ALLAI.

24.02.2022

#2

AIA in-depth

#2 | Prohibited AI Practices

Article 5

This report is the second in a series of in-depth analyses of the European Commission proposal for a Regulation for Artificial Intelligence (AIA).

Authors:
Catelijne Muller, LL.M
Christofer Talvitie
Rosa van Ree

CONTENTS

EXECUTIVE SUMMARY

In this paper, which is the second in a series, we will dive deeper into the main elements of Article 5 of the AIA: Prohibited AI Practices. By evaluating each prohibited practice, we will assess the scope of these prohibitions, also in relation to other legislation.

At the time of publication of this paper, the Slovenian Presidency of the European Council already issued a proposed compromise text for articles 1 - 7 of the AIA. This paper will take this compromise text into consideration where relevant.

Main findings

1 **Three types of AI practices are prohibited:**

- Materially distorting behaviour with subliminal techniques or by exploiting vulnerabilities
- Social Scoring by public authorities
- Biometric Identification for law enforcement purposes

2 **Prohibition of AI-driven manipulation is limited to a very rare and narrow set of practices.**
The AIA presents a grand opportunity to address the wider societal harms that AI-driven manipulation can bring and curb the trajectory towards the Internet-of-Minds.

3 **Prohibition of social scoring is welcome but should be widened to private actors and clarified.** For social scoring to be effectively banned in Europe, a clearer line should be drawn between what is considered ‘social scoring’ and what can be considered an acceptable form of evaluation for a certain purpose.

4 **The prohibition of real time remote biometric identification only covers a narrow set of practices.** Under the current prohibition, many biometric recognition practices (a.o. biometric assessment and biometric categorization) remain allowed, also by law enforcement. Outside of law enforcement all biometric recognition practices could in principle remain possible.

1 PROHIBITING AI PRACTICES

By introducing an escalating ‘risk pyramid’ (from limited to medium-risk, to high-risk, to unacceptable risk) used to categorize a number of general AI practices as well as a number of domain specific AI uses, the Commission acknowledges that not all AI poses risks and not all risks are equal. Given that some risks are too grave to be acceptable, the Commission proposed to ban some AI practices. However, the descriptions of the various prohibited AI practices are at times unclear, multi-interpretable and could lead to legal uncertainty and create loopholes. Moreover, a number of categorisation choices defy the overall objective of the AIA, which is the protection of health, safety and fundamental rights.

Article 5 of the AIA broadly prohibits 3 types of AI practices: materially distorting behavior, social scoring and biometric identification. However, the prohibitions are not always as comprehensive, effective or clear as one would hope. We dive deeper into the exact wordings of Article 5 in order to give a clear picture of what exactly would be prohibited and what, on the other hand, would remain allowed.

TITLE II

PROHIBITED ARTIFICIAL INTELLIGENCE PRACTICES

Article 5.1 (a) & (b)

Materially distorting behaviour

Main elements and conditions:

- Distorting a person's behaviour
- With subliminal techniques or by exploiting vulnerabilities
- (Likely) causing physical or psychological harm

Article 5.1 (c)

Social Scoring

Main elements and conditions:

- Evaluation or classification of trustworthiness
- By public authorities
- Leading to detrimental or unfavourable treatment

Article 5.1 (d)

Biometric Identification

Main elements and conditions:

- Identification (one-to-many)
- Real time and remote
- For law enforcement
- Except search for victims, imminent physical threat, terrorism, extradition

2 MATERIALLY DISTORTING BEHAVIOR

The prohibitions of article 5.1 paragraphs (a) and (b) center around ‘distorting a person’s behavior’, causing psychological or physical harm. According to the Commission, this includes subliminal manipulation, where the AI-system simply selects the person to be targeted, and that person is micro-targeted in a hidden way. The Commission also notes that the prohibitions should cover a very narrow set of AI practices and address only the most severe risks and harmful consequences of manipulation. Hence the condition of ‘physical or psychological harm’.

There have been calls for the inclusion of ‘economic/financial harm’ to these conditions, however the Commission argues that the proposal for the Digital Services Act and Directive 2005/29/EC (“the Unfair Commercial Practices Directive”) regulate this. Moreover, the Commission argues that additional risks that may occur (beyond physical or psychological harm) are to be addressed through existing regulation of fundamental rights and safety. In the next chapters we dive deeper into these positions.

