to the detection, prevention and investigation of crime.
    - Article 45: extension of the right to appeal to civil society organisations and external stakeholders when they have a legitimate interest in the decisions taken by the notified authorities designated by each member state to carry out third-party conformity assessment for AI systems.
- **Accountability:**
    - Article 14: human oversight for all AI systems.

Measures to promote and protect human rights, democracy and the rule of law worldwide

- **Scope of application**:
    - Article 2: additional provisions for (i) the export of AI systems; and (ii) international cooperation with organisations or third-countries.

Measures to enhance workers' rights:

- Additional rights for workers subject to workplace AI surveillance or monitoring systems.

Section 2: Extension of protection for vulnerable groups and Artificial Intelligence practices which should be prohibited

Our key suggestions in this section include:

- **Extend protection for vulnerable groups:**
    - Article 5,1(b): extend the scope of protection to also include specific groups of persons due to their gender, sexual orientation, ethnicity, race, origin, including migrants, refugees and asylum seekers, and religion.
- **Extend the ban on certain AI systems to private actors as well as public authorities, namely**:
    - Article 5,1(c): social scoring;
    - Article 5,1(d): 'real-time' remote biometric identification systems;
    - Article 5,1(e): the use of AI systems categorising individuals from biometrics into clusters according to certain identitary criteria; and
    - Article 5,1(f): AI systems to infer emotions of a natural person, except for special situations.

Section 1

The right framework for AI risk assessment

Introduction

The system of risk assessments laid out in the Proposal for "high risk" systems must become industry-standard across AI development for all AI systems which could pose a risk of harm to health and safety or a risk of adverse impact on fundamental rights.

Whilst it may be tempting to assume that large areas of AI development are of 'minimal or no risk', such as spam filters, that comfort collapses with scrutiny. Avaaz's work investigating and reporting on online bias, hate and disinformation has demonstrated the huge risks in the automated content and moderation systems of social media, which currently fall outside of Annex III.

The pandemic has provided numerous stark illustrations of the effect of AI taking decisions previously supervised by human moderators. As the tech giants were forced to send human moderators home, they may have hoped that their AI could take on the vital work of safeguarding the information ecosphere they now run. But the AI didn't work as intended, with **devastating impacts on human rights of free speech, freedom for discrimination and the right to life** with reliable information on Covid-19.

Reliable health information related to coronavirus was scrubbed[1] from the internet by the AI's overreaction to key words[2] whilst Avaaz's own reports detail how the platforms failed to curb the harm of disinformation accelerated by their own content recommendation AI. They tried valiantly to label it in many cases, but they just couldn't keep up.[3] In vital areas, such as child exploitation and self-harm, the number of removals fell by at least 40 percent in the second quarter of 2020 because of a lack of humans to make tough calls about what broke the platforms' rules, according to Facebook's own transparency report.[4] **Astonishingly, neither child protection nor suicide prevention are listed in Annex III as areas in which AI deployment is seen as high risk.**

The lockdown measures required by the pandemic also gave us a chance to see how AI without human moderation performed in dealing with sensitive questions of free speech for minorities. Between 2020 and 2021 scores of activists' YouTube accounts in Syria, where citizens and the media rely on social media to document potential war crimes, were closed down overnight — often with no right to appeal decisions made by the system's AI moderation.[5] This had a devastating effect on their fight for freedom and on the personal lives of those affected, "It's not

[1] Kevin McKernan. Twitter post revealing how Twitter censored his scientific article. (2020)
[2] Politico. What happened when humans stopped managing social media content. (2020)
[3] Avaaz. Disinformation Hub. (2021)
[4] Politico. What happened when humans stopped managing social media content. (2020)
[5] France 24. Activists in race to save digital trace of Syria war. (2021)

just videos that have been deleted, it's an entire archive of our life," said Sarmad Jilane, a Syrian activist and close friend of Al-Mutez Billah, one of the affected account holders, who was killed at the age of 21. Sarmad said, "Effectively, it feels like a part of our visual memory has been erased."

Furthermore, the bias phenomenon is known to the industries that use AI systems. For example, two studies in 2019 showed that AI trained to identify hate speech can actually amplify racial bias in failing to properly account for the context a human moderator might spot. Researchers have found that leading AI models for processing hate speech were one-and-a-half times more likely to flag tweets as offensive or hateful when they were written by African Americans, and 2.2 times more likely to flag tweets written in African American English (which is commonly spoken by black people in the US).[6] Another study found similar widespread evidence of racial bias against black speech in five widely used academic datasets for studying hate speech that totaled around 155,800 Twitter posts.[7] Finally, we would add in evidence this study which concluded bias is inevitable in AI coding of AI " *We conclude that discrimination can occur in any sociotechnical system in which someone decides to use an algorithmic process to inform decision-making*."[8]

Meanwhile in the same period, AI allowed hate speech to thrive online. In France, for example, during the period in which AI dominated moderation during the pandemic, reports were published indicating an increase of over 40% in antisemitic terms on Twitter was reported[9] and that less than 12 percent of those posts were removed.

Despite Twitter's known problems with an AI bias, Twitter users uncovered a disturbing example of bias in its image-detection algorithm designed to optimise photo previews. This cropped out black faces in favour of white faces. Twitter apologised for this algorithmic prejudice, but the bug remains.[10] This kind of pervasive discrimination across multiple "low risk" applications has proven effects on mental health, particularly amongst our most vulnerable citizens - such as teenagers - who collectively form an experimental subject body for AI with their daily participation in social media.[11] The potential tragic effects of AI, such as seeding social unrest and hate crime, have been well-documented. In 2020, Avaaz reported on how Facebook's AI in Assam worked with inadequate datasets to recognise hate speech against Bengali Muslims,[12] the European Parliament reported on the disproportionate effect of misinformation on minorities and migrant communities in June 2021,[13] and the UN is still uncovering the full effect of Facebook's inadvertent role in fomenting the Myanmar massacres.[14]

[6] Sap et al.The Risk of Racial Bias in Hate Speech Detection. (2019)
[7] Davidson et al. Racial Bias in Hate Speech and Abusive Language Detection Datasets. (2019)
[8] JSTOR. How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications. (2018)
[9] San Juan Partnership. What happened when machines took over social media. (2020)
[10] The Atlantic. How the Racism Baked Into Technology Hurts Teens. (2020)
[11] The Atlantic. How the Racism Baked Into Technology Hurts Teens. (2020)
[12] Avaaz. Megaphone for Hate. (2019)
[13] European Parliament. The impact of disinformation campaigns about migrants and minority groups in the EU. (2021)
[14] Reuters. U.N. investigators cite Facebook role in Myanmar crisis. (2018)

These issues may be the unintended effects of apparently low risk AI like moderation AI but it will take regulation to focus the efforts of those deploying the AI systems to correct the risks of harm it poses.[15] AI used on online media platforms, specifically recommendation and ranking algorithms, as well as machine learning algorithms used to moderate content should fall within the ambit of this proposed regulation where it poses risks of harm to health and safety or a risk of adverse impact on fundamental rights

And finally, if we turn to other areas of industry beyond online content, we see similar concrete risks to our fundamental human rights in systems excluded from in Annex III. For example, what could seem more low risk than the children's toy industry? But researchers on children's rights have demonstrated several risks that current applications pose such as the usage of emotional AI in children's toys and services. In a report entitled "Emotional artificial intelligence in children's toys and devices: Ethics, governance and practical remedies"[16] concerns about the evolution of *human rights infringements* regarding the datafication of childhood, hidden manipulation, increased parental vulnerability, the effect of synthetic personalities on child development were detailed. Even something innocuous, such as an AI spam filter, has been shown to be coded with hidden biases,[17] with biased datasets having a huge impact on our information we see and build our world view on - where those filters obscure and ignore minority and diverse datasets, they can only exacerbate societal division.

