ALLAI.

13.02.2022

#1

AIA in-depth

#1 | Objective Scope Definition

Articles 1 - 4 & ANNEX I

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

Authors:
Catelijne Muller, LL.M
Prof. Virginia Dignum
Maria Avramidou
Mónica Fernández Peñalver

CONTENTS

EXECUTIVE SUMMARY

In this paper, which is the first in a series, we will dive deeper into the main elements of Chapter I of the AIA. By evaluating the objective, scope and technical definitions (of AI and data), we will assess whether Chapter I indeed reflects the overall objective of the AIA, which is to protect health, safety and fundamental rights and support innovation.

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 elements could be considered to strengthen the objective of the AIA, which is the protection of health, safety and fundamental rights against the ill effects of AI:**

- Strengthening the aim and purpose of the AIA by drawing inspiration from existing regulations for potentially harmful products or practices, such as REACH.
- Mandatory Fundamental Rights Impact Assessment for all High-Risk AI systems.
- EU taxonomy aimed at ensuring digitally sustainable economic activities (inspired by the EU taxonomy for environmentally sustainable economic activities).

2 **Multiple (proposed) exclusions of AI systems or domains from the scope of the AIA render the AIA's protection of health, safety and fundamental rights less effective.** Most of these exclusions should be reconsidered, while the R&D exclusion could be replaced with exemptions from certain requirements and specific notification and labelling obligations when AI systems are used in R&D.

AI does not operate in a lawless world. Defining the scope of the AIA should include a clear clarification of its interplay with existing (and upcoming) primary and secondary EU law, UN Human Rights Treaties, Council of Europe Conventions and national laws.

3 **Both definitions of AI (in the AIA and the Slovenian Presidency compromise proposal) focus on AI techniques, while it is better to focus on the characteristics, or properties of a system, that are relevant to be regulated.** The focus on technologies can create loopholes and legal uncertainty, and is not necessary. This paper proposes an alternative definition of AI that provides sufficient legal certainty as to what is covered by it and is future proof, whereas it provides the necessary room for interpretation.

1 OBJECTIVE

In her political guidelines, President von der Leyen announced that the Commission would put forward legislation for a coordinated European approach on the human and ethical implications of AI. The White Paper on AI set out policy options on how to achieve the twin objective of promoting the uptake of AI and of addressing the risks associated with certain uses of the technology. The AIA proposal is the result of this promise and aims to implement an 'ecosystem of trust' by proposing a legal framework for trustworthy AI.

1.1 Health, safety and fundamental rights

The proposal is based on EU values and fundamental rights and aims to give people the confidence to embrace AI-based solutions, while encouraging businesses to develop them. According to the Commission, the use of AI with its specific characteristics (e.g. opacity, complexity, dependency on data, autonomous behavior) can adversely affect a number of fundamental rights enshrined in the EU Charter of Fundamental Rights (the "Charter").

In a paper for the Council of Europe's Ad-hoc Committee on AI (CAHAI), Catelijne Muller described how AI can affect virtually all human rights as enshrined in the European Convention on Human Rights (the "ECHR"), as well as overall democracy and the rule of law.

The decision to approach AI from a health, safety and fundamental rights' perspective is thus a good one. The question whether the AIA indeed reaches this objective will be the guiding light throughout these series of papers.

It is not the first time that the European Commission proposes regulation for potentially harmful products or practices, in order to protect its citizens. While one could (and maybe even should) question whether the chosen instrument of a product safety regulation is the most suitable instrument to regulate such an impactful technology as AI, drawing inspiration from existing product safety regulations could be of value here.

Three elements that could be considered to strengthen the protection of health, safety and fundamental rights:

- Strengthening the aim and purpose of the AIA by drawing inspiration from existing regulations for potentially harmful products or practices, such as REACH.
- Mandatory Fundamental Rights Impact Assessment for all High-Risk AI systems.
- EU taxonomy aimed at ensuring digitally sustainable economic activities (inspired by the EU taxonomy for environmentally sustainable economic activities).[1]

[1] https://ec.europa.eu/info/business-economy-euro/banking-and-finance/sustainable-finance/eu-taxonomy-sustainable-activities_en

RECOMMENDATION

1.2 Drawing inspiration from REACH

For inspiration, we decided to look at the REACH regulation, which is aimed at *'improving the protection of human health and the environment from the risks that can be posed by chemicals, while enhancing the competitiveness of the EU chemicals industry'*. We fully acknowledge that AI, in its entirety and complexity, cannot in any way be compared with a chemical. We do however see parallels between the root aims of REACH and of the AIA, which is protecting people (and the environment) from the hazards of potentially dangerous products or practices.