2.1 The Internet-of-Minds

The combination of data and computing power has enabled the capture of an unprecedented amount of information about us—what we do, where we go, what we (supposedly) think, what we say, how we (supposedly) feel. Ever larger parts of our lives (shopping, socialising, playing, entertainment, working, learning, assembly, navigation, etc.) happen online, where we constantly leave data trails. We have seen a surge in sensors being deployed in public spaces, at home, and on our bodies. As it stands today, over 13 billion devices are connected to the internet and recently the concept of the Internet-of-Bodies, has emerged. The use of AI has made these troves of data a playground for categorisation, sifting, sorting and profiling of our entire lives, behaviour, thoughts, ideas and beliefs, of unprecedented scale and depth. These insights in turn have created ample opportunities to target, nudge, manipulate and deceive us, leading to altered and sometimes harmful beliefs, thoughts, ideas and behaviour. And all this happens ever more covertly, by using dark patterns that lure us into accepting these practices and by deploying subliminal techniques that we hardly notice, at the abstract level of code, algorithms, models and data. Beyond the Internet-of-Things and the Internet-of-Bodies, we are also on a trajectory towards the Internet-of-Minds.

2.1.1 AI-driven manipulation: the impact on democracy and society

AI has become a popular tool for surveillance, categorization and manipulation not only for oppressive regimes but also for companies and politicians alike. The most widely discussed example of this is Cambridge Analytica, which exploited data of 87 million Facebook users to build profiles that could be utilized for political gain. But we need to look beyond Cambridge Analytica to understand the full exposure of our democracies and societies to AI.

Distortion of democracies, public discourse, social cohesion and public trust by AI cannot be pinpointed to one event, scandal nor even a single phenomenon. AI-driven computational propaganda has distorted elections in Ukraine, Estonia, China, Iran, Mexico, the UK, and the US. It has been estimated that during the 2016 US Presidential elections almost one-fifth of the Twitter discussions came from bots[1]. In 2017, two of the most widely followed black activist Twitter accounts, @Blacktivist and @WokeBlack, turned out to be fake accounts run by Russian troll farms. Facebook's algorithm has incited violence against protesters in Myanmar, when it promoted junta misinformation, death threats, and the glorification of military violence. More recently Facebook whistleblower Frances Haugen has said that the company's algorithm is "fanning ethnic violence" in Ethiopia[2].

These are just a few contemporary examples of how AI can impact our democracies and social cohesion, but we cannot even begin to fully grasp how AI could potentially shake democratic societies in future.

"Very large online platforms" (VLOPs), like Twitter, Facebook, Instagram, TikTok and Google, have a systemic role in our societies in shaping, amplifying, directing and targeting information flows online. The recommender systems designed by the VLOPs are crafted for their main clientele: advertisers. To maximize their profits, their business model is relatively simple: they must grab and maintain users' attention in order to maximize the time they spend on the platform. This raises the price of ad space and provides additional data on those users that can be used for further profiling and targeting. Stakeholders — even former supporters — of the largest technology firms have increasingly pointed out that they are designing AI-powered interactive products which systematically exploit human biases in the attempt to maximize profit and hence often come into conflict with fundamental rights and the common good[3].

Political micro-targeting, news aggregation and the amplification of information via the VLOPs utilizes these very same recommender systems. Obviously, a system designed for commercial purposes and profit maximization has entirely different implications when it is employed to 'sell' political ideas, news and opinions. The VLOPs "engagement maximization" model translates, when used in a political context, to recommending and amplifying extreme, hyper-partisan and radicalizing content. Algorithmic logic can create 'filter bubbles' on social networks, leading to polarization in society. As a consequence it jeopardizes user's access to pluralistic and objective information and may undermine shared understanding, mutual respect and social cohesion required for democracies to thrive. If AI-driven micro-targeting is very powerful and effective, it may even undermine the human agency and autonomy required for taking meaningful decisions. Moreover, AI is becoming more capable of producing media footage (video, audio, images) resembling real people's appearance and/or voice (also known as 'deep fakes'), enabling the deceptive practices for various purposes.

