ARTIFICIAL INTELLIGENCE ACT

ANALYSIS & RECOMMENDATIONS

Catelijne Muller
Virginia Dignum

ALLAI.

ALLAI.

August 6, 2021

Submission to the public consultation on:

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

For (press) enquiries:
welkom@allai.nl

TABLE OF CONTENTS

Introduction

In its highly anticipated legislative proposal for AI, the European Commission projects a clear message: fundamental rights and European values are at the core of Europe's approach to AI. Europe is basically saying: when it comes to this technology, 'anything goes' is no longer the norm. We will not allow everything, just because it can be done. And we don't just regulate bits and pieces of the technology, we set EU wide rules that will resonate across the globe. ALLAI welcomes the ambitious proposal for a Regulation for Artificial Intelligence (also referred to as the Artificial Intelligence Act ("AIA")). We particularly welcome the fact that the European Commission took up the courage to actually prohibit certain AI practices and that it draws heavily on the Ethics Guidelines for Trustworthy AI.

By Catelijne Muller and Virginia Dignum

Over the past couple of years AI has found its way into our societies in many forms and shapes. It has shown promise, but also posed serious risks to health, safety and fundamental rights. Apart from the many examples where AI adversely impacts society by interfering with our (fundamental) rights, ethical principles and social values, there are more and more signs that the current state of the art of AI does not live up to the 'human like' capabilities that are attributed to it. In many instances AI has overpromised and underdelivered.

ALLAI welcomes the fact that the Commission proposal for the Artificial Intelligence Act (the "AIA") not only addresses the risks associated with AI. The combination of prohibited AI practices, requirements for high and medium-risk AI plus the measures to promote trustworthy AI innovation including the voluntary compliance model, will help improve the quality, performance and trustworthiness of AI and improve the chances that AI can live up to its promises, so that the benefits of AI can be reaped and the risks be mitigated.

This paper contains an analysis of the main elements of the AIA and, where relevant, proposes textual or conceptual changes.

Amsterdam/Umeå, August 6, 2021

1. EXECUTIVE SUMMARY

While we welcome the AIA, we do see areas for improvement of the text in its current form; in particular regarding the scope and definition, the clarity of the prohibited AI practices, the implications of the categorisation choices made in relation to the 'risk pyramid', the risk-mitigating effect of the requirements for high-risk AI and the relation to existing regulation and other recent regulatory proposals. Below is a summary of the most important areas for improvement.

ALLAI welcomes the fact that the AIA puts health, safety and fundamental rights at the centre of the AIA. We also welcome the external effect of the AIA, making sure that AI that is developed outside of the EU, has to meet the same legal standards if deployed or having an impact within the EU.

For any law to be effective, there must be a clear legal definition of what it is, it attempts to regulate. While the proposal aims to regulate AI-practices rather than the technology itself, it focuses heavily on AI-technologies or methods, and as such runs the risk of organisations evading the regulation, simply by classifying their applications differently. It is more effective to focus on the characteristics, or properties of a system, that are relevant to be regulated.

The AIA lacks more general notions such as the prerogative of human decision making, the need for human agency and autonomy, the strength of human-machine collaboration and the full involvement of stakeholders.

AI does not operate in a lawless world and the AIA should be clear(er) on the fact that existing laws and regulations (beyond the GDPR) apply to AI and the way we use it.

The two first prohibitions centre around 'distorting a person's behaviour', but in their current form only capture rare cases. They however provide a grand opportunity to address one of the most worrying and widespread capabilities of AI: harmful conditioning and manipulation.

It is important that the AIA halts the current trajectory of public and private actors using ever more information to assess, categorise and score us. AIA should attempt to draw a clear line between what is considered 'social scoring' and what can be considered an acceptable form of evaluation for a certain purpose, for example at the point where the information used for the assessment is not reasonably relevant or related to the assessment.

Broadly in line with the EDPB and EDPS, ALLAI calls for a ban on biometrics recognition (which includes biometric identification, but also all forms of 'emotion/behaviour/affect/intent/trait recognition with biometric recognition, which is now categorised as medium-risk and high-risk in some areas) both by private organisations and by or on behalf of (semi-)public authorities.

ALLAI thinks that the ubiquitous tracking of our entire behaviour through our online behaviour, our location data and our IoT data serves no obvious social benefit. While prohibiting these practices might be challenging, ALLAI believes that the further proliferation of these forms of widespread tracking of our entire lives by public and private actors should at least be curbed.

The high-risk approach, consisting of criteria for 'risk of harm to health, safety and fundamental rights', a selection of high-risk AI that supposedly serve a social benefit and a number of requirements that are aimed at mitigating their risks, can normalise and mainstream quite a number of AI practices that are still heavily criticised, often due to their lack of sufficient social benefit. Moreover, this approach assumes that the risks high-risk AI systems pose, also future ones, can be sufficiently mitigated by the requirements, which unfortunately is not always the case.

The AIA lays down the criteria for what can be considered a 'risk of harm to health, safety and fundamental rights'. This limits the broad interpretation of our fundamental rights framework, that allows for the consideration of all relevant circumstances (or 'criteria').

Analysis of the high-rise AI uses on the stand-alone high-risk AI list of ANNEX III shows that the benefits of a number of them not necessarily outweigh the risks and that for some, clear benefits are simply lacking. Moreover, the risks these uses pose cannot necessarily always be sufficiently mitigated by the current requirements for high-risk AI. We also think that a number of AI uses such as in election and voting processes and for content moderation should be added to ANNEX III.

We welcome the fact that the requirements for high-risk AI strongly reflect the requirements for trustworthy AI of the Ethics Guidelines for Trustworthy AI. A number of requirements are however not included in the AIA, which we think is a missed opportunity. (i) Human agency (ii) privacy (iii) diversity, non-discrimination and fairness, (iv) explainability and (v) environmental and social well being are particularly important to avoid the risks AI poses, which are those of privacy, bias, exclusion, inexplicability of the outcomes of AI decisions, the undermining of human agency and the environment.