From housing allocation[18] to our love lives[19] - the bias endemic in AI creates risks to our human rights. Industry players operating AI systems have neither the human rights perspective nor proper incentive to anticipate the effects their AI will have on humanity if they are left to self regulate, especially if their use of AI is motivated primarily by profit. **It is crucial, then, that the system of risk assessments laid out in the Proposal for "high risk" systems become industry-standard across AI development for all AI systems which could pose risks of harm to the health and safety or a risk of adverse impact on fundamental rights.** Each assessment must be subject to external scrutiny by an appropriately resourced and qualified regulator. We know this is a concern, not just for Avaaz but for many civil society voices, and for a considerable number of Members of the European Parliament ("MEPs") who presented an open letter on the issue to Ursula von der Leyen in March this year.[20]

Bringing transparency and accountability to the black box

20 years ago, social media companies fought off transparency and accountability obligations that could have prevented the rise of disinformation with the argument that they were too small

[15] Wired. Why is TikTok creating filter bubbles based on your race? (2020)
[16] Sage Journals. Emotional artificial intelligence in children's toys and devices: Ethics, governance and practical remedies. (2021)
[17] Springer. Bias in algorithmic filtering and personalization. (2013)
[18] Pivigo. AI in Social Housing: Use Cases. (2021)
[19] For example the dating app Coffee Meets Bagel tended to recommend people of the same ethnicity even to users who did not indicate any preferences. See UX Collective. How to mitigate social bias in dating apps. (2019)
[20] European Parliament. MEP Letter on AI and fundamental rights. (2021)

an industry to cope with the burden of regulation.[21] The European Democracy Action plan and the Digital Services Act ("DSA") now has to wrestle global giants that grew in this "free for all" environment, profiting from attention-driven economics with content acceleration algorithms that have wreaked havoc on our democracies, health and trust in each other. **The lessons of online disinformation cannot be ignored. Transparency must be legislated to allow public oversight** of the risks inherent in any AI system, and **accountability to mitigate those risks should be backed up with regulation rather than left to self-regulation**. As with the DSA, there must be a role for civil society in the identification of risks and the ability to contribute to any industry Code of Conduct.

AI is as only good as its coding and the data that goes in and out of it. If the coding has a bias, intentional or not, these will shape and distort its outcomes. The adage of "garbage in, garbage out" is true of all coding, but the thing that makes AI different, and that requires regulatory intervention, is that it has the capacity to make decisions built on its own learning, machine learning. Humanity, the ethical counterbalance to the risks within any automated data driven system, can be removed from the equation. We believe that this ethical counterbalance should be present across all AI systems, not only those that have been defined as high risk. Too often, Avaaz has seen the effects of automated moderation that misses hate speech against minorities due to inadequate datasets.[22] Without human moderators to assess how the automated systems are functioning on a community level, to spot what is missing, and correct its biases as they emerge, the system continues to learn on its own parameters. Accordingly, Avaaz's other concern in terms of the overall framework is the restriction of the requirement for **human oversight**.

Finally, we are also concerned about the lack of provision relating to the **assessment of and transparency on AI's impact on workers' rights and rights of consultation rights**.

A global approach

*"The EU should aim to act as a norm-setter for AI in a hyper-connected world by adopting an efficient strategy towards its external partners, **fostering its efforts to set global ethical norms for AI at international level in line with safety rules and consumer protection requirements, as well as with European values and fundamental rights**."*[23]

Whilst not specifically responding to the AI Act, we believe that this quote from the Opinion of the Committee on the Internal Market and Consumer Protection ("IMCO") expresses important EU values. However, the current draft Proposal falls short of setting global ethical norms for AI in line with European values and fundamental rights. It fails a basic test of ethics as it does not

[21] And we see this tactic repeating, see Center for Data Innovation. How Much Will the Artificial Intelligence Act Cost Europe? (2021)

[22] Avaaz. Megaphone for Hate. (2019)

[23] European Committee on the Internal Market and Consumer Protection. Opinion on artificial intelligence: questions of interpretation and application of international law in so far as the EU is affected in the areas of civil and military uses and of state authority outside the scope of criminal justice. (2020)

ensure that all AI systems exported from the EU cannot be used in contravention of those standards - irrespective of their intended civil or military use. It creates a loophole that could allow public authorities to circumvent the provisions of the AI Act if they rely on third-countries or international organisations operating high-risk or banned AI systems.

Our amendments

Avaaz has suggested various amendments to ensure the Proposal's basic framework is aligned with European fundamental rights.

1) Measures to extend obligations of risk assessment, transparency and accountability across all AI systems that could pose risks of harm to health and safety or a risk of adverse impact on fundamental rights (Articles 7, 13, 14, 52, and 45);

a) Risk assessment (Article 7)

Ex ante regulatory procedures on AI systems should extend beyond those currently classified as "high risk" in Annex III. We are concerned about the limit the AI Act places on the categories of industry that are deemed as high risk. This would allow services that are not in Annex III, but which do pose significant risks of harm to the health and safety or a risk of adverse impact on fundamental rights, to slip out of any regulatory oversight. As we explained in our introductory remarks, **we do not believe that such a powerful tool as AI can be allowed to grow outside of any public scrutiny.** We argued in our first response that all content distribution through AI content acceleration systems should be deemed high risk, given its role in the dissemination of disinformation. We note the Commission did not adopt this perspective but has continued with a relatively narrow assessment of the risks of AI based on its use in certain industries.

The Proposal itself lays out the framework for such assessment - namely the following criteria in Article 7,2 - but it fails to provide any transparency or accountability over such assessments unless the system appears in Annex III. **We urge the Commission to make Article 7's assessment process compulsory for all new AI systems that could pose risks of harm to health and safety or a risk of adverse impact on fundamental rights- with oversight of those risks assessments by national regulators, as proposed in the AI Act**. Accordingly, all services should assess their AI systems in line with the Proposals framework for high risk services ie:

(a) the intended purpose of the AI system[24];

[24] This assessment should include whether there is a purpose limitation for the AI for the intended use.

(b) the extent to which an AI system has been used or is likely to be used;[25]
(c) the extent to which the use of an AI system **has already caused** harm to health and safety or adverse impact on fundamental rights, or has given rise to significant concerns in relation to the materialisation of such harm or adverse impact, as demonstrated by reports or documented allegations submitted to national competent authorities;
(d) the potential extent of such harm or such adverse impact, in particular in terms of its intensity and its ability to affect a plurality of persons;
(e) the extent to which potentially harmed or adversely impacted persons are dependent on the outcome produced by an AI system, in particular because - for practical or legal reasons - it is not reasonably possible to opt-out from that outcome;
(f) the extent to which potentially harmed or adversely impacted persons are in a vulnerable position in relation to the user of an AI system, in particular due to an imbalance of power, knowledge, economic or social circumstances, or age;
(g) the extent to which the outcome produced with an AI system is easily reversible, whereby outcomes having an impact on the health or safety of persons shall not be considered as easily reversible;
(h) the extent to which existing Union legislation provides for: (i) effective measures of redress in relation to the risks posed by an AI system, with the exclusion of claims for damages; (ii) effective measures to prevent or substantially minimise those risks

These assessments should be openly submitted to the relevant national regulator, in much the same way that Data Protection Impact Assessments ("DPIA's") are made available to national data regulators.