The European Parliament recently took a position on the DSA that VLOPs must provide at least one recommender system that is not personalized and that dark patterns will no longer be allowed [note]. While these are steps in the right direction, an amendment to make the non-personalized version the default recommender system, was voted down.

[1] Woolley, S. C., & Howard, P. N. (Eds.). (2018). *Computational propaganda: Political parties, politicians, and political manipulation on social media*. Oxford University Press.

[2] Facebook's role in Myanmar and Ethiopia under new scrutiny | Facebook | The Guardian

[3] McNamee, 2018; LaJeunesse, 2020; Haugen 2021.

2.1.2 **The limitations of transparency and consent**

The Commission also proposed a Regulation on Transparency and Targeting of Political Advertising (TTPA), in an attempt to make these practices more transparent. The TTPA however completely relies on transparency and consent measures, which have already been shown to be insufficient and even inadequate. Obtaining true informed consent has shown to be virtually impossible. In a world where vast amounts of data are collected, combined and reshuffled in the most creative ways, one could not reasonably be expected to oversee all potential future uses and consequences of such consent.

This goes even more for the transparency solution proposed in the TTPA. Apart from acquiring consent, the party doing the political micro-targeting needs to provide additional information on why someone is being targeted, which data was used, the logic involved in the decision making process, the parameters of the AI technique, etc. Other AI-related rules, such as the Platform Directive proposal and the GDPR, include similar transparency and consent measures. They seem to be the holy grail of protecting our online lives. Transparency and consent measures however place the burden of deciding whether someone is willing to be profiled and micro-targeted and how, entirely on the recipient. It has led to slippery slope developments of ever more shortcuts to push towards consent, for instance by making cookie banners ever more confusing, annoying and manipulative. The obligation of transparency is often denied or executed miserably, for example by providing incomprehensible and lengthy information. People should be able to do nothing and still be protected. In other words, not being tracked, profiled and micro-targeted should be the default mode.

2.1.3 **The gaps in current legislation**

A large part of AI-driven manipulation consists of profiling or categorization based on extensive tracking, collection and AI-driven processing of information, and the GDPR indeed deals with profiling. However, it does so predominantly through transparency and consent measures, the limitations of which are described in 2.1.1. Moreover, the GDPR considers automated decision making conditional to profiling. In other words, it only deals with profiling that is part of or leads to an automated individual decision which produces legal effects or significantly affects people. One could question whether the AI-driven manipulation described here would always be covered by this, for example when the manipulation is aimed at a large group of people, having a broader societal effect rather than a direct individual effect.

According to the Charter of Fundamental Rights of the European Union (the “Charter”) and the Ethics Guidelines for Trustworthy AI, every human being possesses an “intrinsic worth”, which should never be diminished, compromised or repressed by others – nor by new technologies like AI. This means that all people are to be treated with respect, as moral subjects, rather than merely as objects to be surveilled, sifted, sorted, conditioned or manipulated. While the AIA aims to protect all fundamental rights, in particular this right to human dignity, but also the right to receive or impart information and the right to privacy, are not sufficiently protected by secondary legislation, when it comes to AI-driven manipulation.

2.4 The AIA's grand opportunity

The powerful effects of AI-manipulation are currently not sufficiently addressed and cannot be curbed by merely imposing transparency measures on private actors. We argue that the AIA provides a grand opportunity to address the legal gaps and the wider societal harms that AI-driven manipulation can bring. A prohibition of AI-practices aimed at deception, material distortion of behavior or exploitation of a person's vulnerabilities would fit well within the larger objective of the AIA.

We propose a new Article 5.1 (a) which is a combination of paragraphs (a) and (b), and includes deep fakes as well as the broader impact on groups of people, democracy and society. We acknowledge that this new prohibition could lead to new discussions on whether such a prohibition is practicable and enforceable. While this will indeed be a challenge, we reify that legislation holds many legal norms that are open to interpretation and pose enforceability challenges. This should however not hold legislators back, as we deem it more important that the AIA halts the current trajectory towards a full "Internet-of-Minds".