If one takes a deeper look at the high-risk AI uses of ANNEX III, one will see that once the requirements for high-risk AI are met, AI can potentially largely replace human decision making in law enforcement, the judiciary, enjoyment of essential services, migration and asylum, recruitment, hiring, firing and worker-assessment, education, democratic processes. Most AI is technically incapable of making these decisions at a human level and it is questionable if it is even desirable. Ultimately, not all decisions can or should be simplified to 'ones and zero's'.

In line with our long advocated **'human-in-command'** approach to AI, ALLAI strongly recommends the AIA to arrange for certain decisions to remain the **prerogative of humans**, particularly in domains where these decisions have a strong moral component as well as legal implications or a societal impact such as in the judiciary, law enforcement, social services, healthcare, housing, financial services, labour relations and education.

2. OBJECTIVE SCOPE & DEFINITION

ALLAI welcomes the fact the AIA puts health, safety and fundamental rights at the centre of the AIA. We also welcome the external effect of the AIA, making sure that AI that is developed outside of the EU, has to meet the same legal standards if deployed or having an impact within the EU. We do however see limitations in the chosen instrument and the definition of AI.

Objective

There are a number of challenges but also opportunities when regulating AI from a human (or fundamental) rights perspective.

To date, there here has been much focus on the the impact of AI on the human rights regarding non-discrimination and (data) privacy, but many other human rights are at play with AI, and sometimes more than one at the same time.[1] AI Regulation should make sure to address this.

The fact that the human rights impact of AI is often 'hidden' or unknown and only discovered 'after the fact' necessitates AI regulation to contain ex ante, ex Durante and ex post assessment and monitoring obligations.[2] AI can have both a negative and a positive impact on the same or different human rights at the same time[3]. This requires AI regulation to leave room to balance these impacts.

Many AI applications are developed and/or used by private actors or used in 'private settings' (workplaces, shops, cars, the internet), who cannot be held directly to account where it comes to human rights violations. In the same manner as the GDPR, AI regulation bring the human rights protections into these private spheres as well.

Scope

The choice to deal with a complicated, general purpose technology, that impacts virtually all aspects of society, by way of a product regulation, makes the proposal very complex and hard to

[1] The overview of human rights at play with AI contains multiple citations of the paper by C. Muller for the CAHAI: "The impact of AI on Human Rights, Democracy and the Rule of Law"
[2] Ibid.
[3] Ibid.

navigate. In this attempt, it 'normalises' and likely even legitimises a number of highly invasive AI-systems and uses and it leaves a number of risks unaddressed.

On a higher level, the proposal lacks more general notions such as the prerogative of human decision making, the need for human agency and autonomy, the strength of human-machine collaboration and the full involvement of stakeholders. AI is not merely a technology neither is it just the sum of its software components. It comprises the entire socio-technical system around it. Limiting the legal framework to a 'product regulation' ignores these important notions.

The Ethics Guidelines for Trustworthy AI state that "AI systems do not operate in a lawless world. A number of legally binding rules at European, national and international level already apply or are relevant to the development, deployment and use of AI systems today. Legal sources include, but are not limited to: EU primary law (the Treaties of the European Union and its Charter of Fundamental Rights), EU secondary law (such as the General Data Protection Regulation, the Product Liability Directive, the Regulation on the Free Flow of Non-Personal Data, anti-discrimination Directives, consumer law and Safety and Health at Work Directives), the UN Human Rights treaties and the Council of Europe conventions (such as the European Convention on Human Rights), and numerous EU Member State laws. Besides horizontally applicable rules, various domain-specific rules exist that apply to particular AI applications (such as for instance the Medical Device Regulation in the healthcare sector)." This is why the Ethics Guidelines for Trustworthy AI specifically mention "lawfulness" as the first pillar of trustworthy AI.

To clarify this, ALLAI recommends to amend Recital (41) or even be insert into the body of the regulation:

Whereas:

(…)

(41) This Regulation should not be interpreted as indicating that the use of any AI-system is necessarily lawful under other acts of Union law or under national law compatible with Union law. Any use of AI should continue to comply with the the European Charter on Fundamental Rights, secondary Union law and national law. This Regulation should not be understood as providing the legal ground for unlawful AI development, deployment or use.

Definition

For any law to be effective, there must be a clear legal definition[4] of what it is, it attempts to regulate. While the proposal aims to regulate AI-practices rather than the technology itself, it contains a technical definition of AI in art. 3 jo. ANNEX I, the latter listing a number of broad "AI-techniques and approaches" that a system should incorporate to fall within the scope of art. 3 sub (1). Moreover for any of these techniques to fall within the scope of the definition, it should be able to (paraphrased) "generate outputs that influence the environments they interact with, when given a set of human-defined objectives".

[4] A legal definition, in particular when regulating such a complicated technology with such broad and deep impact, should meet a number of requirements such as inclusiveness, preciseness, comprehensiveness, practicability and permanence.

Listing AI-techniques could easily create confusion, legal uncertainty and loopholes. First of all, the list of AI-techniques and approaches (ANNEX I) it lacks a number of relevant AI-techniques and approaches such as decision trees, random forests, fuzzy logics, game theory etc. More importantly however, AI-techniques and approaches (and their names for that matter) are constantly evolving and new techniques are being developed as we speak. Moreover, while the regulation should of course be interpreted based on its spirit rather than its letter, adding a list of AI-techniques and approaches limits the room for such interpretation and opens the door to '*a contrario*' reasoning, where it is argued that any AI-technique that is not listed, falls outside of the scope of the AIA.

The Commission does retain the right to update the definition and AI-techniques by delegated act, but can only do so when these new techniques have similar characteristics as the ones already on the list. While this approach promotes flexibility in lawmaking, we also fear that it could lead to a game of 'whack a mole' for the Commission along with discussions and disagreements of what is considered to be a 'similar' AI-technique.