The Proposal also fails to provide a route for comment by either the victims of AI systems, or for civil society to provide context about the environment in which these risks are assessed. Currently, unless an AI service falls into the categories pre identified in Annex III all assessments are hermetically enclosed in the service's development teams - another case of leaving an industry to complete and then mark its own homework.

We would specifically suggest that providers of any AI system that distributes content online through any form of AI-assisted content distribution algorithm should be required to take a risk-based approach by assessing and mitigating the distribution of content that may pose a risk to fundamental rights, public interests, public health, and security.[26] This would require a redraft of Chapter 2, removing where appropriate the restriction relating to risk assessments in Articles 8 and 9 as only required for high risk services. The Proposal should instead be applicable to all AI systems that pose a risk to health and safety, or a risk of adverse impact on fundamental rights.

[25] This assessment should include the extent to which individual data is obtained and used as a basis for the AI to function

[26] This process would bring AI systems used in content distribution into line with Article 26 of the draft DSA to carry out risk assessments at least once a year in relation to the functioning and use of their services. In particular, and in parallel with the provision of the DSA we suggest AI developers are, for example, required to identify systemic risks related to the dissemination of content, any negative effects for the exercise of certain fundamental rights, and the intentional manipulation of their service.

b) Transparency

(i) Amending Article 13 to impose transparency of operation on all AI systems that pose risks of harm to health and safety or a risk of adverse impact on fundamental rights

We noted in our introductory comments that, unlike the DSA, there is no general introduction of transparency into the operation of AI within systems. Transparency of computer operations is becoming a cornerstone of our democracies as our communications systems move into the digital space. We consider this to be a fundamental flaw in the Proposal and suggest that, ideally, Article 13 is amended to remove its restriction on transparency regulation to “high risk” services only. Please also see our specific suggestions for Article 52 below.

(ii) Amending Article 52, 1

Users should have transparency of operations as of right, and it may be particularly important that such transparency is provided in relation to the detection, prevention and investigation of crimes. We see no rationale why the public should not be made aware of the use of AI systems by policing and immigration authorities, and in fact such disclosure seems vital to ensuring the proper balance between citizens’ rights to privacy and to protection from crime.

Article 52 **Transparency obligations for certain AI systems** 1. Providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use. This obligation shall not apply to AI systems authorised by law to detect, prevent, investigate and prosecute criminal offences, unless those systems are available for the public to report a criminal offence.	**Article 52** **Transparency obligations for certain AI systems** 1. Providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an AI system, of the data that has been used in order to generate any decision-making in relation to those natural persons and of the rights and processes to allow natural persons to appeal against the application of such AI to them unless ~~this is obvious from the circumstances and the context of use. This obligation shall not apply to AI systems authorised by law to detect, prevent, investigate and prosecute criminal offences, unless those systems are available for the public to report a criminal offence.~~

(iii) Extending the “right of appeal” to civil society organisations and other external stakeholders when they have a legitimate interest in these decisions (Article 45)

Article 45 of the Proposal establishes that only parties having a legitimate interest can exercise the right of appeal against the decisions taken by the notified bodies that are designed by the competent national authorities with the purpose of providing a third-party conformity assessment on AI systems. As the European Center for Not-For-Profit Law Stichting ("ECNL") has also argued, this article misses an important opportunity to enable civic participation[27]. The current Proposal places business interests above fundamental rights, creating an imbalance between AI systems developers and people subjected to these AI systems.

Avaaz advocates for relevant stakeholder engagement, in particular of marginalised and at-risk groups. Excluding them will exacerbate the inequality that AI systems already create. This right to appeal should be also extended to external stakeholders such as civil society organisations, where they have a legitimate interest for example with expertise on the impact of the AI on communities and or research as to the human rights impacts of the AI systems, as follows:

Article 45	**Article 45**
Member States shall ensure that an appeal procedure against decisions of the notified bodies is available to parties having a legitimate interest in that decision	Member States shall ensure that an appeal procedure against decisions of the notified bodies is available to parties having a legitimate interest in that decision. Civil society organisations should also have the right to appeal when they have a legitimate interest in these decisions.

d) Accountability - Harmonising human oversight for all AI systems (Article 14)

Article 14 of the Proposal establishes human oversight only for high-risk AI systems. The overseeing duties can be extended to all AI systems that pose risks of harm to health and safety or a risk of adverse impact on fundamental rights in order to prevent or minimise such risks. The restriction of this article to high-risk systems should be removed as follows

Article 14	**Article 14**
1. High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use.	1. ~~High-risk~~ AI systems that pose risks to health and safety or fundamental rights shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use.
2. Human oversight shall aim at preventing or minimising the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its	2. Human oversight shall aim at preventing or minimising the risks to health, safety or fundamental rights that may emerge when ~~a~~

[27] ECNL. Position Statement on The EU AI Act. (2021)

intended purpose or under conditions of reasonably foreseeable misuse, in particular when such risks persist notwithstanding the application of other requirements set out in this Chapter. 3. Human oversight shall be ensured through either one or all of the following measures: (a) identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service; (b) identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the user. 4. The measures referred to in paragraph 3 shall enable the individuals to whom human oversight is assigned to do the following, as appropriate to the circumstances: (a) fully understand the capacities and limitations of the high-risk AI system and be able to duly monitor its operation, so that signs of anomalies, dysfunctions and unexpected performance can be detected and addressed as soon as possible; (b) remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system ('automation bias'), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons; (c) be able to correctly interpret the high-risk AI system's output, taking into account in particular the characteristics of the system and the interpretation tools and methods available; (d) be able to decide, in any particular situation, not to use the high-risk AI system or otherwise disregard, override or reverse the output of the high-risk AI system; (e) be able to intervene on the operation of the high-risk AI system or interrupt the system through a "stop" button or a similar procedure. 5. For high-risk AI systems referred to in point 1(a) of Annex III, the measures referred to in paragraph 3 shall be such as to ensure that, in addition, no action or decision is taken by the user on the basis of the identification resulting from the system unless this has been verified and confirmed by at least two natural persons	~~high-risk~~ AI systems that pose risks to health and safety or fundamental rights or AI systems subjected to the transparency obligations ex article 52 are used in accordance with their intended purpose or under conditions of reasonably foreseeable misuse, in particular when such risks persist notwithstanding the application of other requirements set out in this Chapter. 3. Human oversight shall be ensured through either one or all of the following measures: (a) identified and built, when technically feasible, into any ~~high-risk~~ AI systems that pose risks to health and safety or fundamental rights by the provider before it is placed on the market or put into service; (b) identified by the provider before placing the ~~high-risk~~ AI system on the market or putting it into service and that are appropriate to be implemented by the user. 4. The measures referred to in paragraph 3 shall enable the individuals to whom human oversight is assigned to do the following, as appropriate to the circumstances: (a) fully understand the capacities and limitations of the ~~high-risk~~ AI system and be able to duly monitor its operation, so that signs of anomalies, dysfunctions and unexpected performance can be detected and addressed as soon as possible; (b) remain aware of the possible tendency of automatically relying or over-relying on the output produced by an ~~high-risk such~~ AI system ('automation bias'), in particular for ~~high-risk~~ such AI systems used to provide information or recommendations for decisions to be taken by natural persons; (c) be able to correctly interpret the ~~high-risk~~ AI system's output, taking into account in particular the characteristics of the system and the interpretation tools and methods available; (d) be able to decide, in any particular situation, not to use the ~~high-risk~~ AI system or otherwise disregard, override or reverse the output of the ~~high-risk~~ AI system; (e) be able to intervene on the operation of the ~~high-risk~~ AI system or interrupt the system through a "stop" button or a similar procedure. 5. For high-risk AI systems referred to in point 1(a) of Annex III, the measures referred to in paragraph 3 shall be such as to ensure that, in addition, no action or decision is taken by the user on the basis of the identification resulting from the system unless this has been verified and confirmed by at least two natural persons