As with the AIA, REACH aims to ensure that products are used in a safe and responsible manner. As with the AIA, REACH categorises chemicals according to their potential risks and connects requirements to the development and use of such chemicals depending on that risk level. As with the AIA, REACH prohibits or restricts the use of certain substances in the EU. Because of these parallels, and not to re-invent the wheel, using articles and approaches from REACH and 'translating' them into the AIA, could be a logical way to improve it.

Article 1 of the AIA (Subject matter) and particularly paragraphs (c) and (d) have no added regulatory value other than providing a 'table of contents' of some sort. The article could be improved with additional paragraphs inspired by art. 1 of the REACH regulation, to have it better reflect the aim of the AIA, which is to protect against the ill effects of AI while preserving innovation. Below is a table reflecting an improved art. 1, incorporating 3 new REACH-inspired paragraphs (1, 3 and 4). The new paragraph 3 stresses the value of the lifecycle approach to AI right off the bat. The new paragraph 4 provides a clear basis for the obligations of the entire chain of parties involved and (*in fine*) clearly mentions the fact that the AIA is underpinned by the precautionary principle, which already resonates throughout the AIA in its current form.

AIA	REACH	AIA amendment
Article 1	*Article 1*	*Article 1*
Subject Matter	**Aim and scope**	**Aim and subject matter**
This Regulation lays down: (a) harmonised rules for the placing on the market, the putting into service and the use of artificial intelligence systems ('AI systems') in the Union; (b) prohibitions of certain artificial intelligence practices; (c) specific requirements for high-risk AI systems and obligations for operators of such systems; (d) harmonised transparency rules for AI systems intended to interact with natural persons, emotion recognition systems and biometric categorisation systems, and AI systems used to generate or manipulate image, audio or video content; (d) rules on market monitoring and surveillance.	1. The purpose of this Regulation is to ensure a high level of protection of human health and the environment, including the promotion of alternative methods for assessment of hazards of substances, as well as the free circulation of substances on the internal market while enhancing competitiveness and innovation. 2. This Regulation lays down provisions on substances and mixtures within the meaning of Article 3. These provisions shall apply to the manufacture, placing on the market or use of such substances on their own, in mixtures or in articles and to the placing on the market of mixtures. 3. This Regulation is based on the principle that it is for manufacturers, importers and downstream users to ensure that they manufacture, place on the market or use such substances that do not adversely affect human health or the environment. Its provisions are underpinned by the precautionary principle.	**1. The purpose of this Regulation is to ensure a high level of protection of health, safety, fundamental rights and the environment, from harmful effects of artificial intelligence systems ("AI systems" in the Union, while enhancing innovation.** 2. This Regulation lays down: (a) harmonised rules for the placing on the market, the putting into service and the use of artificial intelligence systems ('AI systems') in the Union. (b) prohibitions of certain artificial intelligence practices; **3. These provisions shall apply to AI systems as a product, service or practice, or as part of a product, service or practice.** **4. This Regulation is based on the principle that it is for developers, importers, distributors and downstream users to ensure that they develop, place on the market or use artificial intelligence that does not adversely affect health, safety, fundamental rights, or the environment. Its provisions are underpinned by the precautionary principle.**

RECOMMENDATION

2 SCOPE

While the scope of the AIA appears to be wide and even global at first sight, a closer look learns that there are several important exclusions that could render the protection of the AIA far less effective and impactful. Moreover, the Slovenian Presidency partial compromise proposal adds several areas to be excluded from the scope of the AIA that need a thorough evaluation.

We will discuss the following (proposed) exclusions:

- Wings, wheels, water, rails
- Article 83 AIA
- National Security
- Research and Innovation
- General Purpose AI

We will also discuss the AIA's relation to existing Union Law, as this also defines its scope.

2.1 Wings, wheels, water, rails

Article 2 paragraph 2 AIA holds a noteworthy exclusion of AI systems used in the aviation, vehicle, marine and railroad domains. High risk AI systems in these domains are excluded from the scope of the AIA (with the exception of art. 84).

