Access Now's submission to the European Commission's adoption consultation on the Artificial Intelligence Act

August 2021

TABLE OF CONTENTS

Executive summary

Access Now welcomes the European Commission's pioneering proposal for a regulatory framework for artificial intelligence. We have consistently pointed to the insufficiency of ethics guidelines and self-regulatory approaches, and have long called for regulatory intervention in the field of AI. The current Proposal provides a workable framework to ensure the protection of fundamental rights, but requires significant modifications in a number of areas or it risks failing to achieve the objective of protecting fundamental rights.

Broadly speaking, the main aim of the Proposal is to regulate so-called high-risk AI systems, namely those which pose a risk of harm to health and safety or a risk of adverse impact on fundamental rights. Title III of the Proposal "contains specific rules for AI systems that create a high risk to the health and safety or fundamental rights of natural persons," and a list of high risk AI systems is provided in Annex III. The Proposal also contains provisions to prohibit certain AI practices in Article 5, which the Explanatory Memorandum says "comprises all those AI systems whose use is considered unacceptable as contravening Union values."[1] Additionally, the Proposal contains provisions for transparency obligations for certain AI systems such as chatbots and emotion recognition systems. As we will argue below, a number of shortcomings of the Proposal prevent it from adequately achieving the aim of protecting fundamental rights.

As we explain below, the definitions of emotion recognition and biometric categorisation contain certain technical flaws which must be addressed. Without defining these applications of AI correctly, we cannot hope to adequately address them in regulation. Section II therefore proposes alternative definitions of both of these terms, as well as a discussion of the severe risks they pose to fundamental rights. Having defined these two terms correctly, we then argue in Section III that they should both be prohibited.

Section III begins with a discussion of the four existing prohibitions in Article 5 of the Proposal. We point out flaws in the formulations of the prohibitions in Article 5, paragraphs 1(a), 1(b), and 1(c), and propose alternative formulations. We then provide justifications and proposed formulations for a number of additional prohibitions. Finally, we argue that Article 5 must be supplemented with a list of criteria to define 'unacceptable risk' so that other applications of AI can be added to the list of prohibited practices if evidence emerges that they pose unacceptable risks, and that a mechanism must be added to Article 7 to allow for additional practices to be added to Article 5.

In Section IV we address the relative lack of obligations placed on 'users' of AI systems as compared to those placed upon 'providers.' We suggest that additional obligations be placed upon users, including mandating that some form of impact assessment is carried out for all high-risk AI systems. In Section V,

[1] Proposal for a Regulation Of The European Parliament And Of The Council - Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts (hereafter the Proposal), p. 12

we discuss the need to extend the scope of the publicly viewable database of high risk AI systems proposed in Article 60.

Section VI raises a number of concerns about the current proposal for AI regulatory sandboxes in Articles 53 and 54, and makes recommendations for a number of modifications, including that the use cases under paragraph (i), namely those related to law enforcement applications of AI, be removed from Article 54. Section VIII addresses a number of gaps in the enforcement and redress mechanisms provided by the proposal and, finally, in Section VIII we note a number of additional concerns which need to be addressed in the Proposal.

While the current Proposal provides a workable framework for regulating harmful applications of AI, it requires serious modifications in a number of areas. We look forward to working with the co-legislators in the coming months to ensure that these issues are addressed.

I. Introduction

Access Now 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.

Access Now's European Policy Manager, Fanny Hidvegi, as a member of the European Commission's High Level Expert Group on Artificial Intelligence, advises the EU on its strategy for the development of AI.[2]

With the increasing investment in and proliferation of automation-based technologies, the EU must enforce and develop the highest human rights standards for artificial intelligence systems that are designed, developed, or deployed in the European Union.

Access Now has put forward concrete policy objectives for EU lawmakers on artificial intelligence[3] in our European Human Rights Agenda in the Digital Age.[4] We also provided recommendations in response to the European Commission's consultation on the "White Paper on Artificial Intelligence — a European approach to excellence and trust."[5]

In response to the Commission's consultation on the *Proposal for a Regulation Of The European Parliament And Of The Council - Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts* (hereafter the Proposal), Access Now will submit this document outlining six key issues with the current proposal. This is not intended to be a comprehensive or definitive statement of all our comments on the Proposal, but rather intends to highlight some of the key issues for discussion in the next stage of the legislative process and to offer some preliminary proposals. Those six issues are:

1. The inadequate treatment of biometric categorisation, emotion recognition, and 'AI polygraph' applications;
2. The inadequacy of the current prohibitions under Article 5;
3. The relative lack of obligations on 'users' of AI systems;
4. Extending the scope of Article 60's 'EU database for stand-alone high-risk AI systems'
5. Issues with the proposed regulatory sandboxes
6. Gaps in enforcement and redress

Finally, we include a section to highlight a number of additional issues.

[2] https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

[3] https://www.accessnow.org/cms/assets/uploads/2020/02/European-Human-Rights-Agenda-for-the-Digital-Age-Recommendations-on-Artificial-Intelligence.pdf

[4] https://www.accessnow.org/access-now-presents-the-european-human-rights-agenda-for-the-digital-age/

[5] https://www.accessnow.org/cms/assets/uploads/2020/06/EU-white-paper-consultation_Access_Now_June2020.pdf

Access Now also welcomes the joint opinion on the Proposal from the European Data Protection Supervisor (EDPS) and the European Data Protection Board (EDPB).[6] As we note throughout this submission, we largely support the recommendations made in the joint opinion which we believe are essential to ensure that the Proposal is effective in safeguarding fundamental rights.

In the coming months, Access Now will continue to engage with policymakers on the issues raised by the Proposal, and we look forward to providing further analysis on the issues raised here and on other issues in the Proposal.

II. Address the inadequate definitions of emotion recognition & biometric categorisation in Articles 3 & 52

Definitions

In Section III, we will discuss the need to modify a number of Article 5's existing prohibitions and to add a number of additional prohibitions. Before discussing these prohibitions, however, we must address some issues regarding the definition and treatment of emotion recognition and biometric categorisation. These two applications of AI are defined in Article 3, paragraphs (34) and (35) respectively. Furthermore, Article 52 outlines "harmonised transparency rules for AI systems intended to interact with natural persons, **emotion recognition systems and biometric categorisation systems**, and AI systems used to generate or manipulate image, audio or video content" (p.38).

As we will show in this section, both their definitions in Article 3 and their treatment in Article 52 are problematic. In this section we will discuss the need to modify their definitions, and in Section III we will argue that both of them need to be prohibited. As should be clear, to prohibit something in an efficient manner, it must first be well defined.

In essence, the transparency rules in Article 52 require that people are informed when interacting with, or being subjected to, such systems. There are, however, numerous limitations to these transparency rules, such as an exception for when a system is used in a law enforcement context. But even when they apply, these rules are insufficient to protect people's rights.

We will focus our attention here on the treatment of emotion recognition systems and biometric categorisation systems in Article 52, although much could also be said about the other applications, namely chatbots and 'deepfakes.' Following the logic of the Proposal, we should assume that the transparency obligations laid out in Article 52 are intended to be sufficient to mitigate the risks posed by the systems in question.

[6] https://edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-52021-proposal_en

In the case of chatbots and deepfakes, one can assume that the risk identified is that people would believe themselves to be interacting with a real human being (as opposed to a chatbot) or a real piece of content (as opposed to a piece of synthetic media). In such cases, the obligation to notify people seems to make sense, although in many cases this transparency obligation will be insufficient to mitigate all risks associated with such applications.

In the case of chatbots, we have already seen how an experimental use of OpenAI's large language model, GPT-3, in a medical setting led to the chatbot suggesting that a fictional patient should kill themselves.[7] With deepfakes, there are now numerous forums, apps, and social media groups where people can request deepfake pornography videos of women. As has been well-documented, labelling such a non-consensual, synthetic video as a deepfake does not eliminate the damage it causes; women who have been victims of these non-consensual images and videos have reported how their reputations have been damaged nevertheless and how they no longer feel comfortable using social media.[8] In both cases, a transparency obligation makes some sense, but clearly does not fully address the damage caused.