2) Scope of application: Measures to promote and protect human rights, democracy and the rule of law worldwide

Effective AI regulation in the EU must include exports and international cooperation (Article 2)

We do acknowledge the difficulties imposed by international law when regulating international cooperation and commend the efforts in the Proposal to protect EU Citizens' rights even when providers of AI systems used in the EU are established in a third country. However, we are concerned about two loopholes evident in the Proposal.

(i) Provisions for the export of AI systems

The promotion of human rights worldwide is one of the two main streams of EU human rights policy and action[28]. The EU is "*based on a strong commitment to promoting and protecting human rights, democracy and the rule of law worldwide. Human rights are at the heart of EU relations with other countries and regions*"[29]. This strong commitment cannot be achieved if the EU continues to expose harmful technology to third-countries.

The first loophole therefore we are concerned about is the **lack of provisions addressing the export of AI Systems**. According to Amnesty International[30], a 2020 study in the EU evidenced that "*the current international voluntary due diligence framework is unsatisfactory to significantly change the way businesses manage human rights impacts*".

We recognise that the EU has been making an effort to deal with this scenario, and Avaaz commends all actions in that direction. However, in our opinion, the AI Act falls short in addressing the export of European products when there is a risk they may be used to violate human rights abroad.

Our opinion is largely corroborated by studies and research in the area such as the recent paper Demystifying the Draft EU Artificial Intelligence Act[31]. Its authors analyse the initial draft proposed by the Commission and present their critique that, under the current draft of the AI Act, "*EU vendors can sell biometric systems which would be illegal to use in the EU to oppressive regimes in third countries".*

[28] European Union. Human Rights and Democracy.
[29] European Union. Human Rights and Democracy.
[30] Amnesty International. EU companies selling surveillance tools to China's human rights abusers. (2020)
[31] Demystifying the Draft EU Artificial Intelligence Act published by (i) Michael Veale, from the Faculty of Laws, University College London, United Kingdom; and (ii) Frederik Zuiderveen Borgesius, from the Interdisciplinary Hub for Security, Privacy and Data Governance, Radboud University, The Netherland.

This conclusion affirms an investigation into exports to one territory, by Amnesty International, which identified at least three cases of European companies exporting AI and surveillance technology to China:

- **France (Idemia/Morpho)**: in 2015, Morpho, which specialises in security and identity systems, including facial recognition systems and other biometric identification products, entered into a contract to supply facial recognition equipment directly to the Shanghai Public Security Bureau;
- **The Netherlands (Noldus)**: Noldus sold its "FaceReader" software, which is used for automated analysis of facial expressions that convey anger, happiness, sadness, surprise and disgust, to public security and law enforcement authorities in China. FaceReader was also used in Chinese universities with links to the police, and the Ministry of Public Security. Digital surveillance technology has also been sold by the company to universities in China;
- **Sweden (Axis Communications):** from 2012 to 2019, the company has been listed as a "recommended brand" in Chinese state surveillance tender documents. Axis Communications is specialised in security surveillance and remote monitoring, and has supplied technology to the surveillance programme of Guilin, a city in the south of China, in order to expand it from 8,000 cameras to 30,000.

Amnesty International argues that the companies took insufficient steps to satisfy themselves as to whether sales to China's authorities were of significant risk. In doing so, Amnesty International concludes, those European companies "*totally failed in their human rights responsibilities*."

We agree with these concerns and encourage the Commission to include the export of AI Systems in the scope of application of the AI Act. Article 2 should be amended to include exporters based in the European Union, even if the AI systems they provide are deployed outside the Union.

Additionally we believe the AI Act would be stronger if it incorporated the due diligence obligations set forth in the Regulation (EU) 2021/821 ("Dual Use Regulation")[32], as Article 2,7. As you will notice, the language we suggest is based on the wording of Article 5,2 of the Dual Use Regulation, which establishes due diligence obligations for the export of cyber-surveillance items, establishing:

- Prohibition of EU vendors to export AI systems that are prohibited by the AI Act;
- For all other AI systems, where an exporter is aware, according to its due diligence findings, that AI systems which the exporter proposes to export are intended, in their entirety or in part, for use in connection with internal repression

[32] European Parliament. Regulation (EU) 2021/821 of the European Parliament and of the Council of 20 May 2021 setting up a Union regime for the control of exports, brokering, technical assistance, transit and transfer of dual-use items (recast). (2021)

and/or the commission of serious violations of human rights and international humanitarian law, the exporter shall notify the competent authority;

- That competent authority shall decide whether or not to make the export concerned subject to authorisation. The Commission shall make available guidelines for exporters.

(ii) International cooperation

Our second concern is that the current draft of Article 2,4 creates a potential loophole in which public authorities in the EU could bypass the provisions of the AI Act by relying on third-countries or international organisations operating high-risk or banned AI systems. Our concern about this loophole can be put simply: **no entity, public or private, should be able to accomplish in partnership, with a provider in a third country, what it could not achieve within the EU.**

We are not alone in our opinion as other important organisations have already expressed their worries about the circumvention risks we outlined above. In particular, we draw attention to the Joint Opinion 5/2021 prepared by the European Data Protection Board ("EDPB") and the European Data Protection Supervisor ("EDPS") ("EDPB-EDPS Joint Opinion"). They express their "*(...) serious concerns regarding the exclusion of international law enforcement cooperation from the scope set out in Article 2(4) of the Proposal. This exclusion creates a significant risk of circumvention (e.g., third countries or international organisations operating high-risk applications relied on by public authorities in the EU)".*

This exclusion could severely impact the protection of fundamental rights that the AI Act hopes to safeguard. In this regard, we would like to refer to the IMCO Opinion on artificial intelligence concerning questions of interpretation and application of international law.

EU human rights policy includes "*defending human rights through active partnership with partner countries, international and regional organisations, and groups and associations at all levels of society*"[33]. As evidenced above, the AI Act falls short in ensuring consistency with these values. We encourage the Commission to amend item 2(4) from the final draft of the AI Act to address this international cooperation loophole that poses serious circumvention risks to the scope of application of the AI Acts.

In summary, we would propose to amend the scope of Article 2 to (i) include exporters based in the EU, even if the AI systems they provide are deployed outside the Union; (ii) amend subitem 4 from the final draft of the AI Act to address the international cooperation loophole; and (iii) incorporate due diligence obligations for exporters of AI Systems:

Article 2	**Article 2**

[33] European Union. Human Rights and Democracy.