It is better to focus on the characteristics, or properties of a system, that are relevant to be regulated. By focusing on technologies, or methods, i.e. by regulating systems that are based on 'machine learning, logic, or statistical approaches', we run the risk of organisations evading the regulation, simply by classifying their applications differently.[5]

Borrowing from the definition given in the seminal textbook on AI, AI is the discipline of developing computer systems that are able of perceiving its environment, and to reason about how to best act on it in order to achieve its goals, assuming that the environment contains other agents similar to itself. As such, AI is about the *autonomy* (or better automation) to decide on how to act, the *adaptability* to learn from the changes affected in the environment, and the *interactivity* required to be sensitive to the actions and aims of other agents in that environment, and decide when to cooperate or to compete.[6]

A responsible, ethical, approach to AI goes further than the technology used, and needs to include the social, organisational and institutional context of that technology. It is this socio-technical that needs to ensure *transparency* about how adaptation is done, *responsibility* for the level of automation on which the system is able to reason, and *accountability* for the results and the principles that guide its interactions with others, most importantly with people. In addition, and above all, a responsible approach to AI makes clear that AI systems are artefacts manufactured by people, for some purpose, and that people are responsible for the use and development of AI.[7] We propose to consider the following alternative definition of AI, and removing ANNEX I from the AIA entirely:

Article 3
Definitions

1. *For the purpose of this Regulation, the following definitions apply:*

(1) *"'Artificial intelligence system' (AI system) means software that can, in an automated manner, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations or decisions, influencing the environment it interacts with."*

[5] https://www.linkedin.com/pulse/what-we-need-talk-when-ai-regulatory-purposes-virginia-dignum
[6] Ibid.
[7] Ibid.

3.
PROHIBITED AI PRACTICES

The escalating 'risk pyramid' (from limited and medium-risk, to high-risk, to unacceptable risk) used to categorise a number of general AI practices as well as a number of domain specific AI use cases, acknowledges that not all AI poses risks and not all risks are equal. The descriptions of the various prohibited AI practices and medium and high-risk AI use cases, however, are at times unclear, multi-interpretable and could lead to legal uncertainty and create loopholes. Moreover, ALLAI questions a number of categorisation choices made and sees a great number of still heavily criticised AI uses listed as medium and high-risk, potentially further mainstreaming them.

Materially distorting behaviour

It is not entirely clear what it is the AIA aims to prohibit in article 5.1 paragraphs (a) and (b). What stands out is that they centre around 'distorting a person's behaviour' and that the distorted behaviour should lead to physical or psychological harm for the prohibition to kick in. While these prohibitions likely cover a very narrow set of AI practices, they provide a grand opportunity to address one of the most worrying and widespread capabilities of AI: harmful conditioning and manipulation.

With regards to the right to receive and impart information and ideas, AI used in media and news curation, bringing ever more 'personalised' online content and news to individuals, raises concerns. Search engines, video recommendation systems and news aggregators often are opaque, both where it comes to the data they use to select or prioritise the content, but also where it comes to the purpose of the specific selection or prioritisation.[8] Many business models are based on online advertising revenue. In order to have people spend as much time on a platform or website as possible, they might be selecting and prioritising content that will do only that: keep people on their platform, irrespective of whether this content is objective, factually true, diverse or even relevant.

Beyond commercial motives, political or other motives might lead to AI-systems being optimised to select or prioritise particular content in an effort to coerce and influence individuals towards certain points of view, for example during election processes.[9]

---

[8] Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, 3(1), 2053951715622512.

[9] Cambridge Analytica, Netflix Documentary: The Great Hack

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.

All this can give rise to filter bubbles and proliferation of fake news, disinformation and propaganda, and affects the capacity of individuals to form and develop opinions, receive and impart information and ideas and thus impact our freedom of expression.[10]

According to 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, scored, herded, conditioned or manipulated.[11]

In order to capture what the AIA should prohibit, which is condition, manipulate and exploit people into harmful behaviour, ALLAI recommends to combine the two current prohibitions under article 5.1 (a) and (b) as follows:

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 materially distorting a person's behaviour or exploit a person's vulnerabilities, in a manner that causes or is likely to cause harm to that person's, another person's or group of persons' fundamental rights, including their physical or psychological health and safety, or to democracy and the rule of law*

(b) *DELETE*

Social scoring

ALLAI welcomes the prohibition of 'social scoring' of article 5.1 (c). We do however feel that the prohibition leaves the door open to a great number of social scoring mechanisms that in our opinion deserve scrutiny.

First of all, in the current wording the scoring should be aimed at evaluation or classification of trustworthiness of people. While we like the term trustworthiness for obvious reasons, in this context it is vague. What is considered the trustworthiness of a person? AI used to determine creditworthiness, an element of trustworthiness, is considered high-risk and thus allowed. Secondly, the existence of "detrimental or unfavourable treatment", which is a requirement in the current prohibition, will be difficult if not impossible, to prove.

We have seen many instances of social scoring by public and private actors alike. While they do not consist of 'generalised' schemes of citizen scoring, examples of which we have seen in China,

[10] UN Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, A/73/348

[11] Ethics Guidelines for Trustworthy AI, High Level Expert Group on AI (2019)

they do consist of the scoring of people for various purposes, such as social benefits fraud detection, loan or mortgage eligibility etc.

For 'social scoring' to be effectively prohibited in Europe, the AIA should attempt to draw a clear line between what is considered 'social scoring' and what can be considered an acceptable form of evaluation for a certain purpose. ALLAI believes that this line can be drawn where the information used for the evaluation should no longer be deemed relevant or reasonably related to the goal of the evaluation. We realise that this will lead to discussions as to what is relevant or reasonably related, but we do not see this as a problem, as many legal norms are open to interpretation. We do however feel that it is important that the AIA halts the current trajectory of using ever more information to assess our every move.