In the cases of emotion recognition and biometric categorisation, however, it is difficult to see what risk the transparency obligation is intended to mitigate, as the harms caused by these systems are not primarily due to people being unaware of their operation. To understand why, let's look at how these two applications of AI work, starting with how they are defined in Article 3:

(34) 'emotion recognition system' means an AI system for the purpose of identifying or inferring emotions or intentions of natural persons on the basis of their biometric data;

(35) 'biometric categorisation system' means an AI system for the purpose of assigning natural persons to specific categories, such as sex, age, hair colour, eye colour, tattoos, ethnic origin or sexual or political orientation, on the basis of their biometric data;

First of all, it is important to point out that there are potential flaws in these definitions. In both cases, the use of the term 'biometric data' is problematic. According to Art. 3 (33) of the Proposal (which is identical to Art. 4 (14) of the General Data Protection Regulation), biometric data is defined as "personal data resulting from specific technical processing relating to the **physical, physiological or behavioural characteristics** of a natural person, **which allow or confirm the unique identification of that natural person**, such as facial images or dactyloscopic data."

There are two essential components of this definition: firstly, biometric data relates to the physical, physiological or behavioral characteristics of a person and secondly, it allows the unique identification of that person. This means that personal data relating to the physical, physiological or behavioural characteristics of a natural person which **does not allow for unique identification *is not biometric data***, according to this definition. This is of the utmost importance here because both emotion

[7] https://www.nabla.com/blog/gpt-3/
[8] https://www.huffpost.com/entry/deepfake-porn-heres-what-its-like-to-see-yourself_n_5d0d0faee4b0a3941861fced

recognition and 'biometric categorisation' can be performed using physical, physiological or behavioral data which arguably do not meet the high bar for identification required to be classified as biometric data.

Both emotion recognition and biometric categorisation can be performed using physiological and behavioural data such as gait,[9] heart rate,[10] galvanic skin response,[11] none of which can necessarily be used to uniquely identify a person in all circumstances, unlike a facial template, for example. The danger with tying the definition of these applications to the high bar of biometric data is that the Proposal thereby risks creating a loophole for applications of emotion recognition and biometric categorisation that use data which does not meet the bar of unique identification. Moreover, emotion recognition may be used on groups of people, and not only on individuals, so we recommend changing the definition to reflect this.

In the case of emotion recognition, we would thus recommend modifying the definition in Article 3, (34) to the following:

(34) 'emotion recognition system' means an AI system for the purpose of identifying or inferring emotions or intentions of ~~natural persons~~ ***individuals or groups*** *on the basis of ~~their biometric data~~* ***data relating to their physical, physiological or behavioural characteristics****;*

In the case of biometric categorisation, there are a number of other issues with the definition which still need to be discussed. Let us look at these before proposing a new definition.

Biometric categorisation

According to the definition in Art. 3 (35), biometric categorisation systems work by assigning people to categories based on their biometric data. As discussed above, this limitation to the high bar of biometric data is problematic, so **we can more correctly say that such systems data assign people to categories based on their physical, physiological or behavioural characteristics**. In some cases, this is technically possible. An AI system could, for example, be used to analyse photos to group people according to hair or eye colour. This is possible because hair or eye colour is something an AI system can reasonably infer from physiological data in the form of images of people's hair and eyes.

However, the definition in Art. 3 (35) includes a number of other categories which **cannot and should not be inferred from physical, physiological or behavioural data**, including: sex, ethnic origin and sexual or political orientation. The claim that it is possible to infer any of these attributes from physiological or behavioural data is hugely problematic. A brief overview of the problems involved in

[9] https://arxiv.org/abs/2003.11461
[10] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7038485/
[11] https://imotions.com/blog/gsr/

such inferences should suffice to understand some of the reasons why this processing of data is not legitimate.

The claim that machine learning systems can infer sexual or political orientation from physiological or behavioural data goes back to the work of the controversial psychometrics enthusiast and business school professor, Michal Kosinski. Kosinski's work is, in turn, based on outdated, racist, and eugenicist theories of phrenology and physiognomy which attempted to use physiological data, such as length of nose or shape of skull, to sort people into social hierarchies and groups.

Kosinski's original infamy stems from the use of his work on psychometrics by Cambridge Analytica,[12] but he has continued to publish flawed, dangerous research since then, most notoriously his papers on inferring sexual orientation from facial images (the 'AI Gaydar') and on inferring political orientation from facial images.

One can only assume that the definition in Art.3 (35) of the Proposal is a reference to Kosinki's work, as his is some of the only published work to make such ridiculous and problematic claims. Both of Kosinki's papers (*Deep neural networks are more accurate than humans at detecting sexual orientation from facial images*[13] and *Facial recognition technology can expose political orientation from naturalistic facial images*[14]) have been roundly and devastatingly criticised by scholars in the relevant fields, and are not accepted as serious, not to mention ethical, pieces of research in machine learning (on the AI Gaydar paper, see critiques by Phil Cohen,[15] Greggor Mattson,[16] Shreeharsh Kelkar,[17] and Andrew Gelman,[18] and on the second paper, see this piece by Juan Pablo Pardo-Guerra[19]).

Regardless of whether Kosinski's research is flawed, however, the idea that complex attributes of our personalities and identities can be inferred from, and are therefore determined by, our physiology, is in fundamental conflict with inherent human dignity, and our rights to freedom of thought and opinion. **The claim that 'political orientation', which is already a vague and contentious concept, can be reliably inferred from physiological data is founded on an idea of biological determinism that is in direct and irremediable conflict with the essence of fundamental rights. I can change my political views, but I cannot change my face**.

Work such as Kosinksi's implies that political orientation, along with other complex human attributes, are not subject to modification, but are determined by our biology. **To claim that my facial structure, my gait, or any other piece of physiological data, can be used to determine my voting preferences is thus to deny me the capacity to think freely or form opinions, and attacks the essence of**

[12] https://www.theguardian.com/technology/2018/mar/17/facebook-cambridge-analytica-kogan-data-algorithm
[13] https://psyarxiv.com/hv28a/
[14] https://www.nature.com/articles/s41598-020-79310-1
[15] https://familyinequality.wordpress.com/2017/09/11/on-artificially-intelligent-gaydar/
[16] https://sociologicalscience.com/articles-v5-12-270/
[17] https://scatter.wordpress.com/2017/09/25/guest-post-how-might-the-history-of-ai-help-us-critique-kosinskys-gaydar-study/
[18] https://statmodeling.stat.columbia.edu/2017/09/11/god-goons-gays-3-quick-takes/
[19] https://scatter.wordpress.com/2021/01/19/bad-science-computational-imperialism-and-the-economy-of-attention/

human dignity. Informing a person that they have interacted with a system that inferred their political orientation from their gait does nothing to mitigate the harm caused in this case. **Such systems should simply be prohibited, a point which we will return to in Section III**. For now, let us concentrate on the problem of the definition.

Any definition of biometric categorisation in the Proposal should be strictly limited to grouping people according to physical or physiological categories (including things such as gait or keystrokes), and not include inferences about personality, social grouping, or indeed any aspects of person which are not straightforwardly physiological. Even if there may be some link between a given attribute and physiological characteristics, it should only fall under the definition of biometric categorisation if it is solely determined by such data.

The following revised definition of biometric categorisation is therefore proposed:

(35) 'biometric categorisation system' means an AI system ***that uses data relating to the physical, physiological or behavioural characteristics of a natural person*** *for the purpose of assigning natural persons to specific categories ~~such as sex, age, hair colour, eye colour, tattoos, ethnic origin or sexual or political orientation, on the basis of their biometric data~~ which can be reasonably inferred from such data;*

It is important to note that the suggested modification of this definition is merely for the purpose of technical and legal correctness. The current definition in the Proposal contains technical and legal errors and ambiguities which make it unusable. In correcting these errors, as suggested here, we do not suggest that the risks of biometric categorisation, when defined properly, can be mitigated by the transparency obligations in Art. 52. Indeed, in Section III we will argue that biometric categorisation should be prohibited in publicly accessible spaces, and that any attempt to infer things such as political orientation from these types of data must be prohibited in all circumstances.