1.This Regulation applies to: (a) providers placing on the market or putting into service AI systems in the Union, irrespective of whether those providers are established within the Union or in a third country; (b)users of AI systems located within the Union; (c)providers and users of AI systems that are located in a third country, where the output produced by the system is used in the Union; (...) 4. This Regulation shall not apply to public authorities in a third country nor to international organisations falling within the scope of this Regulation pursuant to paragraph 1, where those authorities or organisations use AI systems in the framework of international agreements for law enforcement and judicial cooperation with the Union or with one or more Member States.	1.This Regulation applies to: (a) providers placing on the market or putting into service AI systems in the Union, irrespective of whether those providers are established within the Union or in a third country; (b) users of AI systems located within the Union; (c) providers and users of AI systems that are located in a third country, where the output produced by the system is used in the Union; (...) (e) providers placing on the market or putting into service AI systems in a third country where the provider, distributor or operator of that AI system originates from the Union; (...) 4. This Regulation shall not apply to public authorities in a third country nor to international organisations falling within the scope of this Regulation pursuant to paragraph 1, where those authorities or organisations use AI systems in the framework of international agreements for law enforcement and judicial cooperation with the Union or with one or more Member States, provided, however, that no EU public authority nor any Member State shall obtain, or otherwise make use of, any data originated by an AI system that is prohibited under this Act when operated by any public authorities in a third country or international organisations. The use of data originated by high-risk applications operated by any public authorities in a third country or international organisations is not permitted for EU public authorities or Member States unless safeguards similar to the ones established in this provision for high-risk AI systems are adopted by those operators.[34] (...) 7. Providers may not export AI systems set out in Article 5. For all other AI systems,

[34] We imagine this will happen inthe same way that the EU confirms the adequacy of third country data provisions Add comparison to agreeing inter country data sharing by ensuring reciprocity of data systems

	where an exporter is aware, according to its due diligence findings, that AI Systems which the exporter proposes to export are intended, in their entirety or in part, for use in connection with internal repression and/or the commission of serious violations of human rights and international humanitarian law, the exporter shall notify the competent authority. That competent authority shall decide whether or not to make the export concerned subject to authorisation. The Commission shall make available guidelines for exporters.

3) Measures to enhance workers rights

We acknowledge that, when it comes to workers' rights, employment is part of Annex iii and includes AI intended to be used for making decisions on promotion and termination of work-related contractual relationships, for task allocation and for monitoring and evaluating performance and behavior of persons in such relationships.

However, we are not satisfied that the conformity assessment procedures are sufficient to safeguard workers’ rights in the workplace as an internal conformity assessment (or self-assessment) can be conducted without external oversight. Potential violations would only be discovered at a later stage by overburdened market surveillance authorities when damages have already occurred. Loosely regulating AI practices that potentially infringe on workers’ rights can be very risky, especially if we consider this regulation will apply to all AI developers targeting the EU market, including non-EU entities that might not share EU values.

We therefore propose inclusion of the following additional rights for workers subject to workplace AI surveillance or monitoring systems (by workplace we mean any place in which an employee is contractually engaged in work for its employer so including the home, and any public spaces including transport systems in which the employee conducts their regular work), namely:

i) A legal duty for employers to consult trade unions on the use of high risk and intrusive forms of AI in the workplace;

ii) A legal duty for employers to ensure that workers are aware of the AI systems at the workplace, including their impact on data, digital footprint and work organisation. This could be achieved by extending the transparency and provision of information requirements ex article 13 to employers as well;

iii) A legal right for all workers to have a human review of decisions made by AI systems so they can challenge decisions that are unfair and discriminatory;

iv) An annual conformity assessment for work place based AI to guard against discrimination by algorithm;

v) A legal right to "switch off" from work so workers can have proper downtime in their lives.

Section 2:

Extension of protection for vulnerable groups and Artificial Intelligence practices which should be prohibited

As detailed above, we believe the Commission needs to step back and review its overall framework for AI - but we do agree with the aspects of the Proposal that identify that urgent action that is needed to protect vulnerable groups and ban certain AI applications. Specifically, we believe that the EU should:

a) Protection from the exploitation of vulnerabilities granted to a specific group of persons to include gender, sexual orientation, ethnicity, race, origin, including migrants, refugees and asylum seekers, and religion (Article 5,1(b))

The current wording of Article 5,1(b) prohibits AI systems that "*exploit any of the vulnerabilities of a specific group of persons due to their* ***age, physical or mental disability***".

This protection granted to certain groups is partially aligned with EU values and fundamental rights, which are the basis of the AI Actbut surprisingly, some groups protected by the Charter of Fundamental Rights ("Charter") are not included in the scope of Article 5,1(b). In order to guarantee a greater harmonisation with the Charter, we suggest extending the scope of Article 5,1(b) to also include specific groups of persons due to their gender, sexual orientation, ethnicity, race, origin, including migrants, refugees and asylum seekers, and religion.

Article 5	**Article 5**
1.The following artificial intelligence practices shall be prohibited: (...) (b) the placing on the market, putting into service or use of an AI system that exploits any of the vulnerabilities of a specific group of persons due to their age, physical or mental disability, in order to materially distort the behaviour of a person pertaining to that group in a manner that causes or is likely to cause that person or another person physical or psychological harm;	1.The following artificial intelligence practices shall be prohibited: (...) (b) the placing on the market, putting into service or use of an AI system that exploits any of the vulnerabilities of a specific group of persons due to their age, physical or mental disability, gender, sexual orientation, ethnicity, race, origin (including migrants, refugees and asylum seekers), and religion, in order to materially distort the behaviour of a person pertaining to that group in a manner that causes or is likely to cause that person or another person physical or psychological harm;

b) Extending the ban on certain AI Systems to private actors as well as public authorities (Article 5,1(c),(d),(e))

We believe that some AI systems, which are currently approved by the AI Act under certain circumstances, should be banned in their entirety. Specifically, these systems are:

- social scoring;
- ‘real-time’ remote biometric identification systems;
- systems that categorise individuals from biometrics into clusters according to certain identity criteria; and
- systems that infer emotions of a natural person, except for special situations.

Please find our detailed rationale for each of the proposed bans below.

(i) Social scoring (Article 5,1(c))

The current draft of Article 5,1(c) of the AI Act prohibits the use of AI systems by *public authorities* or on their behalf for the evaluation or classification of the *trustworthiness of natural persons* over a certain period of time based on their social behaviour or known or predicted personal or personality characteristics. It does so where this “social score” could lead to detrimental or unfavourable treatment of either individual natural persons or groups in social contexts which are unrelated to the contexts in which the data was originally generated or collected; or to detrimental or unfavourable treatment of certain natural persons or groups that is **unjustified or disproportionate** to their social behaviour or its gravity.

It is our belief that these restrictions should be adopted by the AI Act in order to **ban all such AI systems used by private actors as well.**

The EDPB-EDPS Joint Opinion is very clear that the use of AI for social scoring can lead to discrimination, stating “*Private companies, notably social media and cloud service providers, can process vast amounts of personal data and conduct social scoring. Consequently, the Proposal should prohibit any type of social scoring*.”

AlgorithmWatch also expressed its concern that “*the prohibition of AI systems used for social scoring purposes is also limited to those deployed by public authorities. Again, private actors are kept out of the line of fire*”.