Article 5

1. The following artificial intelligence practices shall be prohibited:

(a) The placing on the market (...) of an AI system deployed, aimed at or used for deception or materially distorting a person's behavior or exploit a person's vulnerabilities, in a manner that causes or is likely to cause harm to:
   (i) that person's, another person's or group of persons' fundamental rights, including their physical or psychological health and safety, and/or
   (ii) democracy, the rule of law, or society at large

(b) DELETE

RECOMMENDATION

3 SOCIAL SCORING

The proposed prohibition of ‘social scoring’ of article 5.1 (c) is definitely a step in the right direction. The current wording however could leave the door open to a number of social scoring mechanisms that also deserve stricter scrutiny.

First of all, it remains unclear what would be considered 'the trustworthiness of people’. AI used to determine creditworthiness for example, is considered high risk and thus allowed under conditions, but currently entails practices that assess the broader notion of trustworthiness as well. We have seen many other instances of social scoring by public and private actors alike. While these do not consist of ‘generalized’ schemes of citizen scoring, examples of which we have seen in China, they do consist of the scoring of individuals for various purposes, such as social benefits fraud detection, loan or mortgage eligibility etc. The Slovenian Presidency compromise proposal already rightfully added ‘private actors’ to the prohibition.

Secondly, the condition that ‘the scoring leads to detrimental or unfavourable treatment’, will create legal uncertainty and place the burden of proof on the individual being scored. Deleting this condition and better describing what are considered unacceptable social scoring practices is preferable.

For ‘social scoring’ to be effectively prohibited in Europe, the AIA should broaden the prohibition while at the same time drawing a clearer line between what is considered ‘social scoring’ and what can be considered an acceptable form of evaluation for a certain purpose (provided such evaluation to be considered high-risk). While the AIA already attempts to draw this line, we think that the wording can be clarified. Where the information used for the evaluation is no longer relevant, reasonably related or proportionate to the goal of the evaluation, the scoring should be prohibited. Below we propose an additional definition and and alternative text for Article 5.1 (c) that clarifies this.

Article 3
Definitions

For the purpose of this regulation, the following definitions apply:

(45) ‘social scoring’ means the evaluation or categorisation of EU citizens based on their behavior or (personality) characteristics, where one or more of the following conditions apply:
- (i) the information is not reasonably relevant for the evaluation or categorisation;
- (ii) the information is generated or collected in another domain than that of the evaluation or categorisation;
- (iii) the information is not necessary for or proportionate to the evaluation or categorisation;
- (iv) the information contains or reveals special categories of personal data.

Article 5

1. The following artificial intelligence practices shall be prohibited:

(c) The placing on the market, putting into service or use of AI systems by or on behalf of (semi-)public authorities or by private actors for the purpose of social scoring.

RECOMMENDATION

4 BIOMETRIC IDENTIFICATION

The AIA bans real time remote biometric identification (with for example facial recognition technologies) for law enforcement (with some exceptions) and categorizes it as ‘high risk’ when used for other purposes. Below we dive deeper into what will be prohibited and what will remain allowed under the current text. We also address the Slovenian Presidency compromise proposals for this prohibition.

4.1 Conditions of the ban

The AIA puts some strict conditions on what kind of biometric identification it aims to prohibit: it must be ‘real time’, it must happen ‘remotely’ and in ‘publicly accessible spaces’, it must be aimed at ‘identification’ and it must be used ‘for the purpose of law enforcement’.

4.1.1 **The purpose of law enforcement**

The main condition for the ban to kick-in is the limitation to ‘the purpose of law enforcement’. Only law enforcement authorities (or thier formally entrusted bodies) are prohibited from using biometric identification to prevent, investigate, detect or prosecute crimes, to safeguard public security or to execute criminal penalties. The Slovenian presidency compromise proposal broadens the scope slightly by adding private actors acting on behalf of law enforcement authorities to the mix, but the prohibition remains limited to biometric identification for the purpose of law enforcement. A couple of exceptions are made for (i) the targeted search for victims of crime, (ii) the prevention of a specific, substantial and imminent threat to life or physical safety or a terrorist attack, and (iii) the execution of a European arrest warrant for specific offences.

First of all, exclusion (iii), calls for better wording, so as to refer to extradition activities only.

Article 5

1. The following artificial intelligence practices shall be prohibited:

(d) (...) unless and in as far as such use is strictly necessary for one of the following objectives:

(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, for the purpose of the execution of a European arrest warrant referred to in Article 2(2) of Council Framework Decision 2002/584/JHA62.

