

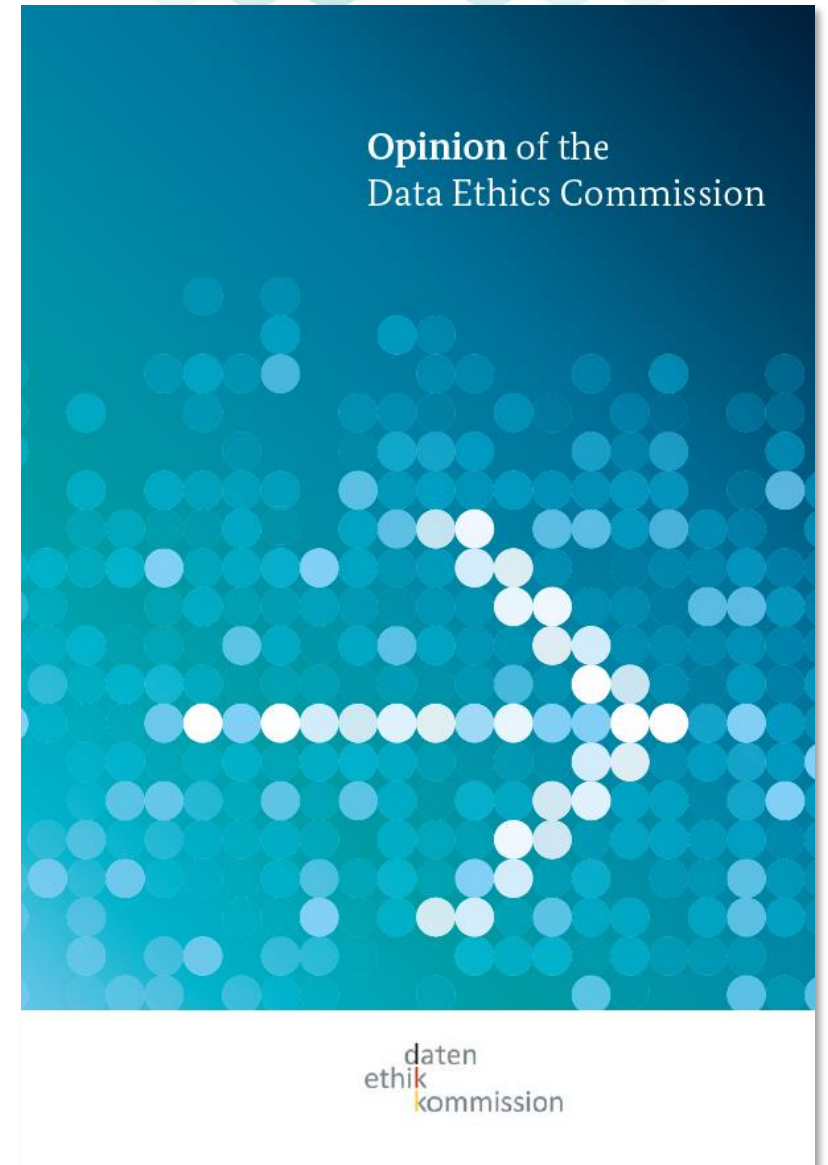
The Commission's Proposal for an Artificial Intelligence Act

The Perspective of the German Data Ethics Commission

Christiane Wendehorst,
Co-Chair of the Data Ethics Commission (2018-2019)

Data Ethics Commission

- Established in mid 2018 with the mission to develop, within one year, an ethical and regulatory framework for data, ADM and AI
- Co-chaired by Christiane Wendehorst and Christiane Woopen
- Opinion presented in Berlin on 23 October 2019
- Includes ethical guidelines and 75 concrete recommendations for action regarding data and algorithmic systems







What is Data Ethics?