Furthermore, the Demystifying the Draft EU Artificial Intelligence Act article also has interesting views on social scoring. Initially, the authors point out that the AI Act does not define trustworthiness, and relying on a 2000 paper on Political Trust and Trustworthiness, go on to define trustworthiness as *“a combination of attributes that indicate that an entity will not betray another due to bad faith such as misaligned incentives, lack of care, disregard for promise-keeping (commitment) or through ineptitude at a task (competence)*”.

If this interpretation is correct, the authors conclude that several scoping practices are allowed by the AI Act.

For the reasons explained above, and due to the risks that social scoring may represent to the fundamental rights of those residing in the EU, we urge the Commission to forbid social scoring by public authorities and private actors.

If the Commission disagrees, and decides to maintain the current wording of Article 5,1(c), we encourage it to include a clear definition of “trustworthiness” in the AI Act to prevent misinterpretations of the provision.

Article 5	**Article 5**
1. The following artificial intelligence practices shall be prohibited: (...) (c) the placing on the market, putting into service or use of AI systems by public authorities or on their behalf for the evaluation or classification of the trustworthiness of natural persons over a certain period of time based on their social behaviour or known or predicted personal or personality characteristics, with the social score leading to either or both of the following: (i)detrimental or unfavourable treatment of certain natural persons or whole groups thereof in social contexts which are unrelated to the contexts in which the data was originally generated or collected; (ii)detrimental or unfavourable treatment of certain natural persons or whole groups thereof that is unjustified or disproportionate to their social behaviour or its gravity;	1. The following artificial intelligence practices shall be prohibited: (...) (c) the placing on the market, putting into service or use of AI systems by private actors, public authorities or on their behalf for the evaluation or classification of the trustworthiness of natural persons ~~over a certain period of time~~ based on their social behaviour or known or predicted personal or personality characteristics, with the social score ~~leading to either or both of the following:~~ ~~(i)detrimental or unfavourable treatment of certain natural persons or whole groups thereof in social contexts which are unrelated to the contexts in which the data was originally generated or collected;~~ ~~(ii)detrimental or unfavourable treatment of certain natural persons or whole groups thereof that is unjustified or disproportionate to their social behaviour or its gravity;~~

(ii) ‘Real-time’ remote biometric identification systems in publicly accessible spaces (Article 5,1(d))

Avaaz, like Amnesty International, AlgorithmWatch, and Article 19 believes that the current draft of the AI Act is too weak to effectively protect human rights.

One commonly held concern is that the AI Act has failed to provide for a total ban on ‘real-time’ remote biometric identification systems in publicly or privately accessible spaces.

AlgorithmWatch notes that “*the narrow applicatory scope of this prohibition of real-time biometric identification does not sufficiently consider that the wide-scale use of such systems may not only violate individuals’ fundamental rights but also pave the way for indiscriminate mass surveillance and undermine fundamental principles of democratic societies”.*

According to Amnesty International**,** *“The EU’s proposal falls far short of what is needed to mitigate the vast abuse potential of technologies like facial recognition systems. Under the proposed ban, police will still be able to use non-live facial recognition software with CCTV cameras to track our every move, scraping images from social media accounts without people’s consent”.* Amnesty International also details how the use of real-time facial recognition on people who are suspected of irregularly entering or living in a European Member State, which is currently allowed by the AI Act, will be used as a weapon against migrants and refugees. The organisation also highlights that the AI Act allows predictive policing and the use of AI systems for border control purposes.

Avaaz has noted reports on the sporadic unregulated use of real time surveillance techniques in environments as unexpected as supermarkets - for example the unlawful use of facial recognition systems in Mercadona, one of the leading supermarkets in Spain, which resulted in a €2.5 million fine imposed by the Spanish Agency for Data Protection[35] and the recent investigation in Holland of the indiscriminate use of facial recognition technologies in supermarkets[36].

Amnesty International also voiced itsconcerns about how the AI Act “*does not go far enough in addressing the risks of AI entrenching and exacerbating racism and discrimination*”. In its own words, “*Not only has research shown that facial recognition software is overwhelmingly less accurate with Black and Brown faces, but systemic racism in law enforcement means this technology can disproportionally be used against these communities and can lead to wrongful arrests. What’s more, the safeguards and transparency obligations outlined in the proposal will fail to meaningfully protect the public”.*

The organisation made a public call to the EU to “*to close the many loopholes in this regulation which leave the door open to rampant abuse and discriminatory practices, including banning the use of all facial recognition systems used for mass surveillance”.*

[35] Olive Press. Mercadona gets a €2.5 million fine for installing facial recognition cameras in Supermarkets in Spain. (2021)

[36] EDPB. Dutch DPA issues Formal Warning to a Supermarket for its use of Facial Recognition Technology. (2021)

Amnesty's views are aligned with those of Article 19, which publicly expressed itsdisappointment that the AI Act does not establish a ban on biometric mass surveillance in publicly accessible spaces. We agree with Article 19's view that, *"'real-time' biometric identification systems remain available to police in some circumstances and other types of biometric systems could still be deployed and used for other purposes, such as migration control. These include the assignment of people into categories based on sex, age, eye colour, political or sexual orientation, among other groupings — and these categorisations are not actually prohibited but merely flagged up as high risk"*.

These concerns are also shared by AlgorithmWatch, which provides,"(..) *the prohibition of 'real-time' remote biometric identification systems only applies to systems which are used for law enforcement purposes in publicly accessible spaces, thus neither to systems used by other public authorities nor to those used by private actors. Evidently, the major risks to fundamental rights such systems come with are not limited to law enforcement purposes – a fact which the proposal does not sufficiently reflect."*

AlgorithmWatch explores in detail all the loopholes that authorities could try to exploit, despite the prohibitions in Article 5. As examples, it lists *"the use of real-time biometric identification systems can be allowed for the "prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or of a terrorist attack", the interpretation of which leaves wide discretionary power to the authorities".* The organisation also notes that due to emergency reasons, the judicial authorisations that may be used for such cases can be postponed.

Even the UK's data protection regulator - the Information Commissioner's Office ("ICO") expressed concern over the indiscriminate use of AI assisted facial recognition technology in its June 2021 report "The use of live facial recognition technology ("LFR") in public places."[37] This report cites numerous issues of concern regarding private companies use of facial recognition, including:

- **The automatic collection of biometric data at speed and scale without clear justification;**
- **The lack of control for individuals and communities**. The ICO stated that *"In most of the examples we observed, LFR deployed in public places has involved collecting the public's biometric data without those individuals' choice or control.";*
- **A lack of transparency**. The ICO stated that "*Transparency has been a central issue in all the ICO investigations into the use of LFR in public places. In many cases, transparency measures have been insufficient in terms of the signage displayed, the communications to the public, and the information available in privacy notices*.";
- **The technical effectiveness and statistical accuracy of LFR systems;**
- **The potential for bias and discrimination.** The ICO cited several technical studies that have indicated that LFR works with less precision for some demographic groups,

[37] ICO. The use of live facial recognition technology ("LFR") in public places. (2021)

including women, minority ethnic groups and potentially disabled people stating that "*These issues often arise from design flaws or deficiencies in training data and could lead to bias or discriminatory outcomes. Equally, there is a risk of bias and discrimination in the process of compiling watchlists (often manual) which underpin an LFR system".* All these issues risk infringing the fairness principle within data protection law, as well as raising ethical concerns on AI systems design;

- The governance of LFR escalation processes;
- The processing of children's and vulnerable adults' data. In many of the examples the ICO observed that LFR was deployed in locations likely to be accessed by children and vulnerable adults, such as retail or public transport settings. Data protection law provides additional protections for children and adults because they may be less able to understand the processing and exercise their rights.