RECOMMENDATION

4.1.2 **Real time**

Another condition for the prohibition on biometric identification to kick in is the fact that it should be ‘real time’.

> (d) the use of **‘real time’** remote biometric identification systems (...)

Recital (8) explains that this means that the biometric identification must occur (near-)instantaneously or without any significant delay, for it to be prohibited. It clarifies that circumvention of this prohibition by inserting a slight delay between the capturing of footage and the actual biometric identification would not be possible.

Biometric identification that takes place in a truly ‘post’ manner, however, for example on footage that is already in a police database or that is collected from CCTV footage or private devices, does not fall within the scope of ‘real time’. The Commission argues that, as post-processing of biometric data is already authorized by Member States under art. 10 of the Law Enforcement Directive, the AIA can only deal with ‘real time’ biometric identification. In practice, this means that law enforcement authorities can keep deploying biometric identification on all footage it holds or captures, as long as it does so at a later stage.

4.1.3 **Remote**

The next condition for the prohibition to kick in is that the biometric identification should happen remotely.

> (d) the use of ‘real time’ **remote** biometric identification systems (...)

The AIA provides no specific definition or explanation of ‘remote’ so one needs to look at the meaning of the word ‘remote’ to interpret the meaning of this condition. Merriam-Webster defines remote as:

> *“being or relating to a means of collecting data about something (such as an object or an area) from a device (as by using radar or photography)"*

This begs the question whether there could be forms of ‘near’ biometric identification that would be excluded from the ban, such as for example the use of facial recognition in police body cams or the use of face scanners.

4.1.4 **Publicly accessible spaces**

The Another condition for the prohibition to kick in is the use in publicly accessible spaces.

> (d) the use of ‘real time’ remote biometric identification systems in **publicly accessible spaces** (...)

Recital (9) and Art. 3 (39) clarify that the notion of a publicly accessible space refers to any physical place that is accessible to the public, irrespective of whether the place in question is privately or publicly owned and regardless of conditions of access apply (such as a ticket for example). This means that spaces such as cinemas, theaters, shops and shopping centers could be deemed publicly accessible. It however also states that whether a given space is accessible to the public should be determined on a case-by-case basis, having regard to the specificities of the individual situation at hand. Moreover, homes, many government offices, workplaces and online ‘spaces’ are not covered while we think they should be.

4.2 What remains allowed?

Looking at all conditions combined, only a very narrow set of biometric recognition practices for an even narrower group of actors would effectively be banned under the AIA. The following biometric recognition practices will be considered high-risk, but they remain allowed:

In law enforcement:

- Biometric categorization/segregation
- Biometric assessment

In domains other than law enforcement:

- Biometric authentication
- Biometric identification
- Biometric categorisation/segregation
- Biometric assessment

4.2.1 **Biometric recognition under the GDPR**

The GDPR prohibits, in principle, the processing of biometric data for identification purposes. It however also holds notable exceptions to that prohibition, such as consent, substantial public interest, employment and public health. These exceptions obviously make the prohibition far less broad and effective.

4.2.2 **Biometric assessment**

The use of biometric recognition to assess micro-expressions, gate, (tone of) voice, heart rate, temperature, etc. is often presented as supposedly being able to determine or predict a person's character, personality traits, mental state, emotions, behavior, intentions, criminality, intelligence, and even leadership skills, future job performance and so on. These AI practices are commonly known under multiple definitions such as 'emotion recognition', 'affect recognition' and more specifically, 'deception recognition', 'behavior analysis' or 'sentiment analysis'.

Article 3 (34) AIA defines 'emotion recognition systems' as "AI systems that can identify or infer emotions or intentions" and categorizes these systems generally as medium-risk, with the exception of a few areas where they are categorized as high-risk.

It should be noted that no sound scientific evidence exists corroborating that a person's inner emotions or mental state can be accurately 'read' from a person's face, gate, heart rate, tone of voice or temperature, let alone that future behavior could be predicted by it. In a recent meta-study, a group of scientists concluded that AI-driven emotion recognition could, at the most, recognise how a person subjectively interprets a certain biometric feature of another person[4]. An interpretation does not align with how that person actually feels, and AI is just labelling that interpretation which is highly dependent on context and culture. Far-fetched statements, that AI could for example determine whether someone will be successful in a job based on micro-expressions or tone of voice, are simply without scientific basis. As many others have already argued, broad beneficial use for these kind of systems, is lacking.

