Prohibit emotion recognition in the Artificial Intelligence Act

What is emotion recognition?

The term 'emotion recognition' covers a range of technologies that claim to infer someone's emotional state from data collected about that person. This can include inferring emotion from a person's facial configurations or expressions, voice, or from granular data from wearable devices, or even neurological data from brain-computer interfaces (BCIs).

Many 'face-based' emotion recognition applications use Paul Ekman's controversial and much-criticised 'basic emotions' theory, which posits 'universal categories' of human emotion and describes how these can be read from facial configurations (i.e. inferring that a smiling person is happy). However, emotion recognition systems also go beyond Ekman's list of basic emotions, to include "Artificial Intelligence Polygraphs" that claim to detect deception.

As ARTICLE 19 has pointed out, emotion recognition is a highly invasive form of surveillance that "involves the mass collection of sensitive personal data in invisible and unaccountable ways, enabling the tracking, monitoring, and profiling of individuals, often in real time."

Moreover, the scientific foundations of emotion recognition systems have been repeatedly called into question, such that there are serious doubts about whether current systems, and even future systems, can actually do what they claim.

A prominent study by researchers in the science of emotion 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**". These systems therefore rely on simplistic, inaccurate inferences and cannot reliably perform their intended function, and fail to capture diverse expressions of emotion across different cultures. Further, devastating criticism of the entire project of emotion recognition has been voiced from many quarters, with even Paul Ekman stating that "**[m]ost of what I was seeing was what I would call pseudoscience**" in emotion recognition technology.

Why we need a prohibition on emotion recognition in the AI Act

In her 2021 annual report on *The Right to Privacy in the Digital Age*, the UN High Commissioner for Human Rights, Michelle Bachelet, notes that the "the use of emotion recognition systems by public authorities, for instance for singling out individuals for police stops or arrests or to assess the veracity of statements during interrogations, **risks undermining human rights, such as the rights to privacy, to liberty and to a fair trial**" and that a "risk-proportionate approach to legislation and regulation will require the prohibition of certain AI technologies, applications or use cases, where they would create potential or actual impacts that are not justified under international human rights law, including those that fail the necessity and proportionality tests."

This call for a prohibition was further supported by the authors of a study commissioned by the AIDA committee of the European Parliament entitled, *Identification and assessment of existing and draft EU legislation in the digital field*, which stated that "Emotion recognition systems powered by AI may have highly undesired discriminatory and dignity consequences, manipulative effects, and risk impact. Therefore, general prohibition might be an option to consider."

In their Joint Opinion on the AI Act, the European Data Protection Board (EDPB) and European Data Protection Supervisor (EDPS) state that the "use of AI to infer emotions of a natural person is highly undesirable and should be prohibited." The EDPB-EDPS statement further notes that exceptions should be made for "certain well-specified use-cases, namely for health or research purposes". However, the fact that these systems are so controversial, and have been shown to be based on flawed scientific premises, suggests that they should not be allowed in sensitive domains such as health unless subject to rigorous clinical validation and the highest level of regulatory scrutiny.

Regardless of whether AI systems can actually infer our emotional state, making unwanted inferences about our emotional state represents an unacceptable intrusion into our private mental life, undermining our rights to privacy and freedom of thought. If security forces use emotion recognition to detect potentially 'aggressive' people in crowds or at protests and proactively apprehend these people before they have committed any aggressive act, it doesn't matter whether the inference was flawed or not; the consequences are real and undermine our rights to freedom of expression and freedom of assembly.

AI systems designed to reveal and manipulate our innermost thoughts and feelings, and which are based on dubious scientific premises, represent one of the greatest threats to the European Union's desire to create an ecosystem of trust and excellence for AI and must be prohibited.

Amendments to the AI Act's treatment of emotion recognition

There are a number of flaws and issues in the treatment of emotion recognition in the AI Act.

Firstly, **the definition of emotion recognition in Article 3, paragraph (34) of the AI Act is technically flawed**. This definition is limited to systems that use **biometric data**, defined in Article 3(33) as data relating "to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person."

The key question is: **does all physiological data allow unique identification?** Emotion recognition systems often use physiological data that arguably **don't meet the high bar for identification required to be classified as biometric data** (e.g. galvanic skin response). In such cases, providers could argue that **their system is not subject to obligations under the AIA.**

To rectify this potential loophole, **a new definition should be added to Article 1 for biometrics-based data** to cover data related to the physical, physiological or behavioural characteristics of a natural person which does not allow or confirm unique identification**.**

An amendment should also be made to Recital 7 to clarify the need for the inclusion of a definition of biometrics-based data, and to clarify that it is intended purely to capture data that falls outside the scope of the existing definition of biometric data in the GDPR.

The definition of emotion recognition should then **be modified so as to clarify that it includes systems which use biometrics-based data, and to clearly include so-called 'AI polygraphs' and other systems that claim to detect deception**.

It is important to note that this definition must specifically refer to systems that make inferences about emotions or states of mind based on physiological and other data, and thus does not cover systems that detect purely physiological traits or behaviours, such as whether a driver is

[falling asleep](#) or [whether a caller to an emergency services line is having a heart attack](#). Recognition of simple facial expressions, gestures or characteristics of a person's voice, such as detecting if a person is smiling, frowning, gesturing with their hands, shouting or whispering, should also not be considered emotion recognition. AI applications that provide information on these for users, such as persons with disabilities, can be considered as ordinary assistive devices and are thus subject to existing legal frameworks for assistive technologies and – insofar as they use high-risk AI – the rules established for such systems in the AI Act.