A solution could be for the AIA to incorporate a definition of social scoring and adjust the prohibition as follows:

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 behaviour or (personality) characteristics, where one or more of the following conditions apply:
- *the information is not reasonably relevant for the evaluation or categorisation;*
- *the information is generated or collected in another domain than that of the evaluation or categorisation;*
- *the information is not necessary for or proportionate to the evaluation or categorisation;*
- *the information contains special categories of personal data, including biometric 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.

Real time biometric recognition for law enforcement

The AIA bans real time remote biometric identification (with for example facial recognition) for law enforcement and categories it as 'high risk' when used for other purposes, except in some specified circumstances.

The AIA puts some strict limitations on what kind or biometric recognition it aims to prohibit, whereas it should:
- is '*real time*'

- *is 'remote'*
- *'identify'*
- *'used for the purpose of law enforcement '*

This leaves 'post' and 'near' biometric recognition allowed. It also leaves biometric recognition not aimed at identifying a person, but rather assessing a person's behaviour from their biometric features (micro-expressions, gait, temperature, heart rate, etc.) allowed. The limitation to 'law enforcement' allows biometric identification, as well as al other forms of biometric recognition not aimed at identification of an individual, including all mentioned forms of 'emotion recognition' for all other purposes, by all other actors, in all public and private places, including at the workplace, shops, stadiums, theatres etc. (as long as it does not serve law enforcement purposes.) This leaves the door wide open to a world where we are constantly being 'emotionally assessed' for whatever purpose the actor assessing us deems necessary.

Biometric recognition has a broad and deep impact on the right to privacy. 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. Art. 8 of the ECHR 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.[12] Different applications and uses of AI can have an impact on these categories, and have received little attention to date.

Facial recognition, involves the capture, storage and processing of personal (biometric) data (our faces)[13], but it also affects our 'general' privacy, identity and autonomy in such a way that it creates a situation where we are (constantly) being watched, followed and identified. As a psychological 'chilling' effect, people might feel inclined to adapt their behaviour to a certain norm, which shifts the balance of power between the state or private organisation using facial recognition and the individual.[14] In legal doctrine and precedent the chilling effect of surveillance can constitute a violation of the private space, which is necessary for personal development and democratic deliberation.[15] Even if our faces are immediately deleted after capturing, the technology still intrudes our psychological integrity.

Other forms of AI-driven biometric recognition have an even greater impact on our psychological integrity. Recognition of micro-expressions, gate, (tone of) voice, heart rate, temperature, etc. are currently being used to assess or even predict our behaviour, mental state and emotions.

These AI practices are commonly known under multiple definitions such as 'emotion recognition' or 'affect recognition' and more specifically as 'deception recognition' or 'behaviour recognition', but are also often presented as supposedly being able to predict a person's personality traits and even leadership skills, future job performance etc. The AIA only defines 'emotion recognition' (as "AI systems that can identify or infer emotions or intentions") and categorises these systems generally as medium-risk, with the exception of a few areas where they are categorised as high-risk.

---

[12] Guidance to art. 8 ECHR, Council of Europe.

[13] The jurisprudence of the European Court of Human Rights (ECtHR) makes clear that the capture, storage and processing of such information, even only briefly, impacts art. 8 ECHR.

[14] Examined Lives: Informational Privacy and the Subject as Object, Julie E. Cohen, 2000.

[15] The chilling effect describes the inhibition or discouragement of the legitimate exercise of a right. It has been shown that once people know that they are being surveilled they start to behave and develop differently; Staben, J. (2016). Der Abschreckungseffekt auf die Grundrechtsausübung: Strukturen eines verfassungsrechtlichen Arguments. Mohr Siebeck.

It should be noted upfront 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 behaviour could be predicted by it. In a recent meta-study, a group of scientists[16] concluded that AI-driven emotion recognition could, at the most, recognise how a person subjectively *interprets* a certain biometric feature of another person. 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.

The GDPR restricts the processing of biometric data only to some extent. Biometric data according to the GDPR is "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. The last part of the sentence is crucial, because if biometric recognition is not aimed at identification (but for example at categorisation, profiling or affect recognition), it might not fall under the GDPR-definition. In fact, recital 51 of the GDPR says that 'the processing of photographs [is considered] biometric data only when processed through a specific technical means allowing the unique identification or authentication of a natural person.' Many biometric recognition technologies are not aimed at processing biometric data to uniquely identify a person, but merely to assess a person's behaviour (for example in class) or to categorise individuals (for example for the purpose of determining their insurance premium based on their statistical prevalence to health problems). These uses might not fall under the definition of biometric data (processing) under the GDPR.

Apart from the right to privacy, biometric recognition also affects other human rights and freedoms 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 recognise, this could lead to people no longer partaking in peaceful demonstrations.[17]

In stead of categorising these AI practices as high and medium-risk, (as the AIA now does), we strongly call for a blanket ban on biometrics recognitions (which includes biometric identification, but also all forms of 'emotion/behaviour/affect/intent/trait recognition) both by private organisations and (semi-)public authorities.

In the exceptional instances already mentioned in art. 5.1 (d) subparagraphs (i), (ii) and (iii), biometric identification might be considered acceptable, however, there is a risk of misuse because the systems need to be in place, even if they are only allowed to be 'switched on' in these circumstances. In any event, the conditions set in the same article should apply. Another situation where some of these AI practices could be allowed is in controlled environments such as for example hospitals, where the technology could serve a scientific purpose (i.e. predicting autism, Alzheimer's etc.). Here also, we warn that these uses should at a very minimum be evidence based, limited in time, proportionate and necessary.

---

[16] Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. Psychological Science in the Public Interest, 20(1), 1–68.
[17] Privacy Impact Assessment Report for the Utilization of Facial Recognition Technologies to Identify Subjects in the Field, 30 June 2011, p. 18.