[4] Barret et al. 2019

4.3 The impact, accuracy and efficacy of biometric recognition

Biometric recognition in all its shapes and forms, has a broad and deep impact on the right to privacy, but privacy discussions around AI currently tend to focus primarily on data privacy and the indiscriminate processing of personal (and non-personal) data. It should however be noted that, while data privacy is indeed an important element, the impact of AI on our privacy goes well beyond our data. The fundamental/human right to privacy encompasses the protection of a wide range of elements of our private lives, that can be grouped into three broad categories namely: (i) a person's (general) privacy, (ii) a person's physical, psychological or moral integrity and (iii) a person's identity and autonomy.

Biometric recognition creates a situation where we are (constantly) being watched, followed, categorized, assessed or identified. This creates a continuous impact on our (general) privacy and our physical, psychological and moral integrity. The psychological 'chilling' effect could cause people to feel inclined to adapt their behavior to a certain norm when they feel watched. This shifts the balance of power between an individual and the state or a private organization using facial recognition. In legal doctrine and precedent the chilling effect of surveillance can constitute a violation of private space, which is necessary for personal development and democratic deliberation. Even if our faces are immediately deleted after capturing, the technology still intrudes on our psychological integrity.

It should be noted that other fundamental rights can become vulnerable too, such as the freedom of opinion and expression. When the protection of 'group anonymity' no longer exists, if everyone in the group could potentially be recognised, this could lead to people no longer freely express themselves or partake in legitimate and peaceful demonstrations.

Moreover, if biometric recognition is only prohibited for public actors (or organizations acting on their behalf) such as law enforcement agencies, this is likely to result in a further concentration of technological expertise, data and power in the hands of the private sector, strengthening their already quite dominant position in our democratic space.

4.3.1 **Accuracy and efficacy of biometric recognition**

While our fundamental rights are not absolute, they can only be surpassed after careful balancing of interests, circumstances, values and other fundamental rights, and then only if such surpassing is proportional and necessary. Obviously, no one principally opposes to measures that help find missing children or prevent a terrorist attack. But this is not the issue at stake. The issue at stake is whether biometric recognition helps us in such a way that we are willing to accept the intrusion of our fundamental rights, and if so, to what extent and under which conditions. In principle, the general rule should not be based merely on extreme and rare cases, whereas this leads to a situation where fundamental rights of everyone are limited only to prevent rare and extreme events, which could also be prevented by other, less intrusive means.

Rare and extreme cases can however form the basis of targeted and well defined exceptions. But this should be under the condition that the technology is indeed effective in those cases. This is an important notion, because we have seen ample examples of biometric recognition being technically flawed, for example when it comes to facial recognition having trouble recognizing people of colour, faces of women, or when operated in certain conditions (e.g. poor lighting, crowdedness, etc.).

But to properly answer the efficacy question we have to look at biometric recognition not only from a technical perspective, but appreciate its workings within the socio-technical environment in which it operates. Technical accuracy is important, but so is its effect on human interpretation and intervention as well as the wider sociological impact. A recent study by Fussey et al. (2020), that investigated several facial recognition trials by UK police departments, shows that, while facial recognition is demonstrably technological in character, human interpretation, and organisational and operational factors constitute key parts of the outcomes and all influence the level of efficacy of the use of facial recognition. The authors conclude that a considerable effort is still required to implement facial recognition in police operating procedures that have a reasonable degree of efficacy and efficiency.

4.4 An alternative legal approach to biometric recognition

Instead of banning only a narrow set of biometric identification practices and categorizing all others high and medium-risk, we echo the various calls (e.g. by EDPS, EDPR, EESC, IBM, and a majority of MEPs) for a ban on biometric recognition (which includes biometric identification, but also all forms of biometric categorization and assessment both by private organizations and (semi-)public authorities. The call for such ban also resonates more and more among the wider public.