According to Recital (29), the Commission deems it more appropriate to amend the EU regulations and directives that underpin these domains rather than have these domains be covered by the AIA. This, supposedly, because the requirements and procedures in these domains are already 'stricter' that what the AIA would require. The Commission also wants to avoid interfering with existing governance, assessments and authorities already established in these domains. The Commission deems it more appropriate to take into account the mandatory requirements for high-risk AI systems laid down in this Regulation when adopting any relevant future delegated or implementing acts on the basis of the rules underpinning these domains.

This reasoning is confusing, because other domains that also have strict requirement and procedures, (such as medical devices, lifts, toys etc.), are not excluded from the scope of the AIA. To avoid interference with existing requirements and procedures in these domains, the AIA simply states that the requirements are to be included in those existing procedures. One would assume that the same structure could be followed for AI in relation to 'wings, wheels, water and rails'.

2.1.1 The implications of the exclusions

To understand the implications of these exclusions, it is good to consider what kind of AI systems would not have to comply with the AIA under this exclusion.

First of all, self-flying planes (e.g. drones), self driving vehicles, self-sailing boats and self-driving trains are excluded from the scope of the AIA. AI systems that are safety components of planes, vehicles, boats and trains (and are subject to a prior third-party conformity assessment) are excluded. But also flight crew assessment and certification and driver behaviour monitoring with AI would likely be excluded. To name just a couple of examples.

As regards the 'wings', the exclusion not only applies to planes, but also to AI that is used to ensure the overall security of civil aviation (on airports and planes and in airport operation services). AI-driven surveillance (beyond biometric identification), profiling, crowd monitoring, behaviour assessment etc. would not be covered by the AIA.

Moreover, the exclusion effectively to allows all prohibited AI practices of art. 5 AIA in these domains.

The idea of not creating a separate set of rules and procedures for AI in these domains, because they are already heavily controlled and regulated, makes sense. But excluding them entirely from the scope of the AIA might create unintended loopholes and runs the risk of leaving certain impactful AI systems and practices in these domains entirely unregulated, at least for now. Only integrating the high risk AI requirements when delegated or implementing acts in these domains are adopted in the future, creates unclarity and legal uncertainty, but also undesirable periods of legal lacunae.

RECOMMENDATION

Avoid loopholes and legal lacunae for AI in the Wings, Wheels, Water and Rails domains

Integrate the prohibitions and high-risk requirements into the Wings, Wheels, Water and Rails regulatory structures by following the same approach as for the products of ANNEX II Section A. To achieve this, article 2 paragraph 2 AIA can simply be deleted.

2.2 The Article 83 exclusions

New EU regulation usually provides for a 24-month transition period to give Member States time to implement the regulation into their national legal systems and to give organizations time to adjust their products, services or processes to the new rules. Article 83 of the AIA however, seems to do the opposite. It (initially) excludes two groups of existing AI-systems from its scope and has it kick in (much) later or, in some cases, not at all:

- AI systems that are part of the EU's centralized information systems for borders and security;
- High-risk AI systems already placed on the market or put into service.

2.2.1 **Significant change to design or intended purpose**

An instance where the AIA would kick in for both these groups (albeit under different additional conditions) is when there are any "significant changes in the design or intended purpose of the system" after 12 months as from the date the AIA becomes applicable. It is thus interesting to have a closer look at what is meant by '*significant changes* in *design* or *intended purpose*' here.

Design is not defined in the AIA, but from a technological point of view, the design phase usually precedes the technical phase and often consists of a 'pen and paper' exercise to illustrate an AI idea that is then used to guide the (technical) development phase. From this perspective, while the design might alter during the development phase to better serve the ultimate result, once the system is finished, the design will not likely change much anymore.

Intended purpose is defined in the AIA and means "*the use for which an AI system is intended by the provider, including the specific context and conditions of use, as specified in the information supplied by the provider in the instructions for use, promotional or sales materials and statements, as well as in the technical documentation*." So when a provider of a system that intended to score creditworthiness does not change the purpose of this system into, let's say, the prediction of crimes, the system in principle remains untouched by the AIA.

This could mean that many of the already existing high-risk AI systems in our society might never fall within the scope of the AIA.

2.2.2 **Border control AI**

The EU uses multiple IT systems to control its outer borders (EU's centralized information systems for borders and security) such as the 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). Any AI that is or becomes part of these systems before the AIA comes into force, is initially excluded from the scope of the AIA. This changes if the legal acts underpinning these systems are amended in such a way that it leads to a significant change in the design or intended purpose of the AI system concerned. We have already discussed the limitations of those conditions.