Broadly in line with the call of the EDPS and EDPR[18], ALLAI calls for the following amendments of article and additions to 5 AIA:

Article 5

1. *The following artificial intelligence practices shall be prohibited:*

   (d) 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 automated recognition of human features in publicly and privately accessible spaces, such as recognition of faces, gait, fingerprints, DNA, voice, keystrokes and other biometric or behavioural signals, in any context

   (e) A ban on AI systems using biometrics to categorise 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;

   (f) A ban on the use of AI to infer emotions, affect, behaviour or intent, of a natural person, except for very specified cases, such as some health purposes, where the patient emotion recognition is important.

New prohibition - Indiscriminate on- and offline tracking

Many public and private actors partake in AI-driven online on- and offline tracking of our entire behaviour, interests, activities, locations, and so on. This is not only done through our online behaviour, such as our social media channels or web visits and search queries, but also through our IoT data and location data. Also this creates a world where are constantly being followed, tracked, profiled, categorised, analysed, sifted and sorted. We already argued for a prohibition on the use of unrelated information to, for example, detect welfare fraud or to receive benefits, a credit, a mortgage, a loan, a job, a promotion, etc., as we find this unacceptable forms of 'social scoring'. This however still leaves room to track us for many other purposes. The GDPR and the proposals for the Digital Services Act proposal ("DSA") and the Digital Markets Act ("DMA") only partially deal with this where it comes to private actors. The European Parliament[19] and the European Data Protection Supervisor[20] have already urged to regulate advertising "more strictly in favour of less intrusive forms of advertising that do not require any tracking of user interaction with content".

Amnesty International has asked to end the use of cross-site tracking, prohibit the use of special categories of data listed under Article 9 of the GDPR for ad targeting, and prohibit further uses of personal data that could expose people to discrimination (e.g. data to infer a person's social or financial situation or mood). The DSA and DMA propose elements of this, but rely on end-users to be provided a specific choice whether or not to consent with these practices. It is however questionable whether such consent will truly meet the GDPR's requirements of being freely given,

[18] https://edps.europa.eu/press-publications/press-news/press-releases/2021/edpb-edps-call-ban-use-ai-automated-recognition_en

[19] "European Parliament resolution of 20 October 2020 with recommendations to the Commission on a Digital Services Act: adapting commercial and civil law rules for commercial entities operating online (2020/2019(INL))"; "European Parliament resolution of 20 October 2020 with recommendations to the Commission on the Digital Services Act: Improving the functioning of the Single Market, (2020/2018(INL))"

[20] European Data Protection Supervisor, Opinion 1/2021 on the Proposal for a Digital Services Act, 10 February 2021

specific, informed and unambiguous. Current practice has shown that breaches of basic GDPR obligations are common and difficult to rectify once they occur and has led to lengthy legal proceedings.[21]

ALLAI thinks that these AI practices do not serve society and have no obvious social benefit. While prohibiting these practices might be challenging, ALLAI believes that the further proliferation of these forms of widespread tracking of our entire lives by public and private actors should at least be curbed.

Article 5

1. *The following artificial intelligence practices shall be prohibited:*

   (g) AI-driven indiscriminate surveillance and tracking by public and private actors of, including but not limited, online behaviour, location data and IoT-data.

[21] noyb, GDPR: noyb.eu filed four complaints over "forced consent" against Google, Instagram, WhatsApp and Facebook, May 2018

4.
HIGH-RISK AI

In deciding whether an AI practice or use that poses a risk to health, safety or fundamental rights should nevertheless be allowed under strict conditions, the AIA contains a 3 step approach: (i) it determines the criteria for 'the risk of harm by AI to health, safety and fundamental right's, (ii) it decides on those AI areas and uses that are high-risk but nevertheless pose social benefits and (iii) it sets a number of requirements that should mitigate this risk.

The chosen approach first and foremost poses the risk of normalising and mainstreaming quite a number of AI practices that are still heavily criticised, often due to their lack of sufficient social benefit.

Moreover, this approach assumes that the risks high-risk AI systems pose, also future ones, can be sufficiently mitigated by the requirements (which are (*paraphrased*: datasets of high quality, documentation, transparency, human oversight, accuracy, cybersecurity and robustness). This is however not always the case. Not for the fundamental rights that are most often mentioned as impacted by AI (the right to non-discrimination and the right to privacy), but neither for the many other fundamental rights of the EU Charter that can be impacted by AI.

But perhaps the most pertinent question this approach raises is: are we ready to allow high-risk AI to largely replace or heavily influence human decision making, even in critical processes such as law enforcement or the judiciary?

The criteria for 'risk of harm to health, safety and fundamental rights'

In art. 7.2 the AIA lays down the criteria to establish whether an AI system's use poses a 'risk of harm to health, safety and fundamental rights' and should be (and likely were) put on the high-risk list.

Codifying a number of criteria and leaving out others, prioritises certain circumstances over others, which might nevertheless be relevant. Setting the intended purpose of the AI system as a criterion, limits the possibility to consider the 'foreseeable use' of the system. Setting the 'ability of a system to affect a plurality of persons' as a criterion undermines the need to consider the effect AI can have only on a small number of persons or even one person. As such the provision limits the broad interpretation of our fundamental rights framework, that allows for the consideration of all relevant circumstances.

Stand-alone high-risk AI (ANNEX III)

Annex III (standalone high-risk AI) lists a number of AI uses in a total of 8 areas that are considered high-risk, but nevertheless bring enough social benefit to justify their use, provided that a number of requirements are met. A brief analysis of these uses shows that the benefits of a number of them not necessarily outweigh the risks and that for some, clear benefits are simply lacking.

1. BIOMETRIC IDENTIFICATION AND CATEGORISATION

As already argues above, we see no societal benefit in the broad acceptance of biometric identification and categorisation and call for moving this AI practice to the prohibitions.