If, in exceptional instances (such as for example those already mentioned in art. 5.1 (d) subparagraphs (i), (ii) and (iii), biometric identification is considered, efficacy (proportionality and necessity) should be well determined and established and measures should be taken to limit its use as regards time and/or scope. Also, such use should be subject to Title III of the AIA (high-risk AI systems). Moreover, the risk of misuse where these systems need to be in place, even if they are only ‘switched on’ only in certain circumstances, should be effectively mitigated. Another situation where some of these AI practices could bring benefits is in controlled environments such as for example hospitals, where the technology could serve a scientific purpose. Here also, we warn that these uses should at a very minimum be evidence based, limited in time, proportionate and necessary.

Biometric verification or authentication, i.e. the verification of an individual’s identity by comparing biometric data of that individual against existing biometric data of that same individual (one-to-one matching), remains allowed. For these uses, one could think of the opening of a device or the door of a lab by using a biometric identifier. Also here, however, it remains crucial to make sure that these uses do not allow for circumvention or create loopholes.

We propose amending various definitions and improving the prohibition of biometric identification (in particular as regards the elements of biometric categorization and assessment) that would also covers the indiscriminate on- and offline biometric tracking of our entire lives, behavior, interests, activities, locations, and so on. The GDPR and the proposals for the Digital Services Act proposal (“DSA”) only partially deal with this. These (proposed) regulations moreover heavily rely on transparency and end-users’ consent, which only provides a limited protection of our fundamental rights.

Amended definition and prohibition of biometric recognition:

Article 3
Definitions

(34) 'biometric recognition' means the use of AI-systems for the purpose of the automated recognition of physical, physiological, behavioral and psychological human features such as the face, eye movement, facial expressions, body shape, voice, speech, gait, posture, heart rate, blood pressure, odor, keystrokes, psychological reactions (anger, distress, grief, etc.) for the purpose of:
  (i) verification of an individual's identity by comparing biometric data of that individual to stored biometric data of individuals in a database (one-to-many identification)
  (ii) categorization of individuals into clusters based on ethnicity, gender, political or sexual orientation, or other grounds on which discrimination is prohibited under Article 21 of the European Charter of Fundamental Rights; and/or
  (iii) assessment of a person's personality, traits, characteristics, behavior, intentions, beliefs or ideas

The definitions of 'emotion recognition system' (34), 'remote biometric identification system' (35), 'real time' (37) and 'post' (38) could be deleted as they would be included in the new definition.

Article 5

1. The following artificial intelligence practices shall be prohibited:

(d) the placing on the market, putting into service or use of biometric recognition systems in publicly and privately accessible on- and offline spaces, with the exception of biometric recognition as described in article 3 (34 (new)) (a) used in 'real time', remotely and in publicly accessible spaces for the purpose of law enforcement as far as such use is strictly necessary for one of the following objectives and under the conditions of Title III:
  (i) [*the text of art. 5.1 AIA continues as from (d) sub (i), with the amendment of sub(iii), see paragraph 4.1.4 above*]

5. Biometric recognition shall be allowed for healthcare purposes, as far as such use takes place under strict conditions including the conditions of Title III, is evidence based, and is in line with the principles of responsible research and innovation.

RECOMMENDATION

ABOUT ALLAI

ALLAI is an independent organisation that aims to foster, promote and achieve the responsible development, deployment and use of AI.

ALLAI's mission is to take a holistic approach to AI, taking into account all impact domains such as economics, ethics, privacy, laws, safety, labour, education, etc. ALLAI aims to involve all stakeholders in its mission: policy-makers, industry, social partners, consumers, NGOs, educational and care instructions, academics from various disciplines.

ALLAI was founded by the three Dutch members of the High Level Expert Group on AI, Catelijne Muller, LLM, Prof. Virginia Dignum and Prof. Aimee van Wynsberghe.

ALLAI refers to Stichting ALLAI Nederland, a foundation under Dutch Law. No entity or person connected to ALLAI, including its Board Members, Advisory Board Members, employees, experts, volunteers and agents, is responsible or liable for any direct or indirect loss or damage suffered by any person or entity relying wholly or partially on this communication.

CONTACT

ALLAI
Prinseneiland 23A
1013 LL Amsterdam
The Netherlands

www.allai.nl
@ALLAI_EU
welkom@allai.nl