According to a recent paper by the EPRS[2], the "EU's centralized information systems for borders and security" are increasingly incorporating biometric technologies, algorithmic profiling, and AI driven behavior assessment. Many of them would be considered high-risk AI systems (as per art. 6(2) jo. ANNEX III, sub 7 of the AIA). The recently adopted European Travel Information and Authorisation System Regulation (ETIAS) for example legally allows algorithmic profiling for the so-called ETIAS watchlist of Europol, containing detailed information on persons of interest about their known or suspected criminal activity. Some safeguards, intended to avoid discriminatory bias and to provide means of redress are included, but they are entirely different than the requirements for high-risk AI in the AIA. Moreover, any AI practice that will eventually be prohibited under the AIA, shall remain allowed at EU border IT systems, if implemented in time and as long as there are no changes to their underpinning rules that lead to significant changes in the design or intended purpose.

For these systems, also the existing ones, the AIA does eventually kick in, but only when the entire IT system is evaluated, which is usually every two to three years. Considering that the Commission expects the AIA to come into force in 2025 at the earliest and becomes applicable 24 months after that, a back of a napkin calculation learns that a prohibited or high-risk AI system at the EU borders that is implemented now, could potentially remain in place until 2031. This could also be the earliest that high-risk AI systems at the EU borders would even have to comply with the requirements of the AIA.

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

2.2.3 **Existing High-Risk AI**

As said, article 83 also excludes existing high-risk AI systems from the scope of the AIA, unless they undergo significant changes in their design or intended purpose after the AIA has been applicable for 12 months.

We already discussed the limitations of the “design and intended purpose-change condition” above, but we would like to give another example here. Imagine an AI system designed for, and with the intended purpose of, surveilling students while they take online exams (a practice that has proliferated over the past two years due to the Corona crisis[3]). While any such system implemented 12 months after the AIA comes into effect would be covered by Chapter II and ANNEX III of the AIA (requirements for high risk AI), any such existing systems would be fully exempted from these requirements (also after the AIA has come into force) as long as their design and intended purpose does not change. These systems can be updated, tweaked, and optimised over time, but that would not necessarily result in a substantial change in the design or intended purpose. If implemented in time, these systems, while on the high-risk list, might not ever have to comply with the AIA.

[3] https://allai.nl/portfolio-item/online-proctoring/

RECOMMENDATION

A fixed transition period for border control AI and existing high-risk AI

Both exclusions in article 83 AIA can lead to fairly undesirable situations and should be dealt with accordingly, i.e. with a fixed transition period during which organizations that provide or use AI, can bring their AI-systems and processes in line with the AI.

2.3 National security exclusion

The Slovenian Presidency compromise proposal makes explicit reference to the exclusion of national security from the scope of the AIA. This approach appears to be similar to the proposed exclusion of national security in the Council’s common position on the ePrivacy Regulation.

First and foremost, there is no commonly accepted definition of national security. The Article 29 Working Group in 2014 already raised the question, in a similar discussion in relation to the GDPR, “to what extent an exemption focused on national security continues to reflect reality, as the work of intelligence services is more than ever before intertwined with the work of law enforcement authorities and pursues several different purposes.” It noted that “data is shared on a continuous and global basis, leaving aside the question which nation’s security is to benefit from the analysis of these data.” Given that AI systems heavily if not primarily depend on data and in fact perform data analysis, the same argument could be made for the current proposed exemption.

The Working Party at the time called upon the Council, the Commission and the Parliament to come to an agreement in order to define the principle of national security and be conclusive as to what should be regarded as the exclusive domain of the Member States. As far as we know, no such definition exists to date.

There is however a more fundamental question to be asked as regards the exclusion of national security from the scope of the AIA. The Council compromise proposal specifically refers to art. 4(2) of the TEU to motivate the exclusion. This article states that national security is the sole competence of the Member States, which means that the EU has no competence in regulation matters of national security.

The AIA however does not specifically regulate national security matter. It is a broad product regulation, aimed at protection of health, safety and fundamental rights. The exclusion of national security from the scope of the AIA could lead to the exclusive competence of Member States to regulate AI (or not) in the area of national security. The EU and Member States' competences are however already defined in the Treaties of the EU, and any changes or additions to those cannot be made by simply excluding certain areas from EU regulation. Such an approach would undermine the workings of the Treaties and even conflict with the principle of sincere cooperation of art. 4(3) of the TEU, stating that Member States should refrain from taking any action – including legislative – that could jeopardize the attainment of the Union's objectives.[4]

[4] Rojszczak (2021) "The uncertain future of data retention laws in the EU: Is a legislative reset possible?"