Emotion recognition and AI polygraphs

Returning to the problem of emotion recognition, the question we asked above was whether the transparency obligation in Art. 52 is sufficient to address the risks posed by emotion recognition systems. It is important to point out that two uses of emotion recognition are also classified as high risk in the Proposal: Annex III 6(b) refers to "AI systems intended to be used by law enforcement authorities as polygraphs and similar tools or to detect the emotional state of a natural person"; and Annex III 7 refers to "AI systems intended to be used by competent public authorities as polygraphs and similar tools or to detect the emotional state of a natural person" in the context of migration, asylum and border control management. The Proposal therefore only lists these two use cases as high risk, suggesting that the risks posed by all other uses of emotion recognition can be mitigated with the transparency obligation as per Art. 52.

Unfortunately, this is far from the case. As we will argue in Section III, not only should the two 'high risk' uses cases outlined above **be prohibited**, but there is a strong case to be made that **all applications of emotion recognition should be prohibited**, an opinion which is shared by the EDPS and EDPB in their joint opinion on the Proposal. To understand why, let's look in more detail at what emotion recognition is.

Broadly, the term 'emotion recognition' covers a range of technologies that attempt to infer someone's emotional state, or intentions, from data collected about that person. This can include using Natural Language Processing (NLP) to analyse text to infer sentiment (often called sentiment analysis), but also more obviously invasive techniques such as using images of a person's face, recordings of their voice, or even more fine-grained physiological and behavioral data from wearable devices to make inferences about emotional states.

Many of the 'face-based' emotion recognition applications typically use some form of Paul Eckman's 'basic emotions' theory, which posits a set of 'universal categories' of human emotion and describes how these can be read from facial configurations.[20] However, emotion recognition systems also attempt to infer emotion from other physiological or behavioural data, such as voice or gait, and therefore go beyond detecting something like Eckman's list of basic emotions, to include applications such as "Artificial Intelligence Polygraphs" that claim to detect deception or other intentions.

One of the main issues with understanding the impact of emotion recognition systems on fundamental rights is that there are serious doubts about whether current systems, and even future systems, can actually do what they claim. For instance, Lisa Feldman Barret and her co-authors carried out a meta-study to assess the evidence for inferring emotion from facial configurations and concluded that despite"[t]echnology companies [...] investing tremendous resources to figure out how to objectively "read" emotions in people by detecting their presumed facial expressions [...] the science of emotion is ill-equipped to support any of these initiatives."[21] Similar concerns apply to emotion recognition systems that use other physiological or behavioural data, such as voice, although no such comprehensive study has been carried out compared to face-based approaches.

A discrepancy therefore exists between the claims coming from those marketing emotion recognition systems, or even researchers developing them, and what the actual systems can do. If emotion recognition systems cannot actually detect emotion, or deception and related 'inner states', but simply make flawed or inaccurate inferences about our emotions, the question of how they impact, and potentially violate, fundamental rights is made complex.

[20] https://journals.sagepub.com/doi/10.1177/1754073911410740

[21] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella et al., *Emotional Expressions Reconsidered: Challenges to Inferring Emotion from Human Facial Movements*, Journal of Psychological Science in the Public Interest (2019), Vol 20, Issue 1: 1-68, DOI: https://doi.org/10.1177/1529100619832930 at page 48.

One way to answer this question is to point to the fact that regardless of whether AI systems can actually infer our emotional state, the fact that they claim to do so already has an impact on our rights. If security forces use emotion recognition to detect potentially aggressive people[22] in crowds or at protests proactively apprehend these people before they have committed any aggressive act, it does not matter whether the inference was flawed or not; the consequences are real. It is also highly likely that such systems will lead to discriminatory impacts on already marginalised and racialised groups. Indeed, a study entitled 'Racial Influence on Automated Perceptions of Emotions' has shown that emotion recognition systems assign more negative emotions to Black people.[23] As Lauren Rhue, the author of the aforementioned study, has noted, the widespread use of emotion recognition "could formalize preexisting stereotypes into algorithms, automatically embedding them into everyday life."[24]

Moreover, if people believe that such systems are in operation, whether it's in workplaces or public spaces, they will feel pressure to modify their behaviour to be positively evaluated by these systems, especially if rewards or punishments are linked to certain emotional expressions. See, for example, how Canon put cameras in its office that only allow smiling workers to enter certain rooms.[25] If the logic of examples like this are extended, we could have employers demand certain levels of overall "happiness" throughout the workday in order to get bonuses, or workers being punished for being in a "bad mood" as evaluated by these systems.

Civil society organisations such as Access Now, Article19 and the AI Now Institute have long pointed to how emotion recognition systems violate a range of human rights, including the right to privacy, right to freedom of expression, right to protest, right against self-incrimination, and the right to equality and non-discrimination. Academics such as Susie Alegre have also pointed to how emotion recognition systems violate our rights to freedom of thought and opinion which are, importantly, absolute rights that admit of no interference.[26] As such, it should be clear that mere transparency obligations are far from sufficient to mitigate the risks posed by these systems. **Indeed, as we will argue in the next section, emotion recognition, AI polygraphs, along with many of the forms of so-called 'biometric categorisation', must be banned in order to safeguard fundamental rights.**

III. Address the issues with the prohibitions under Article 5

In our response to the Commission's 2020 consultation[27] on the "White Paper on Artificial Intelligence — a European approach to excellence and trust," Access Now pointed to the potential limitations of a risk-based approach to regulating AI.[28] We explained that a risk-based approach would not account for

[22] https://thenextweb.com/news/british-police-to-trial-facial-recognition-system-that-detects-your-mood
[23] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3281765
[24] https://theconversation.com/emotion-reading-tech-fails-the-racial-bias-test-108404
[25] The Verge. *Canon put AI cameras in its Chinese offices that only let smiling workers inside*, available online: https://www.theverge.com/2021/6/17/22538160/ai-camera-smile-recognition-office-workers-china-canon, 17 July 2021.
[26] https://link.springer.com/article/10.1007/s12027-020-00633-7
[27] https://www.accessnow.org/trust-and-excellence-the-eu-is-missing-the-mark-again-on-ai-and-human-rights/
[28] https://www.accessnow.org/eu-regulation-ai-risk-based-approach/

the fact that not all risks can be sufficiently mitigated and that, therefore, certain uses of AI systems would have to be prohibited.

Thankfully, in response to concerns raised by a coalition of civil society organisations,[29] including Access Now, and supported by 116 Members of the European Parliament,[30] the Proposal does include provisions, under Article 5, to prohibit certain uses of AI. In line with calls for red lines on certain uses, Article 5 of the Proposal outlines a list of applications of AI that are to be prohibited because their "use is considered unacceptable as contravening Union values, for instance by violating fundamental rights."

While it is an important step for the Proposal to acknowledge the need for prohibitions on certain applications, there are a number of serious flaws with Article 5 in its current form:

1. The current language used to describe the first three prohibited practices (Art. 5, paragraph 1(a), 1(b), and 1(c)) is too vague and contains too many loopholes.
2. The fourth prohibited practice (Art. 5 1(d)) on real-time remote biometric identification by law enforcement is far too limited in scope and contains problematic exceptions.
3. Beyond the problems with the existing prohibited practices in the Proposal, it also omits several important red lines outlined by civil society.[31] Many of the practices flagged by civil society organisations as being incompatible with fundamental rights have only been classified as high risk, and the current obligations on high risk systems are insufficient to protect fundamental rights in these cases.
4. The Proposal also fails to be technically and legally consistent, and is not future proof, because it does not outline clear criteria for why certain practices are prohibited and does not foresee a mechanism for the addition of more use cases to the list of prohibitions in Article 5.