Given these concerns, we urge the Commission to adopt a total ban on AI systems that are incompatible with fundamental rights. We understand that this could be seen to fetter innovation, but given the concerns we share with the NGOs and ICO as detailed above, and many other civil society actors, we ask that at minimum the Commission **establish a review** in order to determine if the total ban on AI systems that use **'real-time' remote biometric identification systems in publicly accessible spaces including online spaces should be extended to private actors as well as public authorities.** To illustrate Avaaz's concerns with private actors, we again highlight the unlawful use of facial recognition systems in Mercadona, one of the leading supermarkets in Spain, which resulted in a €2.5 million fine imposed by the Spanish Agency for Data Protection[38].

Article 5	**Article 5**
1. The following artificial intelligence practices shall be prohibited:	1. The following artificial intelligence practices shall be prohibited:
(...)	(...)
(d) the use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement, unless and in as far as such use is strictly necessary for one of the following objectives:	(d) the use of 'real-time' remote biometric identification systems in publicly or privately accessible spaces or online spaces ~~for the purpose of law enforcement, unless and in as far as such use is strictly necessary for one of the following objectives:~~
(i)the targeted search for specific potential victims of crime, including missing children;	~~(i)the targeted search for specific potential victims of crime, including missing children;~~
(ii)the prevention of a specific, substantial and imminent threat to the life or physical safety	~~(ii)the prevention of a specific, substantial and~~

[38] Olive Press. Mercadona gets a €2.5 million fine for installing facial recognition cameras in Supermarkets in Spain. (2021)

of natural persons or of a terrorist attack; (iii)the detection, localisation, identification or prosecution of a perpetrator or suspect of a criminal offence referred to in Article 2(2) of Council Framework Decision 2002/584/JHA 62 and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least three years, as determined by the law of that Member State. 2.The use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement for any of the objectives referred to in paragraph 1 point d) shall take into account the following elements: (a)the nature of the situation giving rise to the possible use, in particular the seriousness, probability and scale of the harm caused in the absence of the use of the system; (b)the consequences of the use of the system for the rights and freedoms of all persons concerned, in particular the seriousness, probability and scale of those consequences. In addition, the use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement for any of the objectives referred to in paragraph 1 point d) shall comply with necessary and proportionate safeguards and conditions in relation to the use, in particular as regards the temporal, geographic and personal limitations. 3.As regards paragraphs 1, point (d) and 2, each individual use for the purpose of law enforcement of a ‘real-time’ remote biometric identification system in publicly accessible spaces shall be subject to a prior authorisation granted by a judicial authority or by an independent administrative authority of the Member State in which the use is to take place, issued upon a reasoned request and in accordance with the detailed rules of national law referred to in paragraph 4. However, in a	~~imminent threat to the life or physical safety of natural persons or of a terrorist attack;~~ ~~(iii)the detection, localisation, identification or prosecution of a perpetrator or suspect of a criminal offence referred to in Article 2(2) of Council Framework Decision 2002/584/JHA 62 and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least three years, as determined by the law of that Member State.~~ ~~2.The use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement for any of the objectives referred to in paragraph 1 point d) shall take into account the following elements:~~ ~~(a)the nature of the situation giving rise to the possible use, in particular the seriousness, probability and scale of the harm caused in the absence of the use of the system;~~ ~~(b)the consequences of the use of the system for the rights and freedoms of all persons concerned, in particular the seriousness, probability and scale of those consequences.~~ ~~In addition, the use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement for any of the objectives referred to in paragraph 1 point d) shall comply with necessary and proportionate safeguards and conditions in relation to the use, in particular as regards the temporal, geographic and personal limitations.~~ ~~3.As regards paragraphs 1, point (d) and 2, each individual use for the purpose of law enforcement of a ‘real-time’ remote biometric identification system in publicly accessible spaces shall be subject to a prior authorisation granted by a judicial authority or by an independent administrative authority of the Member State in which the use is to take place, issued upon a reasoned request and in accordance with the detailed rules of national~~

duly justified situation of urgency, the use of the system may be commenced without an authorisation and the authorisation may be requested only during or after the use. The competent judicial or administrative authority shall only grant the authorisation where it is satisfied, based on objective evidence or clear indications presented to it, that the use of the ‘real-time’ remote biometric identification system at issue is necessary for and proportionate to achieving one of the objectives specified in paragraph 1, point (d), as identified in the request. In deciding on the request, the competent judicial or administrative authority shall take into account the elements referred to in paragraph 2. 4.A Member State may decide to provide for the possibility to fully or partially authorise the use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement within the limits and under the conditions listed in paragraphs 1, point (d), 2 and 3. That Member State shall lay down in its national law the necessary detailed rules for the request, issuance and exercise of, as well as supervision relating to, the authorisations referred to in paragraph 3. Those rules shall also specify in respect of which of the objectives listed in paragraph 1, point (d), including which of the criminal offences referred to in point (iii) thereof, the competent authorities may be authorised to use those systems for the purpose of law enforcement.	~~law referred to in paragraph 4. However, in a duly justified situation of urgency, the use of the system may be commenced without an authorisation and the authorisation may be requested only during or after the use.~~ ~~The competent judicial or administrative authority shall only grant the authorisation where it is satisfied, based on objective evidence or clear indications presented to it, that the use of the ‘real-time’ remote biometric identification system at issue is necessary for and proportionate to achieving one of the objectives specified in paragraph 1, point (d), as identified in the request. In deciding on the request, the competent judicial or administrative authority shall take into account the elements referred to in paragraph 2.~~ ~~4.A Member State may decide to provide for the possibility to fully or partially authorise the use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement within the limits and under the conditions listed in paragraphs 1, point (d), 2 and 3. That Member State shall lay down in its national law the necessary detailed rules for the request, issuance and exercise of, as well as supervision relating to, the authorisations referred to in paragraph 3. Those rules shall also specify in respect of which of the objectives listed in paragraph 1, point (d), including which of the criminal offences referred to in point (iii) thereof, the competent authorities may be authorised to use those systems for the purpose of law enforcement.~~

(iii) ban the use of AI systems categorizing individuals from biometrics into clusters according to ethnicity, gender, as well as political or sexual orientation, or other grounds for discrimination (Article 5,1(e))

One of the aims of the AI Act is to protect fundamental rights. In order to effectively achieve this purpose, we suggest the inclusion of item (e) under Article 5,1, in order to expressly set forth a ban on the use of AI systems categorizing individuals from biometrics into clusters according to ethnicity, gender, as well as political or sexual orientation, or other grounds for discrimination under Article 21 of the Charter.

This opinion is also shared by other organisations. Notably, the EDPB and the EDPS, in their EDPB-EDPS Joint Opinion expressly voiced their concerns and pushed for "*a ban, for both public authorities and private entities, on AI systems categorizing individuals from biometrics (for instance, from face recognition) into clusters according to ethnicity, gender, as well as political or sexual orientation, or other grounds for discrimination prohibited under Article 21 of the Charter, or AI systems whose scientific validity is not proven or which are in direct conflict with essential values of the EU (e.g., polygraph, Annex III, 6. (b) and 7. (a))*."