Data

Data-driven technologies (such as AI)

Wider framework

Ethics of handling
personal data

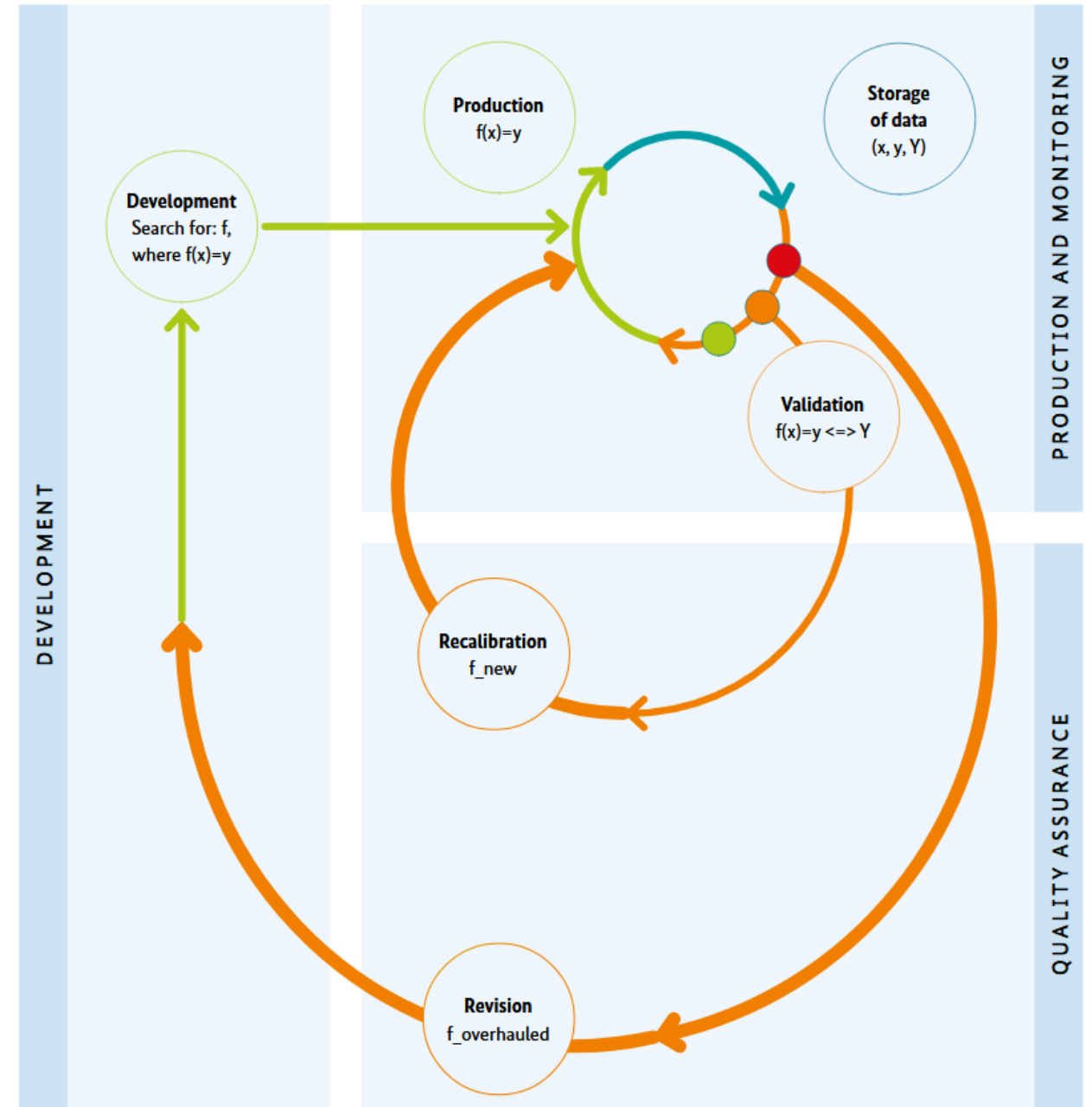
Ethics of handling data in general
(including non-personal data)

Ethics of handling data and data-driven technologies
(including algorithmic systems, such as AI)

Ethics of the digital transformation in general
(including issues such as the platform economy or the future of work)