2. MANAGEMENT AND OPERATION OF CRITICAL INFRASTRUCTURE

Missing from this 'area' are AI systems used in the management and operation of the telecom and internet infrastructure:

> 2. *Management and operation of critical infrastructure:*
>
> (a) *AI systems intended to be used as safety components in the management and operation of telecom, internet, road traffic and the supply of water, gas heating and electricity The management and operation of the telecom and internet infrastructure.*

3. EDUCATION AND VOCATIONAL TRAINING[22]

AI systems to determine access to education and evaluate students, in particular where these applications use biometrics and behaviour recognition, pose a number of risks of harm to student health, safety and fundamental rights.

Algorithmic grading used during the pandemic caused serious uproar in the UK as it incorrectly downgraded students having detrimental effects on their life opportunities.[23] Online proctoring tools track multiple things such as eye movement, mouse movement, key board strokes, sound scapes, copy and paste behaviour, online search behaviour, body movement, etc. to supposedly flag 'suspicious behaviour' and 'indications of cheating' during online exams. Scientific evidence of the effectiveness of these techniques in correctly indicating suspicious behaviour is generally absent, while students have experienced the use of these systems as truly invasive and having a negative effect on their ability to conduct the exams in a dignified manner.[24]

[22] EESC INT/940 Artificial Intelligence Act, Rapporteur Muller C. (2021)
[23] https://allai.nl/portfolio-item/algorithmic-grading/
[24] https://allai.nl/portfolio-item/online-proctoring/

4. EMPLOYMENT, WORKERS MANAGEMENT AND ACCESS TO SELF-EMPLOYMENT[25]

The use of AI systems for monitoring, tracking and the evaluation of workers, has been causing serious concerns as regards workers' fundamental rights to work, to fair and just working conditions, to information and consultation and to justified dismissal. The addition of AI systems used in labour relations to the high-risk list is likely to cause conflicts with national labour laws aimed at (un)fair dismissal, healthy and safe working conditions and the right to worker information.[26] Particularly in labour relations, where there is a power imbalance, these types of systems should not be implemented without proper worker consultation and the involvement of social partners.

5. ACCESS AND ENJOYMENT OF ESSENTIAL PRIVATE SERVICES AND PUBLIC SERVICES AND BENEFITS[27]

Depending on the kind or information used, we consider the use of AI systems in relation to access and enjoyment of public and private services a form of unacceptable social scoring. For assessments with AI in these domains, the description of the use of AI systems in relation to the access and enjoyment of public services is more broad than the use of AI systems in relation to access and enjoyment of essential private services, where for the latter, only credit(worthiness) scoring by AI is considered high risk. We propose the following amendments:

5. *Access to and enjoyment of essential private services and public services and benefits:*
   (a) *AI systems intended to be used by or on behalf of (semi-)public authorities or private parties to evaluate the eligibility of natural persons for public assistance, benefits and services and essential private services, as well as to grant reduce, revoke, or reclaim such benefits and services;*
   (b) *DELETE*

6. LAW ENFORCEMENT

Many of the listed AI uses in the area of law enforcement pose a risk of harm to human dignity, privacy, the right to non-discrimination, the presumption of innocence, the right of defence and the right to a fair trial.[28]

Lie detection and emotion detection with biometric recognition is scientifically flawed and truly invasive. Many AI systems that make individual risk assessments merely seek correlations between characteristics found in other 'cases', on characteristics that a person happens to share with other convicted criminals (such as address, income, nationality, debts, employment, behaviour, behaviour of friends and family members and so on). Moreover, it is often impossible for legal professionals to understand the reasoning behind the outcomes of the system, which affects the right to the right to a fair trial. Predictive policing has led to undesirable feedback loops where the same communities are surveilled more often invading the right to privacy.

---

[25] Ibid.
[26] https://allai.nl/portfolio-item/home-office-surveillance/
[27] [27] EESC INT/940 Artificial Intelligence Act, Rapporteur Muller C. (2021)
[28] Open letter by 61 organisations calling for legal limits on AI risk assessment systems in the criminal justice context

We strongly recommend to consider strongly limiting the use of these types of AI uses by law enforcement. Exceptional uses could be considered, however these should be evidence based, necessary and proportionate, limited in time and have a legal basis.

7. MIGRATION, ASYLUM AND BORDER CONTROL MANAGEMENT

AI used in migration, asylum and border control management for making individual (criminal or security) risk assessments, pose a risk of harm to the presumption of innocence, the right of defence and the right to asylum. According to a recent paper by the EPRS[29]: "The EU's centralised information systems for borders and security[30] are increasingly incorporating biometric technologies for the purpose of identity verification or identification, automated fingerprint identification. Facial recognition is expected to be used by all systems except one in the near future. A number of EU-funded projects and initiatives have explored and piloted emotion recognition technologies at the EU border."

Art. 83 of the AIA however excludes AI systems which are components of these systems from its scope, if they are put into service before the application of the AIA. This enables EU's centralised information systems for borders and security to implement (to be) prohibited and high-risk AI systems in the near future, only having to comply with AIA when the entire system is evaluated.

We propose to move the AI uses listed in ANNEX III point 7 under (a) and (b) to the prohibitions.

8. THE ADMINISTRATION OF JUSTICE AND DEMOCRATIC PROCESSES[31]

The use of AI in the administration of justice and democratic processes is particularly sensitive and should be approached with more nuance and scrutiny than it is done now. Merely putting systems to assist a judicial authority in researching and interpreting facts and the law *and* in applying the law to a concrete set of facts overlooks the fact that judging is so much more than finding patterns in historical data (which is in essence what current AI systems do).

The text also assumes that these types of AI will only assist the judiciary, leaving fully automated judicial decision making out of scope.