We urge the Commission to take these suggestions into consideration and ban the clustering of individuals into aspects of their identity that can represent grounds for discrimination.

Article 5	**Article 5**
1. The following artificial intelligence practices shall be prohibited: (...)	1. The following artificial intelligence practices shall be prohibited: (...) (e) the use of AI systems categorizing individuals from biometrics into clusters according to ethnicity, gender, as well as political or sexual orientation, or other grounds for discrimination.

(iv) AI systems that infer emotions of a natural person, except for health or research purposes or other exceptional purposes, and subject to full regulatory review and with full and informed consent at all times (Article 5,1 and 52,2)

The current draft of the AI Act does not forbid the use of AI systems to infer emotions of a natural person, although it does establish under article 52,2 some transparency obligations towards users of those systems. We believe this is insufficient if the AI Act truly means to protect fundamental rights, and that the AI Act should establish a total ban on these emotion inferring systems, except for in certain specific circumstances, such as for health or research purposes and be subject to full and informed consent at all times.

Our position is also aligned with other organisations.

Article 19 also expressed its concerns about these AI systems, remarking that "*Worryingly, emotion recognition is also not prohibited, despite its discredited scientific basis and fundamental inconsistency with human rights. Instead, it is subject to weak 'transparency obligations' that are largely ineffective for protecting human rights*".

AlgorithmWatch holds the same view and draws attention to the fact that these systems "*are likely to come with a high potential of severe harm on individuals and democratic societies*". Moreover, the organisation stresses that "*the scientific basis of especially emotional recognition systems is highly disputed.*"

Another point that is important to stress is that emotion-recognition systems are evidenced biased against minorities, specially against Black people. The findings of a study[39] showed that "*facial recognition software interprets emotions differently based on the person's race*". Based on the comparison of the emotional analysis of two different facial recognition services, Face and Microsoft's Face API, the study found that "*services interpret black players as having more negative emotions than white players; however, there are two different mechanisms. Face consistently interprets black players as angrier than white players, even controlling for their degree of smiling.*"

It is not hard to find other examples showing how emotion-recognition software is biased against minorities. Google, for instance, tagged Black people as gorillas[40]. Facial recognition programs have misidentified gender in 35 percent of darker-skinned females[41].

However, the use of AI to infer emotions should be allowed for very specific reasons, such as health, for example. The EDPB and the EDPS, in their EDPB-EDPS Joint Opinion express their views "*that the use of AI to infer emotions of a natural person is highly undesirable and should be prohibited, except for certain well- specified use-cases, namely for health or research purposes (e.g., patients where emotion recognition is important), always with appropriate safeguards in place and of course, subject to all other data protection conditions and limits including purpose limitation*".

We agree with their views on health purposes. Studies are being conducted, for instance, to determine how emotion recognition technologies can be used to teach children with autism spectrum disorder to identify and express emotions[42]. Similarly, research is being carried out to investigate the use of EEG-based brain-computer interface systems for emotion recognition in patients with disorders of consciousness, with promising results[43].

For these reasons, we urge the Commission to consider the ban of the use of emotional recognition systems, except for in specific circumstances such as for health or research purposes, and subject to full and informed consent at all times.

[39] Lauren Rhue. Racial Influence on Automated Perceptions of Emotions. (2018)
[40] The Wall Street Journal. Google Mistakenly Tags Black People as 'Gorillas,' Showing Limits of Algorithms. (2015)
[41] The New York Times. Facial Recognition Is Accurate, if You're a White Guy. (2018)
[42] Springer. Using emotion recognition technologies to teach children with autism spectrum disorder how to identify and express emotions. (2021)
[43] Emotion-Related Consciousness Detection in Patients With Disorders of Consciousness Through an EEG-Based BCI System. (2018)

Article 5,1 1. The following artificial intelligence practices shall be prohibited: (...)	**Article 5,1** 1. The following artificial intelligence practices shall be prohibited: (...) (f) the placing on the market, putting into service or use of AI systems to infer emotions of a natural person, except for health or research purposes or other exceptional purposes, and subject to full regulatory review and with full and informed consent at all times.

Should the Commission not accept these arguments we consider it essential that citizens are informed about the use of these AI systems. Accordingly, we propose the following alternative changes to **Article 52** in order to introduce greater transparency to users of AI in the event the ban is not accepted by the Commission.

Article 52 (...) 2. Users of an emotion recognition system or a biometric categorisation system shall inform of the operation of the system the natural persons exposed thereto. This obligation shall not apply to AI systems used for biometric categorisation, which are permitted by law to detect, prevent and investigate criminal offences. 3. Users of an AI system that generates or manipulates image, audio or video content that appreciably resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful ('deep fake'), shall disclose that the content has been artificially generated or manipulated. However, the first subparagraph shall not apply where the use is authorised by law to detect, prevent, investigate and prosecute criminal offences or it is necessary for the exercise of the right to freedom of expression and the right to freedom of the arts and sciences guaranteed in the Charter of Fundamental Rights of the EU, and subject to appropriate safeguards for the rights and freedoms of third parties. 4. Paragraphs 1, 2 and 3 shall not affect the requirements and obligations set out in Title III of this Regulation.	**Article 52** (...) 2. Users of an emotion recognition system or a biometric categorisation system shall inform of the operation of the system the natural persons exposed thereto. ~~This obligation shall not apply to AI systems used for biometric categorisation, which are permitted by law to detect, prevent and investigate criminal offences.~~ 3. Users of an AI system that generates or manipulates image, audio or video content that appreciably resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful ('deep fake'), shall disclose that the content has been artificially generated or manipulated. However, the first subparagraph shall not apply where the use is ~~authorised by law to detect, prevent, investigate and prosecute criminal offences or it is~~ necessary for the exercise of the right to freedom of expression and the right to freedom of the arts and sciences guaranteed in the Charter of Fundamental Rights of the EU, and subject to appropriate safeguards for the rights and freedoms of third parties. 4. Users of any AI system that generates or distributes content, whether on a commercial or non commercial basis shall be fully informed of

	the data used and criteria adopted for that content's delivery. 5. Paragraphs 1, 2 ~~and~~ 3, and 4 shall not affect the requirements and obligations set out in Title III of this Regulation.

In conclusion

We look forward to discussing these comments and amendments and urge the Commission, Parliament and Council to reconsider its overall framework in addition to these specific issues. If "*On artificial intelligence, trust is a must, not a nice-to-have*"[44] we must acknowledge that this law comes at a time when there is no standardised trusted AI assurance ecosystem[45]. If the current AI Proposal cherry picks a few egregious harms, and in the name of innovation allows non-human centered AI decision making to run rings around our fundamental human rights without any oversight, it will fail before it begins.

Please contact us in the event of any queries at:

Sarah Andrew at sarah.andrew@avaaz.org;
Ana Paula Rodrigues at anapaula@avaaz.org
and
Luana Lo Piccolo at luana@support.avaaz.org

[44] Margrethe Vestager, Executive Vice President of the European Commission for A Europe Fit for the Digital Age, for BBC. EU artificial intelligence rules will ban 'unacceptable' use. (2021)
[45] Centre for Data Ethics and Innovation Blog. The European Commission's Artificial Intelligence Act highlights the need for an effective AI assurance ecosystem. (2021)