Secondly, **a general prohibition on emotion recognition should be added to the list of prohibited practices.**

On the need for exceptions to the prohibition on emotion recognition

While we firmly believe that a general prohibition on emotion recognition is necessary to safeguard people's fundamental rights, **we wish to engage in a constructive dialogue with any stakeholders who believe that exceptions should be made for certain limited use cases**. The use cases most commonly presented as requiring such an exception are related to healthcare, assistive technologies and in related therapeutic contexts for persons with disabilities such as visual impairment or certain people on the autism spectrum who may have difficulty recognising emotions. However, these arguments are not supported by any widespread call from any group of disabled people for the development of such technologies, but are mainly driven by certain researchers and companies.

Groups representing autistic people, such as the [European Council of Autistic People](#), and scholars who have investigated issues related to autism and AI, such as [Os Keyes](#), are sceptical of both the need and the feasibility of such technology for any significant proportion of the autism spectrum, noting that a small number of autistic people may have a specific impairment in this area, but it is not a defining feature of the autism spectrum. According to the idea of the "[Double Empathy Problem](#)", people who differ from each other neurologically have equal difficulty interpreting each others' emotions. This means that the 'problem' does not lie with neurominorities only, and therefore any emotion detection system trained to recognize so-called 'normal' expressions of emotion would privilege the neurotypical person only. This could in fact exacerbate problems. These arguments should be considered when discussing the needs of other neurominorities and neurological impairments, as well.

Given that there are serious doubts about the potential effectiveness of emotion recognition systems, any exceptions for such uses must be specific and limited, and subject to the strictest safeguards and highest standards of testing and scientific validation to ensure that they do not cause serious harm in sensitive cases such as healthcare, or give misleading information about emotions to people with any impairment that may limit their ability to detect emotions.

There is a serious risk that unscientific, flawed emotion recognition systems used in such contexts, or ones based on a medical-model interpretation of disability only and failing to consider the complexity of interactions where disability arises, would lead to more harm for the people they are intended to help. Where certain uses of emotion recognition technologies are permitted, measures should be put in place to inform users of the limitations, potential risks, including flawed outcomes, of such technologies, as well as to ensure that such technologies are not used against people with disabilities or in ways that undermine their rights.

Moreover, while there are claims that emotion recognition systems can help persons with

disabilities, the UN Special Rapporteur on the rights of persons with disabilities, Gerard Quinn, in his 2021 report on artificial intelligence and the rights of persons with disabilities, notes that:

> ***emotion-processing algorithms may misinterpret the facial expressions of autistic persons, people with Williams syndrome or others with atypical facial expressions, such as persons who have experienced stroke, Parkinson's disease, or Bell's palsy***

As such, the use of emotion recognition systems to help one group of persons with disabilities may inadvertently undermine the rights of other persons with disabilities. If emotion detection systems are combined with attempts to teach autistic people about the 'correct' ways to show emotion, there is also the risk of embedding ableist stereotypes and reinforcing a problematic image of autistic people as "suffering from a disorder, one characterised by stunted or absent social skills and emotional awareness" as opposed to the neurodiversity perspective that "seek[s] to portray autism and autists in a more positive light." It is important to guard against AI applications being used to serve manipulation or coercion in the name of rehabilitation, signalling to any group of disabled people that their natural emotional expressions are wrong or unacceptable and need to be masked and eradicated by imposing more "normal" expressions.

If, despite all this evidence about the lack of usefulness and danger posed by emotion recognition technologies as assistive technologies, certain groups representing persons with disabilities provide evidence that emotion recognition could in fact be useful for their needs, then consideration could be given to **an exception for these use cases. Any such uses of emotion recognition in assistive technologies must be developed in consultation with the full range of affected groups and relevant experts**, such that it ensures that any emotion recognition systems that are permitted for these limited use cases:

- undergo strict testing to ensure scientific and clinical validity;
- contain clear advice to anyone that may procure / use them about the limitations of such technologies and their potential risks, including of flawed or potentially harmful outcomes;
- are developed with the active participation and input of the groups they are intended to benefit, are subject to approval by those groups, and also consult those with expertise in the range of fundamental rights that could be deliberately or inadvertently impacted (for example privacy and data protection);
- are developed and deployed in a manner that respects the rights of all persons likely to be affected by them; and
- if a specific system cannot be developed in a way that proves compliance with the above criteria, then the risks of such a use case will be clearly unacceptable, and it will remain strictly prohibited under the AI Act.

If a specific emotion recognition system cannot fulfil these criteria, then it should not be made available on the market.

For a comprehensive position on stopping biometric mass surveillance in the AI Act, see also the amendment documents on **banning problematic uses of biometric categorisation** and **prohibiting all remote biometric identification**. Further modifications must also be made to Annex III to rectify issues with the treatment of biometric categorisation and other biometric technologies. For more information, see the related amendment documents on **regulating non-prohibited uses of biometrics in AI systems**. For more information on any of these issues, please contact Daniel Leufer (daniel.leufer@accessnow.org) and Ella Jakubowska

(ella.jakubowska@edri.org).

The authors of this paper also wish to thank the European Disability Forum (EDF) and the European Council of Autistic People (EUCAP) for invaluable comments on this paper.