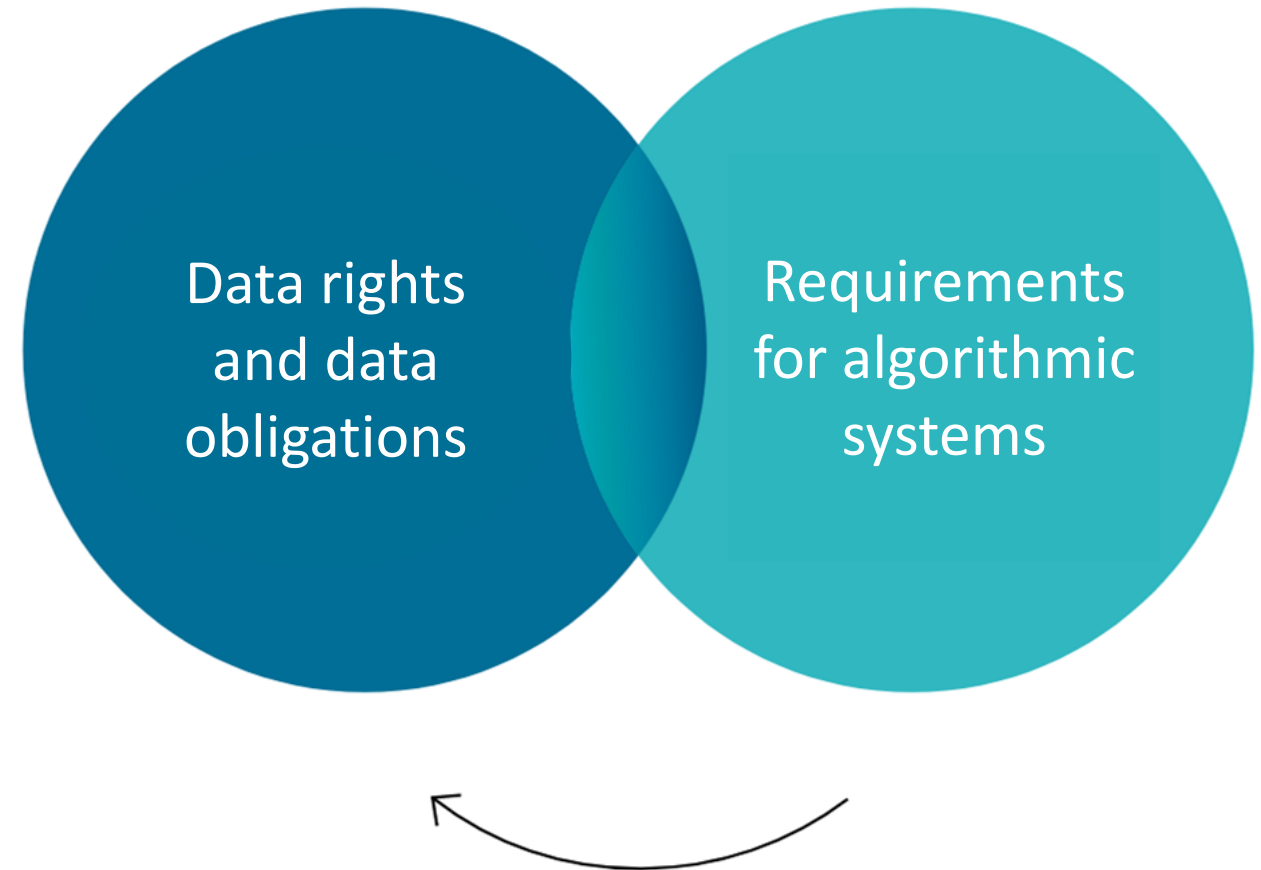
AI vs Algorithmic System

- **‘De-mystify’ the technology** and do away with popular misconceptions that may be inspired by science fiction rather than by science
- Not useful to quarrel about the proper definition of ‘Artificial Intelligence’,
- Ethical and legal implications may follow more from the existence of an **‘algorithmic system’** rather than on how the algorithms are created



Data perspective and algorithm perspective

- Two mutually dependent and overlapping discourses
- In part also reflected in different legal instruments
- Also recognisable from the German Federal Government's guiding questions





Introduction



Ethical and legal principles



Technical foundations



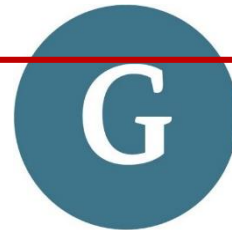
Multi-level governance of complex data ecosystems