While the 'header' mentions democratic processes as an area, there are no AI uses listed in the realm of elections. We recommend the following amendment to ANNEX III:

---

[29] "Artificial Intelligence at EU borders", EPRS https://www.europarl.europa.eu/RegData/etudes/IDAN/2021/690706/EPRS_IDA(2021)690706_EN.pdf

[30] These systems include (a.o.): Schengen Information System (SIS), the Visa Information System (VIS), Eurodac, the Entry/Exit System (EES), the European Travel Information and Authorisation System and the European Criminal Records Information System on third-country nationals and stateless persons (ECRIS-TCN)

[31] EESC INT/940 Artificial Intelligence Act, Rapporteur Muller C. (2021)

8. *Administration of justice and democratic processes:*
   (a) *AI systems intended to be used for or assist a judicial authority in researching and interpreting facts and the law and in applying the las to a concrete set of facts.*
   (b) *AI systems or uses in the realm of elections, voting and counting processes and election news aggregation and provision.*

MISSING HIGH-RISK AI DOMAINS/USES

Missing from the High-Risk list is AI used for content moderation:

X. *AI systems intended to be used for online content moderation (as described in articles 12, 26, 27 of the DSA)*

Harmonised products with AI

The 'interwoven system' between the AIA and Union Harmonisation Legislation is a smart approach from a lawmaking point of view, however, only AI safety components of a product, or AI that is itself a product, that is covered by Union Harmonisation Legislation listed in ANNEX II, are currently covered by the AIA. AI however not only can pose risks when used as safety components of these products, nor is the AI system itself always a product. When AI is used for example as part of a diagnostic or prognostic tool in the medical field or an AI driven thermostat that regulates a boiler (both covered by Union harmonisation legislation listed in ANNEX II), this should also be covered by the AIA.

ALLAI realises that this will cover many AI systems, also those that pose no risk whatsoever, but a case can be made that the fact alone that these products are covered by harmonised rules and already go through rigorous pre-market assessment, merits prior conformity assessment of all AI elements of these products as well.

Article 6
Classification rules for high-risk AI systems

1. *Irrespective (…)*

   (a) the AI system is intended to be uses as a ~~safety~~ component (…);

   (b) the product whose ~~safety~~ component is the AI system (…).

Delegated acts - limited to predefined areas

Art. 7 of the AIA allows for the Commission to add AI uses to the high-risk list by delegated act, but only in the 8 areas that are already listed in ANNEX III. This excludes the possibility to add risky AI uses in other areas. Moreover, amendments to ANNEX II are not possible under the AIA, while potential new Union Harmonisation Legislation for products with AI is feasible. We propose the following amendments:

Article 7

Amendments to ANNEX II and ANNEX III

1. *The Commission is empowered to adopt delegated acts in accordance with Article 73 ~~to update the list in~~ to make additions to the lists in ANNEX II and ANNEX III ~~by adding high-risk AI systems~~ where ~~both of the following~~ ~~conditions are fulfilled~~:*

   (a) DELETE

Do the requirements mitigate the risks?

ALLAI welcomes the alignment of the the requirements for high risk AI with elements of the Ethics Guidelines for Trustworthy AI (the "Ethics Guidelines"), however, five important requirements of the Ethics Guidelines are not specifically dealt with in the AIA namely: (i) human agency (ii) privacy (iii) diversity, non-discrimination and fairness, (iv) explainability and (v) environmental and social well being. ALLAI considered this a missed opportunity, because many of the risks AI poses are those of privacy, bias, exclusion, inexplicability of the outcomes of AI decisions, the undermining of human agency and the environment, and are all reflected in our fundamental rights.

While some of these missing requirements are (partially) dealt with in existing legislation such as the GDPR (privacy and explainability) and the European Charter on Fundamental rights (non-discrimination), this is not sufficient. Moreover, by adding the missing requirements, the AIA can bring many human rights protections into the private sphere as well.

The GDPR offers some frameworks for the right to an explanation of AI decisions (in the recitals)[32] and the right to human intervention. However, these frameworks are insufficient to ensure adequate explainability in automated decision-making. The right to human intervention in the GDPR for example only exists in the event of a fully automated decision, which is easy to circumvent. Moreover, the GDPR deals with situations where personal data is involved and it is well known that a decision based on non-personal data can also have a major impact.[33]

What is more important is that the AIA works from the premise that the requirements for high-risk AI will sufficiently mitigate the risks of harm to health, safety and fundamental rights. This is however expected not to be the case in all circumstances.

As for discriminatory outcomes for example, it overlooks the fact that AI biases are the result of low-quality data, lack of human oversight or inaccuracy. The design of any artefact is in itself an accumulation of choices that can cause biases to creep in. And even with high quality data, documentation, human oversight, accuracy, robustness and cybersecurity, AI can still pose risks. Often where it comes to less mentioned human rights such as freedom of expression, freedom of assembly, presumption of innocence, right of defence, right to fair trial etc.

Merely focussing on technical and procedural solutions does always not suffice to avoid this. Rather the socio-technical processes around AI need to be organised in such a way that that any

[32] Lokke Moerel and Marijn Storm, Law and Autonomous Systems Series, "Automated Decisions Based on Profiling - Information, Explanation or Justification, That is the Question!", Oxford University, Faculty of Law

[33] "Artificiele Intelligentie - Human in Command" ALLAI (2019)

discriminatory, unfair or otherwise harmful outcomes of AI can be adequately identified and addressed.