If one were to allow such an exclusion for AI legislation, one could argue excluding national security from any and all EU legislation that (even remotely) has a link with national security. Think of the (proposed) NIS Directive, ECI Directive and REACH[5], but also EU legislation related to motor vehicles. As much as we do not (have to) exclude national security from such legislation, we should not exclude it from the AIA.

[5] See ENISA: https://www.enisa.europa.eu/topics/threat-risk-management/risk-management/current-risk/laws-regulation/national-security

Exemptions in the AIA when national security is at stake

National security could be an area where exemptions to the AIA might be considered. These exemptions should however be underpinned by the principles of necessity and proportionality, limited in time and limited by the protection of fundamental rights. The CJEU[5] has established multiple times that "the powers of public authorities face an insurmountable barrier, namely the fundamental rights of individuals."[6] Given the fact that the mere aim of the AIA is to protect our fundamental rights, renders this notion even more pertinent.

[5] Opinion of Advocate General delivered on 15 January 2020, Joined Cases C–511/18 and C–512/18, EU:C:2020:6, para. 132

[6] Rojszczak (2021) "The uncertain future of data retention laws in the EU: Is a legislative reset possible?"

RECOMMENDATION

2.4 R&D exclusion

The Slovenian Presidency compromise proposal further makes explicit reference to the exclusion of AI used solely for research and development from the scope of the AIA. This proposal raises some serious concerns. Although the newly proposed Recital 12a states that "any research and development activity should be carried out in accordance with recognized ethical standards for scientific research", such ethical standards may fall short and may not cover the scope of what AI research can do, leaving loopholes to unethical research and development. We also worry that such exclusion could lead to a declining incentive to approach AI research in a multidisciplinary manner, where multiple domains are involved.

AI does not tend to ‘stay’ in the lab. Research in the field of AI is often open-access and open-source as codes are shared and uploaded to platforms such as GitHub, which is accessible and free to use. This gives anyone the opportunity to use new AI techniques and codes for any purpose or practice, also high-risk ones, the latter possibly without any adherence to the requirements for high-risk AI. Any ‘downstream’ user would have to make sure the high-risk AI system is in line with the AIA, but this is almost in all cases merely a self-assessment obligation.

While we fully respect the freedom of research, such freedom is not absolute. It never is. And again, the issue at hand is not new. Also, here we can draw inspiration from REACH, which holds specific arrangements for when hazardous material is used for Scientific Research and Development.[6]

[6] [Guidance in a Nutshell - SR&D and PPORD: https://echa.europa.eu/documents/10162/2324906/nutshell_srd_ppord_en.pdf/14675e6c-b2cf-4049-81ad-3d1bc41ace6d”

In the same spirit, we recommend that rather than fully excluding scientific research and development from the AIA, specific arrangements are included to facilitate it. Such arrangements could first and foremost include a definition of “Scientific Research and Development" along the lines of the definition in REACH.

REACH	**AIA amendment**
Article 3 *Definitions*	*Article 3* *Definitions*
(23) scientific research and development is any scientific experimentation, analysis or chemical research carried out under controlled conditions in quantities of less than 1 tonne per year.	(x) scientific research and development means: any scientific development, experimentation, analysis, testing or validation carried out under controlled conditions.

In this context, “controlled conditions” can be understood to mean that procedures and measures are in place to minimise or control potential risks of harm to health, safety and fundamental rights.

As a next step, obligations and/or exemptions could be added to Chapter 3 specifically aimed at scientific researchers and developers via a new Article 30.

A new article 30 could include the following (partial) exemptions and obligations for AI R&D:

Exemptions regarding:

- Quality management system
- Technical documentation of the high-risk AI systems conformity assessment procedure
- CE-marking obligation
- Specific (sub-)requirements for high risk AI, provided that the Scientific Research and Development institution provides suitable justification for not considering the requirement(s)

Obligations could include:

- An obligation to notify the national competent authorities of the Member States in which the scientific research and development is performed, and/or
- Prior authorization by the national competent authorities of the Member State(s) in which the scientific research and development is performed
- A time-limit for the exemptions (3 – 5 years, renewable upon request);
- An “AI R&D exemption notice/label" including relevant information on the AI system(s) researched that should be attached to the (interim) results/information

RECOMMENDATION

2.5 General Purpose AI exclusion

The Slovenian Presidency compromise proposal further proposes to exclude ‘general-purpose AI systems’ from the scope of the AIA by adding a new Title (IVA) and Article (52a) to the AIA as well as by excluding ‘general purpose AI systems’ from the definition of ‘intended purpose’.