Data




Algorithmic systems



A European path





**The Position of the German
Data Ethics Commission on
Algorithmic Systems**

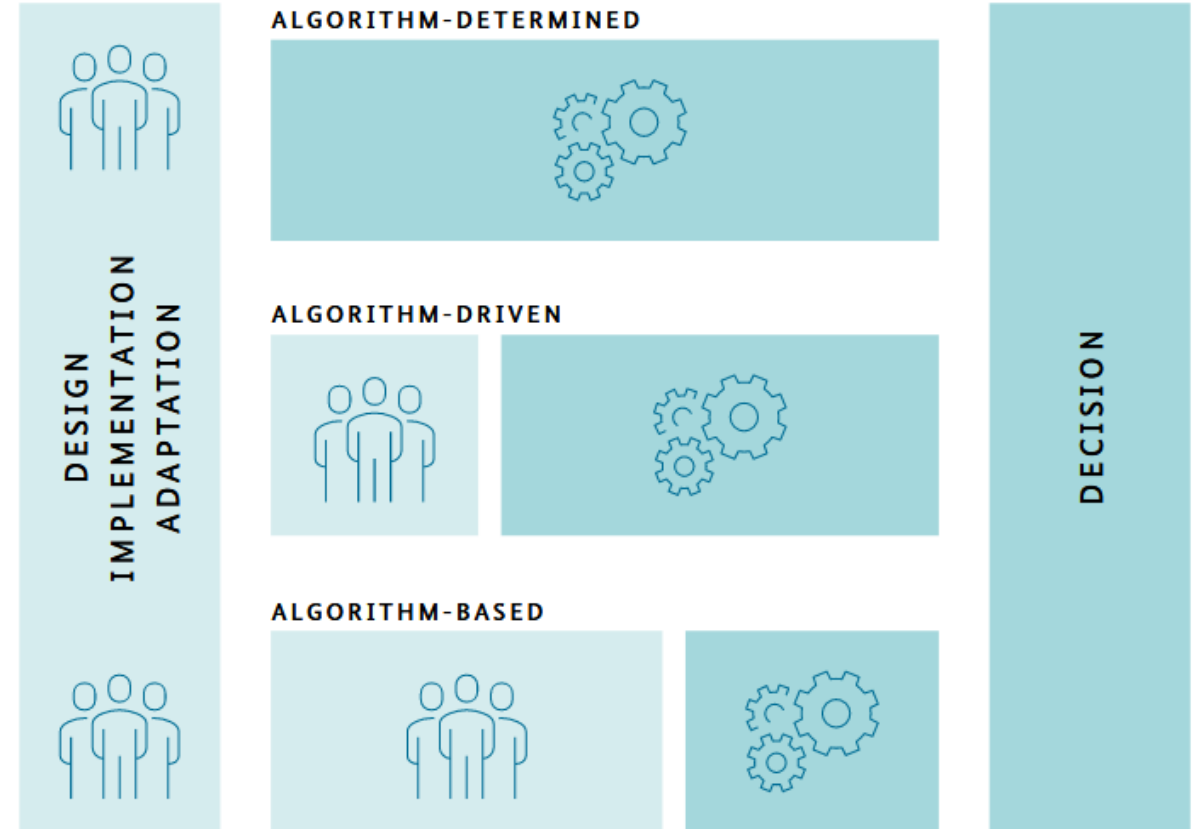


Algorithmic Systems



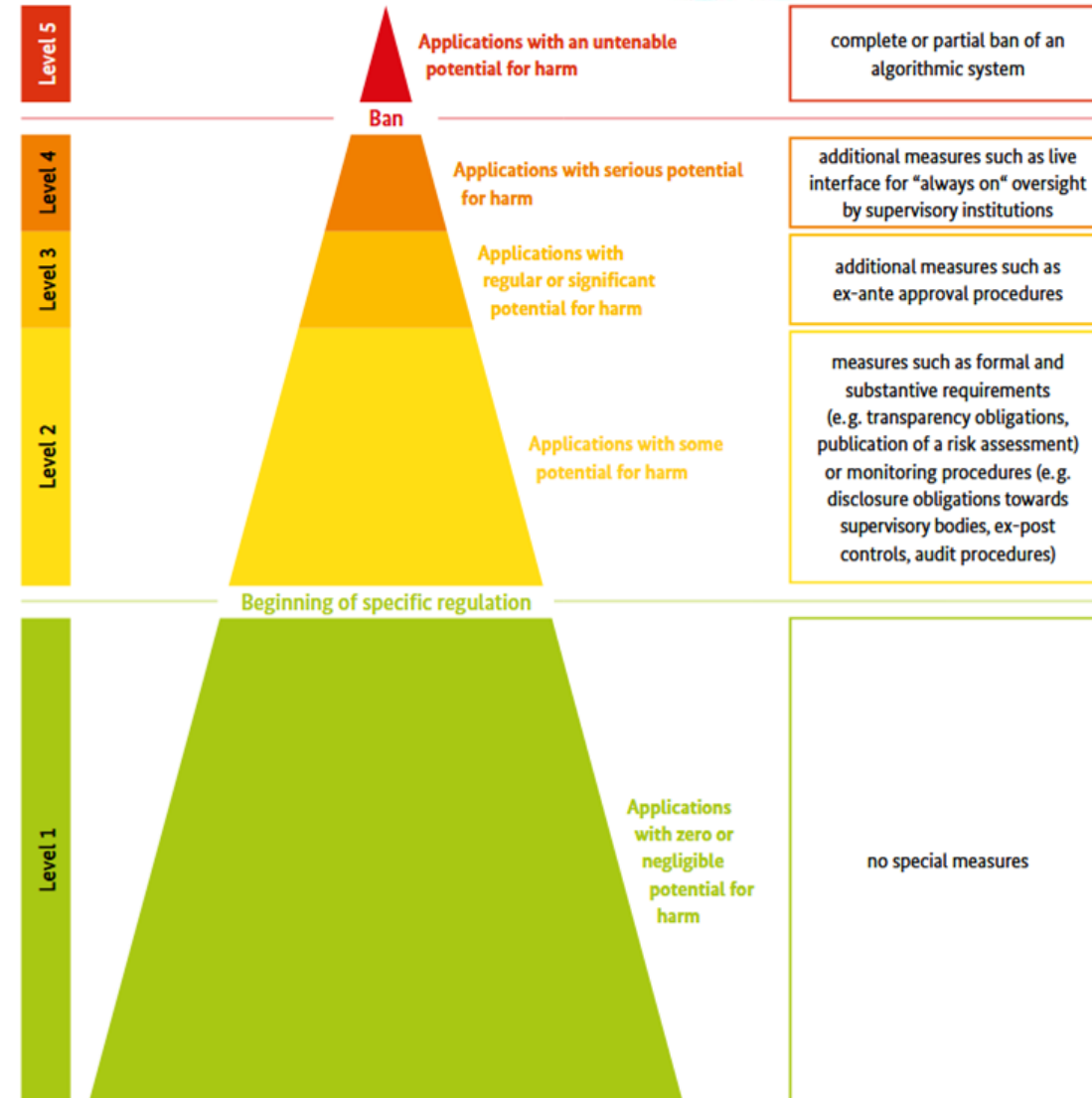
Algorithmic Systems

- AI only as a subset of algorithmic systems
- Differentiation of algorithm-based, -driven and -determined decisions
- General requirements for algorithmic systems



A risk-based regulatory framework

- „Criticality pyramid“: different levels of potential for harm (risk)
- No need for any regulation with regard to most algorithmic systems
- Ban on systems involving an unacceptable potential for harm



A risk-based regulatory framework

- Horizontal Regulation at EU level and sector specific legislation at both EU and national levels

Europäische Union



Verordnung für Algorithmische Systeme (EUVAS)

Zentrale Grundprinzipien für algorithmische Systeme, allgemeine materielle Regelungen zur Zulässigkeit und Gestaltung algorithmischer Systeme. Regeln zu Transparenz, organisatorischen und technischen Absicherungen und Institutionen und Strukturen der Aufsicht.

Bundesregierung und Europäische Union



Sektor 1

Ergänzende/
konkretisierende
Regeln und
Vorgaben

Sektor 2

Ergänzende/
konkretisierende
Regeln und
Vorgaben

Sektor 3

Ergänzende/
konkretisierende
Regeln und
Vorgaben

Sektor 4

Ergänzende/
konkretisierende
Regeln und
Vorgaben

Instruments

Depending on the level of criticality:

- **Labelling** requirements, **information** duties, and duties to **explain**
- **Risk assessment**, documentation and logging
- Ensuring **quality** from a technical and mathematical-procedural perspective
- Ex-post control – **licensing** procedures – continuous audits up to ‘always on’ **oversight** via a live interface
- Individual protection even below the level of Article 22 GDPR
- Rethinking **anti-discrimination law**

Institutions

- **Sectoral supervisory** authorities should normally be in charge (but be better equipped, and have advisory councils representing civil society and a diverse range of players)
- Support to be provided by **national centre of competence at federal level**
- Technical standards, **co-regulation and self-regulation**
- **Algorithmic Accountability Codex**
- **Quality seals**
- **Contact persons** in companies and government authorities
- Rights to file an action on the part of competitors and consumer organisations