In response to the issues, we will do the following below:

1. Provide a criticism of the Art. 5, paragraphs 1(a), 1(b), and 1(c) to explain their shortcomings and suggest improvements or alternatives.
2. Explain the problems with Art. 5 1(d) and offer an alternative formulation that adequately addresses the risk to human rights posed by biometric recognition systems.
3. Suggest additional prohibitions to account for civil society's red lines.
4. Outline criteria for determining why a given practice should be prohibited, and propose a mechanism for the addition of new practices to the list of prohibitions in Article 5.

[29] https://edri.org/our-work/civil-society-call-for-ai-red-lines-in-the-european-unions-artificial-intelligence-proposal/
[30] https://edri.org/our-work/meps-agree-we-need-ai-red-lines-to-put-people-over-profit/
[31] https://edri.org/our-work/civil-society-call-for-ai-red-lines-in-the-european-unions-artificial-intelligence-proposal/

The shortcomings of Art. 5, paragraphs 1(a), 1(b), and 1(c)

As pointed out in the EDPS-EDPB Joint Opinion, rather than safeguarding fundamental rights, the current formulation of the prohibitions in Article 5 risks paying mere lip service to fundamental rights, and limits their scope "to such an extent that it could turn out to be meaningless in practice."[32]

Let us begin with Art.5, 1(a) to see why this is the case. This is a prohibition on the use of an AI system "that deploys subliminal techniques beyond a person's consciousness in order to materially distort a person's behaviour **in a manner that causes or is likely to cause that person or another person physical or psychological harm.**" In our reading, and the reading of other commentators,[33] the entire bolded section of this prohibition is not only unnecessary and vague, but actively undermines the prohibition itself.

It should be abundantly clear that we need to prohibit the use of an AI system "that deploys subliminal techniques beyond a person's consciousness in order to materially distort a person's behaviour." There is no way to materially distort a person's behaviour that is not a gross violation of fundamental rights, including the right to freedom of thought, conscience and religion, and human dignity, which, according to the EU Charter of Fundamental Rights, is inviolable and "is not only a fundamental right in itself but constitutes the real basis of fundamental rights."[34]

Adding the further condition that such a distortion must be done "in a manner that causes or is likely to cause that person or another person physical or psychological harm" suggests, erroneously, that a person's behaviour can be materially distorted in a way that benefits them. **Whether or not some net benefit can be argued to result from this type of manipulation, the fact remains that the person was thereby robbed of their agency and dignity in a manner that undermines any potential benefit they may have received.**

Furthermore, the burden of proof for such manipulation or distortion of behaviour must not fall on the victim. Indeed, if the manipulation is very successful, the victim may not even be aware that they were manipulated. We therefore recommend modifying the language of Article 5, paragraph 1(a), as follows:

- the placing on the market, putting into service or use of an AI system that deploys subliminal techniques beyond a person's consciousness ~~in order to materially distort a person's behaviour in a manner that causes or is likely to cause that person or another person physical or psychological harm~~**, including when such techniques may distort a person's behaviour**

Subliminal techniques in advertising are already prohibited under Article 3e, paragraph 1(b) of Directive 2007/65/EC which states that "audiovisual commercial communications shall not use

[32] EDPS-EDPB Joint Opinion, p.10
[33] https://osf.io/preprints/socarxiv/38p5f
[34] EU Charter of Fundamental Rights, Article 1

subliminal techniques." The language we suggest here would be very much in line with this existing prohibition and could be considered as updating for the age of artificial intelligence.

Art. 5, 1(b) suffers from similar shortcomings to the previous example. In its current form it calls for a prohibition on the 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**." We have highlighted three parts of the current prohibition where we see significant problems. First of all, it is not clear why this prohibition should be limited to age and physical or mental disability. If an AI system exploited the vulnerabilities of a group due to their gender, health, sexual orientation, or any other category, we would find it to be equally harmful. Secondly, any such application would require the identification of people as belonging to these groups, which in the case of physical or mental disability is a highly intrusive form of processing.

Secondly, the bar seems to be set too high by requiring that the system not only exploit a group's vulnerabilities, but also "materially distort" their behaviour *and* do so in a manner likely to cause harm. As stated above, materially distorting someone's behaviour is already problematic enough from a fundamental rights perspective, so we likewise recommend dropping the clause about causing harm in this case. However, it also seems clear that actively setting out to use an AI system to exploit the vulnerabilities of a group should be in itself sufficient reason to prohibit a practice. **If an AI system is designed or used to exploit the vulnerabilities of children, or people with a mental disability, it seems clear that it should be prohibited.** Again, the burden of proof should not fall on potential victims: it should be sufficient to show that the design or use of the system exploits vulnerabilities of certain groups, whether intentionally or not.

As such, Article 5; paragraph 1(b) should be modified to the following:

- the placing on the market, putting into service or use of an AI system that exploits, **intentionally or not,** any of the vulnerabilities of a specific group of persons due to their age, physical or mental disability, **or any other grounds on which discrimination is prohibited under Article 21 of the Charter of Fundamental Rights, or on the basis of mental health status, migration status or gender identity** ~~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~~;

From a fundamental rights perspective, the creation, sale, or use of an AI system which aims to exploit - intentionally or not - the vulnerabilities of people according to "any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation" - or on the basis of mental health status, migration status or gender identity **-** is an affront to the core values of the European Union and must be prohibited.

Finally, regarding Art. 5, 1(c), we again see a similar situation of vagueness and problematic loopholes. The current prohibition refers to the use, by public authorities only, of AI systems 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" where the social score leads to either "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**" or " detrimental or unfavourable treatment of certain natural persons or whole groups thereof that is **unjustified or disproportionate to their social behaviour or its gravity**."

The first issue to note is that no justification is given for the limitation to public authorities, the limitation to 'trustworthiness,' or the limitation to such a practice occurring "over a certain period of time." All three of these limitations undermine the prohibition to the extent of rendering it meaningless. Firstly, as the EDPS-EDPB Joint Opinion notes, "[p]rivate companies, notably social media and cloud service providers, can process vast amounts of personal data and conduct social scoring."[35] As such, they recommend that the Proposal prohibit **any type of social scoring**; we support this call.

Secondly, it is entirely unclear why the prohibition is limited to social scoring that establishes 'trustworthiness.' A social score could in theory be used to measure things beyond just trustworthiness, such as eligibility for certain services or benefits, likelihood to engage in 'antisocial behavior', or even potential future 'criminality'. All of these things would have at least as serious an impact on fundamental rights, and in some cases even greater impact, than estimating trustworthiness. As such, this limitation should be removed.

Thirdly, the limitation to the practice occurring "over a certain period of time" is both unnecessary and damaging. If a system was used to provide me with an instantaneous social score based on an analysis of my biometric features, or my response to a survey question, the harm to fundamental rights would be no different to the case where that social score had been determined by analysing my activities "over a certain period of time." As such, this clause should also be removed.

Finally, the current prohibition demands that "detrimental or unfavourable treatment" must result from the social score that is either in contexts that "are unrelated to the contexts in which the data was originally generated or collected" or "that is unjustified or disproportionate to their social behaviour or its gravity." Both of these conditions again set the bar far too high, and make no sense from a fundamental rights perspective: on the first condition, the harm to fundamental rights does not change if detrimental or unfavourable treatment occurs in contexts that are related to those in which the data was collected or generated; on the second condition, any detrimental or unfavourable treatment resulting from an AI-generated social score will always be unjustified and disproportionate,

[35] EDPS_EDPB Joint Opinion, p.11

as, to quote the EDPS-EDPB Joint Opinion, it "affects human dignity to be determined or classified by a computer as to future behavior independent of one's own free will."[36]

As such, and following the suggestion of the EDPS-EDPB Joint Opinion that "the Proposal should prohibit any type of social scoring," we suggest the following revised language for Art. 5, 1(c) and the deletion of Art. 5, 1(c) (i) & (ii):

- the placing on the market, putting into service or use of AI systems ~~by public authorities or on their behalf~~ for **the calculation or establishment of a 'social score' resulting from** the evaluation or classification of ~~the trustworthiness of~~ natural persons ~~over a certain period of time~~ based on their **physical attributes**, social behaviour or known or predicted personal or personality characteristics ~~, with the social score leading to either or both of the following:~~

Amending Article 5 paragraph 1(d)

The fourth prohibition in Article 5 is, unfortunately, not the prohibition on remote biometric identification that it appears to be. In reality, it is severely limited in scope, and provides wide exceptions that render it all but ineffective in safeguarding fundamental rights from the grave threat posed by this application of AI. The current version of the prohibition is formulated as applying to "the use of '**real-time**' remote biometric identification systems in publicly accessible spaces **for the purpose of law enforcement**." It provides exceptions for three uses, which are (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 of natural persons or of a terrorist attack; and (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/JHA62 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. Let us begin with the general formulation of the prohibition, and then discuss the exceptions.