2.5.1 **What are general purpose AI systems?**

According to a newly proposed Recital (70a), general-purpose AI systems are understood as AI systems that do not have an intended purpose but are able to perform generally applicable functions such as image/speech recognition, audio/video generation, pattern detection, question answering, translation, etc. These systems are trained on broad data that can be adapted to a wide range of downstream tasks and applications.

The proposal suggests no definition for general purpose AI systems, but looking at the above recital, they could cover multiple (if not all) AI models and techniques and perhaps also the widely used ‘benchmark’ datasets. We assume that they would in any event include the recently emerging ‘foundation models’. Self-supervised AI models that, according to the recently founded Centre for Research on Foundation Models at Stanford University, warrant caution and extensive research because, while they have demonstrated impressive behavior, they can fail unexpectedly, harbour biases, and are poorly understood. Examples of such ‘foundation models’ are BERT, GPT-3 (natural language processing), DALL-E, CLIP (computer vision). There is a lack of agreement on basic questions such as when ‘foundation models’ are “safe” to release.[7]

[7] Bommasani et al (2021) “On the Opportunities and Risks of Foundation Models

The fact that these models are being deployed at scale means that they can become singular points of failure that can radiate harms (e.g., security risks, inequities) to countless downstream AI applications. For natural language processing models such as BERT and GPT-3, it remains an open research question to understand whether it is possible to make foundation models that robustly and equitably represent language with both its major and subtle variations.[8] Of particular relevance to this last question is the fact that foundation models are trained on observed (often historical) data and will thus represent the same errors, gaps and biases as the data it observed.

[8] See also: Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchel (2021): On the dangers of stochastic parrots: Can language models be too big? In proceedings of FAcct 2021.”

Another category of 'general purpose AI' systems could be the so-called 'no-code AI development platforms', such as DataRobot, Google AutoML, Lobe, and Amazon SageMaker.[9] These platforms provide easy to use solutions to train an AI model simply by uploading data, or to find the 'best' AI model based on indicators set by the user. These platforms present the same 'singular point of failure' problems, when the model they develop or select holds errors or biases.

[9] See also: "How no-code Ai development platforms could introduce model bias" Kyle Wiggers, 2022 Venturebeat

As with any AI system, general purpose systems do not necessarily inform us about causality, neither do they have common sense. We should be aware that while a generic model (i.e., one that is not specific to a user or community) could be sufficient in some cases, there have been ample examples in recent years where the mere generalization induced by AI systems resulted in unfair outcomes for groups or individuals and at times in devastating personal harm. General purpose AI systems, whatever these are exactly, should not be viewed as simple AI systems that perform easy, manageable tasks. Not in the least because of the highly adaptable and multipurpose use of these systems and their impact can be paramount.

2.5.2 **Benchmark datasets**

Apart from ‘general’ AI-models, there is a wide practice of using so-called ‘benchmark’ datasets that form the backbone of machine learning research and development. Recent critical inquiry into these datasets have however revealed biases, poor categorization and offensive labelling[10] in these datasets. Koch et al. have found increasing concentration on fewer and fewer datasets in the field of AI research.[11] Despite widespread recognition that datasets are critical to the advancement of the field, careful dataset development is often undervalued and disincentivized, especially relative to algorithmic contributions.[12] Even many of the fairness in ML researchers use datasets ‘as is’ without checking them for completeness, representativeness and overall fairness (Propublica’s COMPAS dataset is widely used in this field while there is literature that suggests that a data processing error was made that resulted in a recidivism rate inflation of over 24%).[13]

[10] Koch et al. (2021)
[11] Ibid.
[12] Morgan Klaus Scheuerman, Emily Denton, and Alex Hanna (2021): Do datasets have politics? Disciplinary values in computer vision dataset development; Nithya Sambasivan et al. (2021): Everyone wants to do the model work, not the data work: Data cascades in high-stakes AI.
[13] Mathias Barenstein (20-19) ProPublica’s COMPAS Data

2.5.3 **Homogeneity**

The issues described above around general purpose AI (consisting of ever fewer and more general models and benchmark datasets) can be referred to as the ‘homogeneity problem'. Machine learning by its nature results in more homogeneous decision making compared to human decisions. Human decisions have a lot of “noise”, and while removing the noise is one of the main attractions of statistical decision making as done with AI, there are also risks. If statistical decision-making results in similar decisions being made by many decision makers, otherwise individual biases could become amplified and embedded to the point where they create structural drawbacks.[14]

[14] Creel and Hellman (2021): The Algorithmic Leviathan: Arbitrariness, Fairness, and Opportunity in Algorithmic Decision Making Systems, *Virginia Public Law and Legal Theory Research Paper*, no. 2021–13.