ALLAI proposes to add the 'missing' requirements from the Ethics Guidelines for Trustworthy AI to the requirements of Chapter 2 of Title III of the AIA, to improve the ability of the AIA to effectively protect our health, safety and fundamental rights from adverse impact of AI:

- *Human Agency (part of EGTAI requirement "Human Agency and Oversight")*
- *Privacy (part of EGTAI requirement "Privacy and Data Governance")*
- *Diversity, Non-discrimination and Fairness*
- *Explainability (part of EGTAI requirement "Transparency")*
- *Societal and Environmental Well-being*

Finally, the AIA is unclear as regards the situation where the requirements for high-risk AI are not met. Art. 67 of the AIA deals with "compliant AI systems which present a risk", but only after these systems have entered the market. The potential gravity of the adverse effects of high-risk AI, justifies a precautionary approach, rather than a post-hoc 'reparatory' approach. This can be arranged for by adding a paragraph to art. 16 AIA:

Article 16
Obligations of providers of high-risk AI systems

Providers of high-risk AI systems shall:

(h) refrain from placing a high-risk AI system on the market that:

(i) is not in conformity with the requirements set out in Chapter 2 of this Title; or

(ii) poses a risk fo harm to health, safety or fundamental rights despite its conformity with the requirements set out in Chapter 2 of this Title.

A FIDUCIARY DUTY FOR PROVIDERS AND USERS

The AIA could further mitigate the risks of high-risk AI by establishing a 'fiduciary duty' on AI providers and users to acknowledge, anticipate and protect the fundamental rights of those at the receiving end of the AI system, accompanied by an obligatory "human rights impact assessment". A known characteristic of high-risk AI is the "imbalance of power, knowledge, economic or social circumstances, or age" between those affected by the system and those designing or running it (see also art. 7.2 (f) AIA). It thus makes sense to place this special duty of care on providers and users of high-risk AI systems, where they prioritise the interests of those affected by the AI systems even over their own potential interests:

Article [x] (new)

The providers and users of high-risk AI systems have a fiduciary duty towards those affected by the the AI systems they are deploying or using to act in the interest of potentially harmed or adversely impacted persons.

INCLUSIVITY AND MULTI-DISCIPLINARITY

Another important tool to assess, address and mitigate potential risks of high-risk AI that is missing from the AIA is the involvement of relevant stakeholders, domain experts and representative bodies. We propose the following addition to art. 9 AIA:

Article 9
Risk management system

1. *(…)*
2. *The risk management (…) updating, continuously involving and consulting relevant stakeholders, including but not limited to domain experts, representative bodies and the social partners.*

The prerogative of human decision making

In line with our long advocated **'human-in-command'** approach to AI, ALLAI strongly recommends the AIA to arrange for certain decisions to remain the **prerogative of humans**, particularly in domains where these decisions have a strong moral component as well as legal implications or a societal impact such as in the judiciary, law enforcement, social services, healthcare, housing, financial services, labour relations and education.

If one takes a deeper look at the high-risk AI uses of ANNEX III, one will see that once the requirements for high-risk AI are met, AI can potentially largely replace human decision making in law enforcement, the judiciary, enjoyment of essential services, migration and asylum, recruitment, hiring, firing and worker-assessment, education, democratic processes. Most AI is technically incapable of making these decisions at a human level and it is questionable if it is even desirable. Ultimately, not all decisions can or should be simplified to 'ones and zero's'.

ALLAI proposes the AIA to arrange for certain decisions to remain the ultimate prerogative of humans, particularly in domains where these decisions have potentially severe legal implications (such as in the judiciary, law enforcement, social services, healthcare, housing, financial services, labour relations and education).

5.

ACCOUNTABILITY GOVERNANCE & ENFORCEMENT

ALLAI welcomes the governance structure set up by the AIA. We do see areas for improvement though, in particular where it comes to inclusivity and multi-disciplinarity, complaints and redress, and the exclusion (for now) of the applicability of the AIA to 'legacy high-risk AI' and components of large scale European IT systems in the realm of "freedom, security and justice" already put into service before the application of the AIA. Moreover, the accountability structure is quite complex.

Complex self-assessment structure

The AIA contains a complex accountability structure with a lot of text on third party conformity assessment by notified bodies, while only biometric recognition and categorisation outside law enforcement is subject to third party prior conformity assessment. Prior conformity of all other standalone high-risk AI on ANNEX III are self-assessed.

The complexity of the AIA and the requirements in combination with self-assessment runs the risk of this process simplifying being reduced to check lists where a simple 'yes' or 'no' could suffice to meet the requirements. This would make the AIA fail to achieve its initial objectives of promoting and driving innovation of AI that is trustworthy and in line with our fundamental rights and societal values.

ALLAI recommends to have all high-risk AI assessed by a third party.

Inclusivity and multi-disciplinarity

We recommend the AI Board to not only include representatives from the national authorities, but from wider society as well, including the social partners and NGO's. Also, the board should consist of or be advised by expert advisors from various fields such as AI, data science, the law, ethics, social sciences, psychology, economics, labour relations, healthcare and education.

Complaints and redress

A complaints and redress mechanism for organisations and citizens that have suffered harm from the use of any AI-system that falls within the scope of the AIA is lacking and should be arranged for in the AIA.

Legacy high-risk AI

Article 83 of the AIA excludes high-risk AI systems already placed on the market or put into service before application of the AIA (unless after that date the AI system undergoes significant changes). ALLAI stresses that these 'legacy high-risk AI systems' should also be covered by the AIA, in order to avoid deployers to fast track any high-risk AI to avoid compliance requirements.

* * *

Authors

Catelijne Muller is co-founder and president of ALLAI; former member of the European High Level Expert group on AI; Rapporteur on AI for the European Economic and Social Committee; expert advisor for the Council of Europe on AI & Human Rights, Democracy and the Rule of Law; member of the OECD Network of Experts on AI (ONE.AI); former Dutch lawyer, admitted to the Amsterdam Bar.

Virginia Dignum is co-founder and board member of ALLAI; professor of Artificial Intelligence at Umeå University; scientific director of the Wallenberg AI, Autonomous Systems and Software Program – Humanities and Society (WASP-HS); former member of the European High Level Expert group on AI; member of the World Economic Forum AI Board; expert advisor on AI for UNICEF; member of GPAI's working group on Responsible AI.

ALLAI.

Prinseneiland 23A
1013 LL Amsterdam
welkom@allai.nl
www.allai.nl