First of all, the distinction between 'real time' and 'post' remote biometric identification (RBI) is all but meaningless from a fundamental rights perspective. As the EDPS-EDPB Joint Opinion notes, "[p]ost remote biometric identification in the context of a political protest is likely to have a significant chilling effect on the exercise of the fundamental rights and freedoms, such as freedom of assembly and association and more in general the founding principles of democracy." Perhaps the most notorious facial recognition provider in the world, Clearview AI, only provides post RBI, so a system such as theirs would fall outside the scope of this prohibition.

Secondly, the limitation to law enforcement uses overlooks the fact that any use of such systems in publicly accessible spaces poses the same, profound, and irremediable threat to fundamental rights. The EDPS-EDPB Joint Opinion notes that "the intrusiveness of the processing does not necessarily

[36] Ibid P. 12

depend on its purpose [as t]he use of this system for other purposes such as private security represents the same threats to the fundamental rights of respect for private and family life and protection of personal data."

With all this in mind, we therefore support the EDPS-EDPB Joint Opinion's "call for a general ban on any use of AI for an automated recognition of human features in publicly accessible spaces - such as of faces but also of gait, fingerprints, DNA, voice, keystrokes and other biometric or behavioral signals - in any context" and for a ban on the use of "AI systems for large-scale remote identification in online spaces."

Regarding the exceptions in the Proposal, we also follow the EDPS-EDPB Joint Opinion observation that "the potential number of suspects or perpetrators of crimes will almost always be "high enough" to justify the continuous use of AI systems for suspect detection, despite the further conditions in Article 5(2) to (4) of the Proposal." We also support their claim that "when monitoring open areas, the obligations under EU data protection law need to be met for not just suspects, but for all those that in practice are monitored" and that such obligations cannot be met for the use of these technologies in publicly accessible spaces.

Finally, we wish to add that the inclusion of exceptions to a prohibition in the case of RBI in publicly accessible spaces is deeply problematic from a fundamental rights perspective. One of the most serious impacts of these systems on fundamental rights is the chilling effect they create on freedom of expression and freedom of assembly and association. If people are aware that these surveillance systems are installed in public spaces, and can be turned on in certain cases, they will have no expectation of anonymity in public, and will be constantly aware that the system may be turned on and monitoring their behaviour. This creates an undeniable chilling effect on people's rights, and will have a serious, detrimental effect on people's ability to protest and to fully enjoy public space.

As such, we recommend that Art. 5, 1(d) be changed to the following formulation which is in line with the recommendation of the EDPS-EDPB Joint Opinion:

- **any deployment and use of AI for an automated recognition of human features in publicly accessible spaces - such as of faces but also of gait, fingerprints, DNA, voice, keystrokes and other biometric, physiological or behavioral signals - for any purposes, including law enforcement purposes.**

In line with our discussion of the potential limitations of the term "biometric" in section II above, we also add the term "physiological" to ensure that the prohibition achieves its aim. Finally, we support all the comments and recommendations made regarding Art. 5, 1(d) made by European Digital Rights (EDRi), of which Access Now is a member, in their response to this consultation.[37]

Additional prohibitions

In the above analysis of Article 5, we have focused on correcting the formulations of the existing four prohibitions. However, we already noted that these four prohibitions, even in the amended form suggested here, do not adequately address the range of AI practices which need to be prohibited to safeguard fundamental rights. We therefore suggest that the following prohibitions be added to Article 5:

1. A prohibition on using AI to categorise people on the basis of physiological, behavioural, or biometric data, where such categories are not fully determined by that data
2. A prohibition on emotion recognition
3. A prohibition on a number of uses of AI in the context of policing, migration, asylum and border management

Regarding, the first prohibition, the EDPS-EDPB Joint Opinion calls 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." We fully support this call, however, we believe that limiting the scope to biometric data will leave open a number of possibilities for equally harmful systems that use physiological or behavioural that does not meet the bar to be classified as biometric data (i.e. that it allows or confirms the unique identification of a natural person).

For example, an AI system could be used to infer gender from gait or even length of hair, and the argument could be made that the data used to make this inference would not allow or confirm unique identification. If that argument was accepted, such a system would fall outside the scope of the prohibition. To avoid such a loophole, we suggest the following formulation:

- the placing on the market, putting into service or use of AI systems that use physiological, behavioural or biometric data to infer attributes or characteristics of persons or groups which are not solely determined by such data or are not externally observable or whose complexity is not possible to fully capture in data, including but not limited to:
    - Gender & gender identity
    - Race
    - Ethnic origin
    - Political orientation

[37] See EDRi's submission to this consultation for more details: https://edri.org/wp-content/uploads/2021/08/European-Digital-Rights-EDRi-submission-to-European-Commission-adoption-consultation-on-the-Artificial-Intelligence-Act-August-2021.pdf

- Sexual orientation
- Mental health status
- Migration status
- Or other grounds on which discrimination is prohibited under Article 21 of the EU Charter of Fundamental Rights

In essence, this prohibition fulfills the aim of the recommendation from the EDPS-EDPB Joint Opinion that ""biometric categorisation" should be prohibited under Article 5."

We also support the recommendation from the EDPS-EDPB Joint Opinion that the "use of AI to infer emotions of a natural person is highly undesirable and should be prohibited." We do not agree, however, with the additional statement that would allow exceptions for "certain well-specified use-cases, namely for health or research purposes." While research into the science of emotion should be permitted, of course, the use of emotion recognition for health purposes, such as use on patients, should be considered a particularly sensitive domain and subject to prohibition.

Moreover, the idea of allowing the use of emotion recognition in health scenarios is further undermined by the contentious scientific basis of current emotion recognition. It is not a case of not allowing a potentially beneficial technology to be used in health contexts; rather, emotion recognition in its present form does not work, so prohibiting it in a health context is about protecting patients from a pseudoscientific product. As outlined in Section II, there are serious doubts about whether current systems, and even future systems, can actually do what they claim.

Further, devastating criticism of the entire project of emotion recognition has been voiced from many quarters, with even Paul Eckman, whose theories the majority of face-based emotion recognition technology is based on, recently stated in an interview that he "[m]ost of what I was seeing was what I would call pseudoscience" in emotion recognition technology.[38] The International Biometrics + Identity Association (IBIA), described as the "leading voice for the biometrics and identity technology industry", has also stated that applications such as emotion recognition or biometric categorisation are unscientific: "Facial recognition algorithms as a source of information about an individual's characteristics is not science. One cannot infer emotion, patriotism, criminal inclinations, sexual orientation, or other characteristics from a mathematical template of the face."[39]

For all of these reasons, we believe that emotion recognition should be added to the list of prohibited practices, and suggest the following formulation for a prohibition to be added to Article 5:

- the placing on the market, putting into service or use of 'emotion recognition', meaning an AI system for the purpose of identifying or inferring emotions or intentions of individuals or

[38] https://www.ft.com/content/c0b03d1d-f72f-48a8-b342-b4a926109452

[39] https://www.ibia.org/download/datasets/5356/IBIA-CommentsonPendingPortlandFacialRecognitionProposals-1%20(1).pdf

groups on the basis of data relating to their physical, physiological or behavioural characteristics;

If there is, at a future date, scientific proof that some form of emotion recognition systems do indeed work and can be shown not to undermine fundamental rights, exceptions to the above prohibition could be added. However, for the time being, no such proof exists, and a blanket ban on emotion recognition should be added.