Homogeneity can occur at dataset level, when many machine learning systems use the same training data. They could result in the same classifications, even if the algorithms are adapted to the new situation. Or at model level, when multiple actors use the same (general purpose) AI system, resulting in the same (groups of) people repeatedly affected by multiple actors.[15]

[15] O’Neill describes examples of job seekers being repeatedly rejected on the basis of personality tests, all offered by the same vendor: “How Algorithms Rule Our Working Lives,” *The Guardian* 16 (2016).

Given the above, general-purpose AI systems, if we even succeed in defining these properly, should not be excluded from the scope of the AIA.

2.5.4 **Intended purpose or reasonably foreseeable use**

The Slovenian Presidency proposal was apparently motivated by the notion that it would be impossible for providers of 'general purpose AI systems' to comply with the requirements for high-risk AI. Art. 8 paragraph 2 obligates that providers take the 'intended purpose of the AI system into account when adhering to the requirements, for 'general purpose AI' this would simply not be possible, given the many purposes these systems could have.

While this might be true, this notion alone does not warrant a full exclusion of these systems from the scope of the AIA. Introducing this exclusion would break with the overall Union objective of the safety and liability legal frameworks, which is to ensure that all products and services, including those integrating emerging digital technologies, operate safely, reliably and consistently and that damage having occurred is remedied efficiently.[17]

[16] The AIA currently only holds the notion of 'reasonably foreseeable misuse'.

[17] COM (2020) 64 final Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics

An alternative approach, which is common in Union legislation regarding product safety, would be to add the notion of 'reasonably foreseeable use' to the notion of 'intended purpose'. Given the potential impact of these 'general purpose AI systems' it is not unreasonable to ask from their providers to try to foresee (imagine) the potential uses of the AI system and address the risks to health, safety and fundamental rights these uses could bring, while appreciating that not all uses can be foreseen.

2.5.5 Excluding general purpose AI could stifle innovation

We also warn that fully excluding general purpose AI systems from the AIA runs the risk of in fact stifling innovation rather than supporting it. The exclusion would mean that the burden of bringing these systems in compliance with the AIA falls entirely on 'downstream' users of the general purpose AI systems. They would be the ones that have to bring the systems in line with the requirements for high risk AI, which might be too much of a burden, especially for SME's and micro enterprises, or perhaps even prove to be technically impossible. Even if the general purpose AI developer would help 'downstream users' with the technicalities of complying with the AIA, it places the latter in a fully dependent position, without having the appropriate means to seek redress when the general purpose AI system causes damage.

As a result, it could lead to a limited uptake of general purpose AI systems on the one hand, and a (further) concentration of AI innovation with general purpose AI developers on the other. The latter could thus gain competitive advantage over downstream users, as they would in principle not have to comply with the AIA and only see it kick in when their general purpose AI system is adapted to have an intended purpose.

Do not exclude 'general purpose AI' from the scope of the AIA, but include the notion of 'reasonably foreseeable use' for deployers of general purpose AI

Article 3
Definitions

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

(12a) 'reasonably foreseeable use' means the use of an AI system in a way that is or should be reasonably foreseeable

'Global' amendment

Amend 'intended purpose' to '**intended purpose or reasonably foreseeable use**' throughout the AIA

RECOMMENDATION

2.6 The AIA's relation to existing or new EU or national law

AI systems do not operate in a lawless world. A large 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, 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 e.g. in the field of social services and benefits, administration, labour, social security, pensions, insurance, law enforcement, to name but a few.

While it is uncertain whether the AIA aims at laying down minimum or maximum harmonization rules for AI, the mere fact that AI impacts virtually all legal domains imaginable, calls for more clarity on the position of the AIA within the full EU and national *acquis*. Moreover, the AIA aims to broadly protect health, safety and fundamental rights. All this supports the case for considering the AIA a minimum harmonization legislation. Some scholars even argue that maximum harmonization should be an instrument only used in exceptional cases, justified by sector-specific conditions[18].