Finally, a prohibition must be introduced to cover a range of law enforcement and border management uses of AI systems. A number of applications of AI in these contexts, which correspond to paragraphs 6 & 7 of Annex III, touch the essence of human dignity, and must be prohibited. As noted in the EDPS-EDPB Joint Opinion, applications such as those currently listed as high risk under Annex III, 6. (a) & (e), if "**used according to their intended purpose will lead to pivotal subjection of police and judicial decision-making, thereby objectifying the human being affected. Such AI systems touching the essence of the right to human dignity should be prohibited under Article 5.**"[40]

We therefore suggest that additional prohibitions be added for AI systems used in **predictive policing and uses of AI to assess risk for future criminality, offending or re-offending**, as well as the use of **AI systems at borders and in migration control**. For further details on precisely which applications should be prohibited, please see the response to this consultation from European Digital Rights (EDRi), of which Access Now is a member.[41]

This does not aim to be a comprehensive list of prohibitions, and indeed further prohibitions may be needed to cover applications such as the monitoring, surveillance (including of biometric and other human features) and algorithmic management in employment and educational contexts, as noted by European Digital Rights (EDRi), of which Access Now is a member, in their response to this consultation.[42] Access Now will coordinate with other civil society actors, including unions, to produce a more complete list of prohibitions in the coming months.

Criteria for prohibited AI practices and a mechanism for updating Article 5

While adding these additional prohibitions to Article 5 would vastly improve the Proposal in terms of its protection of fundamental rights, they may not be sufficient to cover the full range of threats posed by AI systems now, and in the future. Artificial intelligence is a fast moving domain. Any regulation that aims to prohibit, or apply other obligations to, a defined list of practices, must have clear mechanisms and processes to update that list to keep pace with technological development. In

[40] EDPS-EDPB Joint opinion, p.12

[41] See EDRi's submission to this consultation for more details: https://edri.org/wp-content/uploads/2021/08/European-Digital-Rights-EDRi-submission-to-European-Commission-adoption-consultation-on-the-Artificial-Intelligence-Act-August-2021.pdf

[42] See EDRi's submission to this consultation for more details: https://edri.org/wp-content/uploads/2021/08/European-Digital-Rights-EDRi-submission-to-European-Commission-adoption-consultation-on-the-Artificial-Intelligence-Act-August-2021.pdf

Article 7, the Proposal has such a mechanism to update the list of high-risk uses in Annex III; however, no such mechanism has been foreseen to update the list of prohibited practices.

As such, it is necessary both to specify criteria that explain both why certain systems are currently subject to a prohibition, and to give guidance to assess whether other current, or not yet existing, AI systems may need to be added to the list of prohibited practices in Article 5. If clear criteria are outlined for why certain practices are prohibited, this will also give legal certainty to developers, companies and researchers by indicating which directions of development are desirable and which not.

We propose the following list of criteria for the prohibition of a practice to be added to the beginning of Article 5:

- An application of AI shall be banned where:
    - The intended purpose of the system is in direct conflict with or fundamentally undermines fundamental rights (including children's rights), democracy or the rule of law, such as by curtailing people's possibilities to exercise their rights; or
    - It poses unacceptable risk(s) of adversely impacting fundamental rights (including children's rights), democracy or the rule of law; or
    - The risk(s) it poses cannot be mitigated by technical fixes or procedural and legal safeguards, including the obligations proposed under this legal instrument for high-risk systems; or
    - The risk(s) it poses are likely to fall disproportionately on marginalized groups; or
    - The risk it poses to fundamental rights (including children's rights), democracy or the rule of law is so severe that even a low probability is intolerable.
    - The application infringes an absolute right.

Furthermore, a provision should be added to Article 5 to allow for the list of prohibition practices to be updated. The question remains open regarding which body should be empowered to do so. At present, Article 7 allows the Commission to adopt delegated acts to update the list in Annex III. However, the question has been raised as to whether the proposed European Artificial Intelligence Board (EAIB) should rather be given such the power to make binding decisions in this regard (providing its independence is ensured). We foresee a scenario where a properly independent EAIB could be empowered to adopt binding decisions to update the list of prohibited practices in Article 5, although this is conditional on a number of modifications to the way the Proposal envisages the structure of the EAIB. If decisions to update such a list shall be taken by delegated act, regulators shall be empowered to give binding opinions to the European Commission on any such proposed act.

Regardless of which body is empowered to update the list of prohibited practices, the fact remains that such a possibility of updating the list of prohibitions must be incorporated into the Proposal, along with a set of criteria which can guide any decision to add a new practice to the list.

IV. Address the relative lack of obligations on 'users' of AI systems

Central to the Proposal is the distinction between 'providers' and 'users' of AI systems. According to Article 3 providers are defined as "a natural or legal person, public authority, agency or other body that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark, whether for payment or free of charge," while a user is defined as "any natural or legal person, public authority, agency or other body using an AI system under its authority, except where the AI system is used in the course of a personal non-professional activity."

The Proposal currently envisages little to no obligations on users, with the entire conformity assessment procedure falling on providers. While we welcome the obligations placed on providers, and would even suggest that they could be strengthened, we believe that a significant loophole exists in the relative lack of obligations on users. To understand why, let us look at what obligations currently exist for users.

Article 29, on *Obligations of users of high-risk AI systems*, list a number of obligations for users of high-risk AI systems, including:

- to use systems in accordance with the instructions provided (29(1));
- where relevant, to "ensure that input data is relevant in view of the intended purpose of the high-risk AI system" (29(3));
- where they believe that "risks to the health or safety or to the protection of fundamental rights of persons are concerned" (as defined in Art 65(1)), or find "any serious incident or any malfunctioning", they must "inform the provider or distributor and suspend the use of the system" (29(4))
- "shall keep the logs automatically generated by that high-risk AI system, to the extent such logs are under their control" (29(5));
- and, finally, "use the information provided under Article 13 to comply with their obligation to carry out a **data protection impact assessment** under Article 35 of Regulation (EU) 2016/679 or Article 27 of Directive (EU) 2016/680, where applicable" (29(6)).

The deployment of an AI system, in particular circumstances and with particular aims, will have a significant impact on the risk it poses to fundamental rights. The idea that a provider of an AI system can foresee all possible risks in abstraction for the concrete context of use is thus flawed. This is echoed by the EDPS-EDPB Joint Opinion, which notes that "it will not always be possible for a provider to assess all uses for the AI system," and that "the initial risk assessment [performed by the provider] will be of a more general nature than the one performed by the user of the AI system."[43] Unfortunately,

[43] EDPS-EDPB Joint Opinion P. 9

there is currently no obligation within the Proposal to ensure that users will perform a risk assessment.

Currently, the only obligation which would require a user to perform any kind of impact, or risk, assessment is the obligation under Art 29(6)requiring that a data protection impact assessment (DPIA) should be carried out, only where applicable according to Article 35 of Regulation (EU) 2016/679 or Article 27 of Directive (EU) 2016/680. This means that in circumstances where a DPIA is not applicable, a user may procure and deploy a high-risk AI system without being obliged to perform any form of impact assessment.

Regarding high-risk systems that do involve the processing of personal data, the EDPS-EDPB Joint Opinion also notes that the Proposal "requires the providers of the AI system to perform a risk assessment, however, in most cases, the (data) controllers will be the users rather than providers of the AI systems (e.g., a user of a facial recognition system is a 'controller' and therefore, is not bound by requirements on high-risk AI providers under the Proposal)."[44]

The current set of obligations imposed upon users, as listed in summary above, is therefore far from sufficient to ensure the protection of fundamental rights. In particular, we believe that the obligation in Art 29(6) falls short of what is necessary. This obligation states that a data protection impact assessment (DPIA) should be carried out, only where applicable according to Article 35 of Regulation (EU) 2016/679 or Article 27 of Directive (EU) 2016/680. This means that in circumstances where a DPIA is not applicable, a user may procure and deploy a high-risk AI system without being obliged to perform any form of impact assessment.