[18] Weatherill (2020): The Fundamental Question of Minimum or Maximum Harmonisation

Recital (41) of the AIA emphasises that a high-risk classification does not necessarily mean that using the AI system is lawful under Union or national law, which could be read as aiming for the AIA to achieve minimum harmonisation. A reference in the recitals only however, seems insufficient to clarify the position of the AIA within the EU legal *acquis*. Moreover, the limitation to high-risk AI seems too narrow, since medium-risk AI might in certain circumstances also be unlawful under Union or national law.

Add two new paragraphs to Article 2 to clarify the interplay with existing and new EU and national laws

Article 2
Scope

6. An AI-system or practice that is in line with this Regulation, should also continue to comply with the European Charter on Fundamental Rights, existing and new secondary Union law and national law.

7. Member States may adopt or maintain in force more stringent provisions, compatible with the Treaty in the field covered by this Directive, to ensure a higher level of protection of health, safety and fundamental rights.

RECOMMENDATION

3 TECHNICAL DEFINITIONS

Some technical definitions in the AIAI and in particular the one for AI, could be improved or altered for the purpose of clarity and technical correctness.

3.1 Definitions of AI

The definition of AI in art. 3(1) of the AIA has already caused significant discussion among scientists on whether it is over inclusive or rather under inclusive and whether the list of AI-techniques of ANNEX I is complete or incomplete. On top of that, the Slovenian Presidency compromise proposal suggests a completely different definition of AI, with the aim of preventing the inclusion of more traditional software systems that are not considered to be AI.

Both definitions raise a number of questions but as we have argued before, listing AI-techniques, or rather “approaches” could easily create confusion, legal uncertainty and loopholes. First of all, the list of AI-techniques and approaches (ANNEX I) 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. 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.

As we have argued before, 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 organizations evading the regulation, simply by classifying their applications differently.

3.1.1 **Human-defined objectives**

Both definitions refer to systems that can achieve a given set of human-defined objectives. This overlooks the fact that not all AI systems need objectives or even human-defined objectives. Unsupervised learning for example, aims to identify patterns in raw (unlabelled) data without any predefined objective. Unsupervised learning methods are computationally more complex, less accurate, and less trustworthy than supervised learning methods. Thus, it is important that these are not excluded from the scope of the AIA.

3.1.2 **Alternative definition for AI**

Borrowing from the definition of AI developed by the High Level Expert Group on AI, one could consider the an alternative definition. The table below also indicates some suggested changes in other technical definitions e.g. for several types of data.

Amended technical definitions AIA

Article 3
Definitions

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

(1) ‘artificial intelligence (AI)' means computer systems that act in the physical or digital world and that, in an automated manner:
(i) decide on action(s) to take according to predefined parameters by perceiving their environment and analysing the collected structured or unstructured information from that environment; and/or
(ii) can adapt their decisions by analysing how the environment is affected by their previous actions.

(29) ‘training data’ means data used for training an AI system to fit its learnable parameters;

(30) ‘validation data’ means data used for providing an evaluation of the trained AI system. The process evaluates whether the model is under-fitted or overfitted; The validation dataset should be a separate dataset of the training set for the evaluation to be unbiased. If there is only one available dataset, this is divided into two parts, a training set and a validation set. Both sets should still comply with art. 10 (3) to ensure appropriate data governance and management practices.

(31) ‘testing data’ means data used for providing an independent evaluation of the trained and validated AI system to confirm the expected performance of that system before its placing on the market or putting into service. Similar to art 3 (30), the testing dataset should be a separate dataset from the training set and validation set. This set should also comply with art. 10 (3) to ensure appropriate data governance and management practices.

RECOMMENDATION

3.1.3 **ANNEX I**

From a legal perspective, ANNEX I would no longer be necessary if this definition is used. This definition refers to the characteristics and properties of AI rather than its technological make up. It provides sufficient legal certainty as to what is covered by it. It is future proof, whereas it provides the necessary room for interpretation, thus also eliminating the need for Article 4 (delegated acts to update ANNEX I).

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 Associate Prof. Aimee van Wynsberghe. Collectively, the founders have a broad expertise in AI: AI sciences, social impact, national and international policy, legal implications, and ethical impact.

CONTACT

ALLAI
Prinseneiland 23A
1013 LL Amsterdam
The Netherlands

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