Given that the list of AI uses in Annex III have been explicitly identified as posing a high risk to fundamental rights, we believe that **all users of high-risk AI systems should be obliged to perform either a DPIA, or, where are DPIA is not applicable according to Article 35 of Regulation (EU) 2016/679 or Article 27 of Directive (EU) 2016/680, they should be required to carry out a human rights impact assessment (HRIA)**. This would ensure that in all cases, a user procuring a high-risk AI system will have performed some form of impact assessment to address how their use of this system, in these particular circumstances, will impact fundamental rights.

It is worth noting here that the EDPS-EDPB Joint opinion does not that "the classification of an AI system as posing "high-risk" due to its impact on fundamental rights does trigger a presumption of "high-risk" under the GDPR, the EUDPR and the LED to the extent that personal data is processed."[45] This means that for all high-risk AI systems that process personal data, a DPIA will be required. The further obligation to perform a HRIA, as suggested above, would only apply to cases of high-risk AI systems that **do not process personal data**. In the coming months, Access Now will provide more

[44] Ibid P. 9

[45] Ibid P. 9

detailed guidance on what such an impact assessment can look like in the context of high risk AI systems.

A further important clarification is that any impact assessments conducted in the context of procuring a high-risk AI system, whether it is a DPIA, HRIA or whatever other form, must also fall under the scope of Article 64(3). Article 64(3) states that "[n]ational public authorities or bodies which supervise or enforce the respect of obligations under Union law protecting fundamental rights in relation to the use of high-risk AI systems referred to in Annex III shall **have the power to request and access any documentation created or maintained under this Regulation** when access to that documentation is necessary for the fulfilment of the competences under their mandate within the limits of their jurisdiction." As such, these bodies must have access to all impact assessments carried out by users in the circumstances described.

V. Extend the scope of Article 60's 'EU database for stand-alone high-risk AI systems' to users

One of the most valuable provisions of the Proposal is Article 60's EU database for stand-alone high-risk AI systems. According to Article 60, this database would contain publicly accessible information on all standalone high-risk AI systems which are placed on the market or put into service in the Union. This database will be highly valuable for all actors in the AI ecosystem, as it will provide a clear, publicly accessible overview of all high-risk systems on the market in the Union.

However, this important transparency measure **must also be extended to how stand-alone high-risk systems are used**. As noted above, the context of use of an AI system has a huge impact on how it affects fundamental rights. From the perspective of civil society, and from the people affected by and interacting with AI systems, what matters most is knowing **where a high-risk AI system is actually deployed or in use**. To use an example of a high-risk system from Annex III, knowing that an AI system "intended to be used for the purpose of determining access or assigning natural persons to educational and vocational training institutions" is on the market in the EU does not provide sufficient information for people whose rights may be impact by that system. What they need to know, and what regulators and civil society organizations need to know, is where that system is in use. As such, the EU database for stand-alone high-risk AI systems **must be expanded to include not just a list of all existing standalone high-risk AI systems but also all their uses.**

Special consideration must be given to the use of high-risk AI systems by public authorities. People interacting with public services do not typically have a choice to 'take their business elsewhere' and will be obliged to interact, or be subjected to, any AI system put into service by a public authority. As such, they should have a right not only to know if and when a public authority is using a high-risk AI system, but also to have access to the impact assessment carried out for the procurement of that system. We therefore recommend that all impact assessments carried out by public authorities in the context of procuring a standalone high-risk AI system should furthermore be **publicly viewable in their full, unredacted form**.

In addition to making publicly available all impact assessments carried out by public authorities, the database should also contain all impact assessments, whether DPIAs or HRIAs, carried out by private sector users of high-risk AI systems. Where necessary, these may be redacted, but the full unredacted version should also be contained in the data **and be accessible to the public on request**.

VI. Address the issues with sandboxes in Articles 53 & 54

Regarding the proposal for AI regulatory sandboxes in Articles 53 and 54, Access Now supports a number of its provisions including:

- that to the extent that "the innovative AI systems involve the processing of personal data or otherwise fall under the supervisory remit of other national authorities or competent authorities providing or supporting access to data, the national data protection authorities and those other national authorities are associated to the operation of the AI regulatory sandbox;"
- that the sandboxes "shall not affect the supervisory and corrective powers of the competent authorities," and;
- that participants in the sandbox "shall remain liable under applicable Union and Member States liability legislation for any harm inflicted on third parties as a result from the experimentation taking place in the sandbox."

AI regulatory sandboxes must not become a "law free zone" where dangerous experimentation can be conducted outside the bounds of existing law and ethical standards. Rather, they should be seen as opportunities for regulatory scrutiny over the development of new systems.

However, despite the positive provisions listed above, the Proposal introduces a number of highly problematic measures related to the AI regulatory sandboxes. In fact, the mention of sandboxes in the Proposal shall not be construed as an encouragement to develop more of these practices. Most problematic of all are the provisions outlined in Article 54 regarding "[f]urther processing of personal data for developing certain AI systems in the public interest in the AI regulatory sandbox."

Article 54(1)(a) lists three areas where an "innovative" AI system may be used to process personal data collected for other purposes for "safeguarding substantial public interest":

> (i) the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security, under the control and responsibility of the competent authorities. The processing shall be based on Member State or Union law;

(ii) public safety and public health, including disease prevention, control and treatment;

(iii) a high level of protection and improvement of the quality of the environment;

While we are generally concerned about establishing any grounds for the further processing of personal data, due to the risk that such provisions pose to undermining the protections offered by existing data protection law, we are particularly concerned about paragraph (i). Perhaps the riskiest, and most problematic uses of AI are in the "prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security." Such uses of AI systems should, rather than being incentivised or freed from scrutiny, be subject to the utmost caution, if they are to be pursued at all. **We therefore recommend that the use cases under paragraph (i) be removed from Article 54.** We also recommend that the term "innovative" be deleted from Article 53, paragraph 2:

- Member States shall ensure that to the extent the ~~innovative~~ AI systems involve the processing of personal data or otherwise fall under the supervisory remit of other national authorities or competent authorities providing or supporting access to data the national data protection authorities and those other national authorities are associated to the operation of the AI regulatory sandbox

AI systems should fall under the remit of the authorities regardless of whether they are innovative, so this word should be deleted to prevent any loophole.

Regarding Article 54 in general, we support the points raised in the EDPS-EDPB Joint Opinion that "relationship of Article 54(1) of the Proposal to Article 54(2) and recital 41 of the Proposal and thus also to existing EU data protection law remains unclear." As noted in the Joint Opinion, "the GDPR and the EUDPR already have an established basis for 'further processing,'" and we fully support the additional point that "balancing between the controller's interests and the data subject's interests do not have to hinder innovation." The legal basis of the grounds for further processing must be clarified, and the necessity of any such grounds for further processing be considered against the risk they pose to undermining existing protections.

Finally, Access Now believes that the use of AI regulatory sandboxes must be subject to the highest level of public scrutiny and transparency. The intention of such sandboxes is to facilitate the development of AI systems "for safeguarding substantial public interest": as such, the public must be able to know exactly what types of systems are being developed in them. In line with initiative such as Findata,[46] Finland's innovative Social and Health Data Permit Authority that promotes secondary use of health and social data, the AI regulatory sandboxes should provide publicly available information regarding all requests to make use of the sandbox, all accepted and rejected applications, information about project currently in development in the context of the sandbox, and follow up information on what happens with projects afterwards. Finally, sandboxing involves close scrutiny from regulators,

[46] https://findata.fi/en/

including data protection authorities, which inherently involve a use of their already stretched ressources. To avoid abuses and limit risk of authorities' resources being diverted from monitoring and enforcing regulations to monitor sandboxing, the relevant authorities must be provided with additional, dedicated staff to oversee and regulate such exercises and the number of sandboxing exercises authorised per year and per country shall be limited.

VII. Address the gaps in enforcement & redress

There are a number of issues to raise regarding the enforcement of the proposed regulation, and the (lack of) possibilities it presents for redress for individuals whose rights are affected by AI systems. As we have previously noted, the proposed enforcement mechanism which includes the **EU Artificial Intelligence Board** and the **appointment of national supervisory authorities** lacks clarity. The creation of a new board and appointment of supervisory authorities with responsibilities and competencies that may overlap with the European Data Protection Board and existing Data Protection Authorities could cause confusion, and in the worst case **undermine the authority of the EDPB and the DPAs on matters which are central to their competencies**. The role of DPAs and the EPDB should be clarified to ensure the smooth functioning alongside, and in cooperation with the existing data protection authorities.

Regarding the appointment of the national supervisory authority, we follow the recommendations in the EDPS-EDPB Joint Opinion that the "designation of DPAs as the national supervisory authorities would ensure a more harmonized regulatory approach, and contribute to the consistent interpretation of data processing provisions and avoid contradictions in its enforcement among Member States."[47] As the EDPS-EDPB Joint Opinion notes, DPAs are already working on AI systems in the context of their current competencies, and have "an understanding of AI technologies, data and data computing, fundamental rights, as well as an expertise in assessing risks to fundamental rights posed by new technologies." They also note, importantly, that for the vast majority of high-risk AI systems, some processing of personal data will be involved, meaning that "provisions of the Proposal are directly intertwined with the data protection legal framework."

Access Now also has serious concerns about the role, structure and independence of the proposed European Artificial Intelligence Board. We support the following recommendations made in the EDPS-EDPB Joint Opinion:

- Given the dominant role assigned to the European Commission in the EAIB, Access Now support the recommendation that "the future AI Regulation should give more autonomy to the EAIB, in order to allow it to truly ensure the consistent application of the regulation across the single market"
- We support the recommendation "that the cooperation mechanisms between national supervisory authorities be specified in the future AI Regulation" and to ensure a "mechanism

[47] EDPS-EDPB Joint Opinion p.15

guaranteeing a single point of contact for individuals concerned by the legislation as well as for companies, for each AI system, and that for organisations whose activity covers more than half of the Member States of the EU, the EAIB may designate the national authority that will be responsible for enforcing the AI Regulation for this AI system."

- We also support the recommendation that "considering the independent nature of the authorities that shall compose the Board, the latter shall be entitled to act on its own initiative and not only to provide advice and assistance to the Commission" and further that "the EAIB shall have sufficient and appropriate powers, and its legal status should be clarified", including that the "EAIB should be empowered to propose to the Commission amendments of the annex I defining the AI techniques and approaches and of the annex III listing the high-risk AI systems referred to in article 6(2). The EAIB should also be consulted by the Commission prior any amendment of those annexes."[48]
- Finally, we support the recommendation that the Fundamental Rights Agency should be made one of the observers of the board.

On redress, it has been noted by many commentators that the Proposal lacks any clear measures to facilitate complaints from, or redress for, individuals or groups who are impacted by high-risk AI systems covered under the Proposal. As Melanie Fink notes, the Proposal "does not include obligations of AI users to explain or justify the decisions they reach towards those affected by them, even less a corresponding right on the part of individuals to demand that." She further notes that this is particularly problematic in the case of public authorities, whose use of AI systems is steadily increasing, because the "exercise of state power, such as law enforcement or adjudication, brings particular fundamental rights risks [and is] for that reason [...] subject to stronger safeguards against abuse of that power—transparency, accountability, oversight." These safeguards are threatened by the use of AI systems which can introduce problematic forms of opacity into decision-making processes.

Access Now supports Fink's suggestion that "the applicability of [the right to a reasoned decision] in the AI context should be made explicit [and that] the benchmarks used to assess compliance with the right to a reasoned decision in the AI context should be clarified." Moreover, an individual complaints mechanism, analogous to the ability to lodge complaints before the European Data Protection Supervisor under Article 57 of the GDPR, should be added to the Proposal.

This call is backed up further by the EDPS-EDPB Joint Opinion which notes that "[w]hether they are end-users, simply data subjects or other persons concerned by the AI system, the absence of any reference in the text to the individual affected by the AI system appears as a blind spot in the Proposal," and they "urge the legislators to explicitly address in the Proposal the rights and remedies available to individuals subject to AI systems."

[48] EDPS-EDPB Joint Opinion, p.15-16 passim

VIII. Additional issues to be remedied

Finally, we wish to highlight a number of additional issues with the Proposal. As mentioned in the introduction, this document does not claim to address all aspects of the Proposal, but only to present an initial assessment of some aspects of it which will require serious consideration and modification in the next stages of the legislative process. Access Now will continue to engage with legislators throughout the process to offer expert, fundamental-rights focused input on the many contentious issues the Proposal aims to tackle.

On high-risk AI systems

There are a number of additional points to be raised regarding the current treatment of high-risk AI systems in the Proposal.

First of all, as the EDPS-EDPB Joint Opinion notes, "the classification of an AI system as high-risk does not necessarily mean that it is lawful per se and can be deployed by the user as such." It further notes that the notion that "unlike prohibited systems, the high-risk systems may be permissible in principle is to be addressed and dispelled in the Proposal, especially since the proposed CE marking does not imply that the associated processing of personal data is lawful." Based on this reasoning, we support the EDPS-EDPB Joint Opinion in demanding that "the compliance with legal obligations arising from Union legislation (including on the personal data protection) should be precondition to being allowed to enter the European market as CE marked product," and support the recommendation of "including in Chapter 2 of Title III of the Proposal the requirement to ensure compliance with the GDPR and the EUDPR."

A final point on the treatment of high-risk AI systems in the proposal is that there is currently no mechanism foreseen to update the list of headings in Annex III. As it stands, new systems can be added to Annex III according to the procedure outlined in Article 7, but only if they "are intended to be used in any of the areas listed in points 1 to 8 of Annex III." This condition severely limits the flexibility of the Proposal to adapt to the fast-moving developments of AI technology. Article 7 should thus be amended to allow not only for the addition of new AI systems to the existing headings in Annex III, but also to allow for the modification of existing headings, or the addition of entirely new headings.

Related to this point, heading 1 on Annex III, is currently formulated as "Biometric identification and categorisation of natural persons." This is a problematic limitation, as it means that no form of biometric authentication system can be considered high risk according to the Proposal. We also point to the limitations of the term "biometric" as outlined in Section II of this document. We therefore propose amending heading 1 to the following: "**Physiological, behavioural, and biometric authentication, identification and categorisation of natural persons.**"

The final additional point to be made concerns the role of harmonised standards as outlined in Articles 40 and 41. As highlighted by Michael Veale and Frederik Zuiderveen Borgesius in their article, 'Demystifying the Draft EU Artificial Intelligence Act,' "[a]rguably the most important actors in the Draft AI Act are the double-act of CEN (European Committee for Standardisation) and CENELEC (European Committee for Electrotechnical Standardisation)."[49] As the authors note, according to Articles 40 and 41, "if these organisations adopt a standard relating to the Draft AI Act, providers can follow this standard, rather than interpreting the essential requirements. If following the standard, providers enjoy a presumption of conformity." This provides a worrying loophole, because as Veale and Borgesius note, providers are incentivised to follow harmonised standards rather than the Proposals essential requirements. This is made even more problematic by the fact that civil society organisations have little experience or impact in influencing such bodies, and "struggle to participate in arcane private standardisation processes, yet the outputs are important standards Member States must recognise". Furthermore, Veale and Borgesius note that "the European Parliament has no binding veto over harmonised standards mandated by the Commission." The significant role according to these bodies in the Proposal risks undermining the fundamental rights protections it currently accords, and allowing a significant way for AI providers to lobby for weaker obligations that the Proposal itself mandates.

VIII. Conclusion

While the Proposed AI Act offers a number of potentially useful elements to ensure the protection of fundamental rights in the context of AI systems, it will fail to provide effective protection unless the issues raised above are addressed. We look forward to working with the co-legislators, other civil society organizations, representatives of affected communities, and other stakeholders, to ensure that the necessary modifications are made in the next steps of the legislative process.

For more information, please contact:
Daniel Leufer, Europe Policy Analyst at Access Now, daniel.leufer@accessnow.org

Access Now 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.

[49] https://osf.io/preprints/socarxiv/38p5f p.14