Use of algorithmic systems by state bodies

- Particular sensitivity and **enhanced criticality**
- Different situation for **law-making and dispensation of justice** on the one hand and **administration** on the other
- **Transparency and explainability** requirements
- Ethical and legal **limits to automated ‘total’ enforcement**

Liability for algorithmic systems

- Existing liability regimes need a **‘digital fitness check’** and may have to be reconsidered
- **No recognition of ‘electronic personhood’**
- **Operators’ liability along the lines of vicarious liability** of principals for their auxiliaries



**Comparing the AIA Proposal with
the Position of the German Data
Ethics Commission**



Brussels, 21.4.2021
COM(2021) 206 final

2021/0106 (COD)

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**

Article 3
Definitions

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

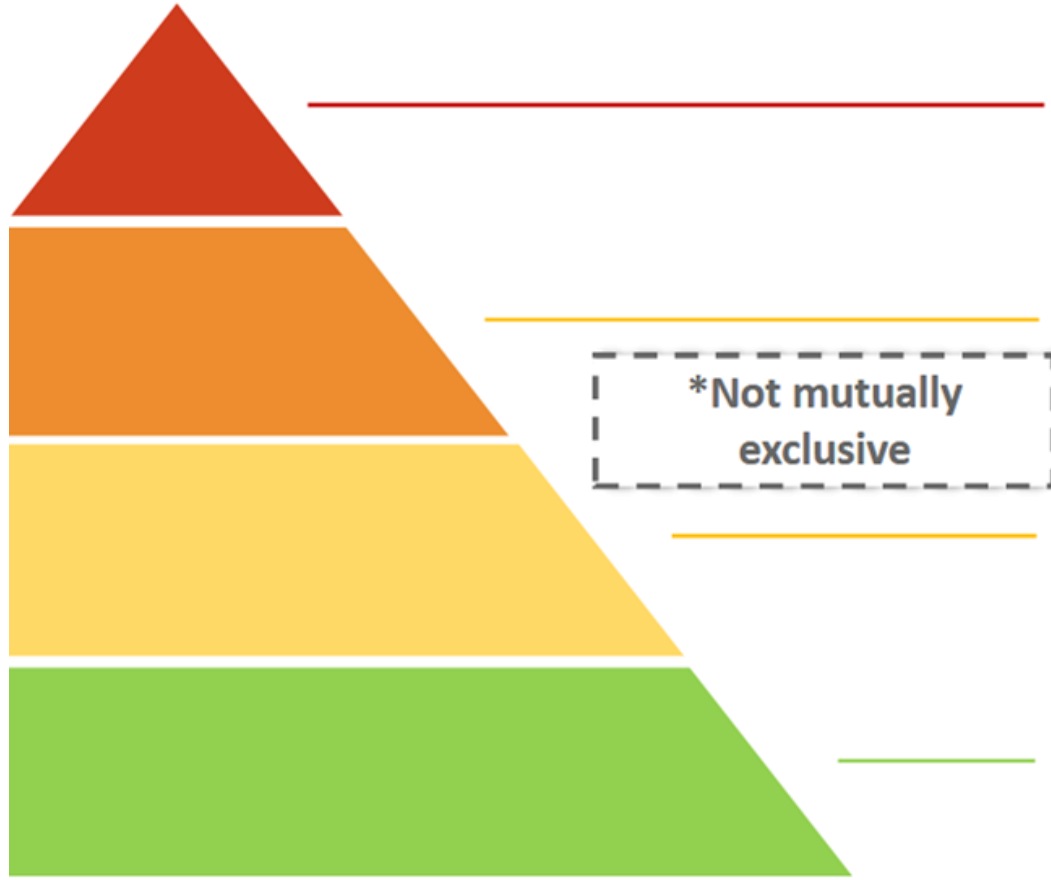
- (1) ‘artificial intelligence system’ (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with;

Extremely broad and flexible definition of AI

ANNEX I
ARTIFICIAL INTELLIGENCE TECHNIQUES AND APPROACHES
referred to in Article 3, point 1

- (a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;
- (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;
- (c) Statistical approaches, Bayesian estimation, search and optimization methods.

Risk-based approach



Unacceptable risk
e.g. social scoring

Prohibited

High risk
e.g. recruitment, medical devices

Permitted subject to compliance with AI requirements and ex-ante conformity assessment

*Not mutually exclusive

'Transparency' risk
'Impersonation' (bots)

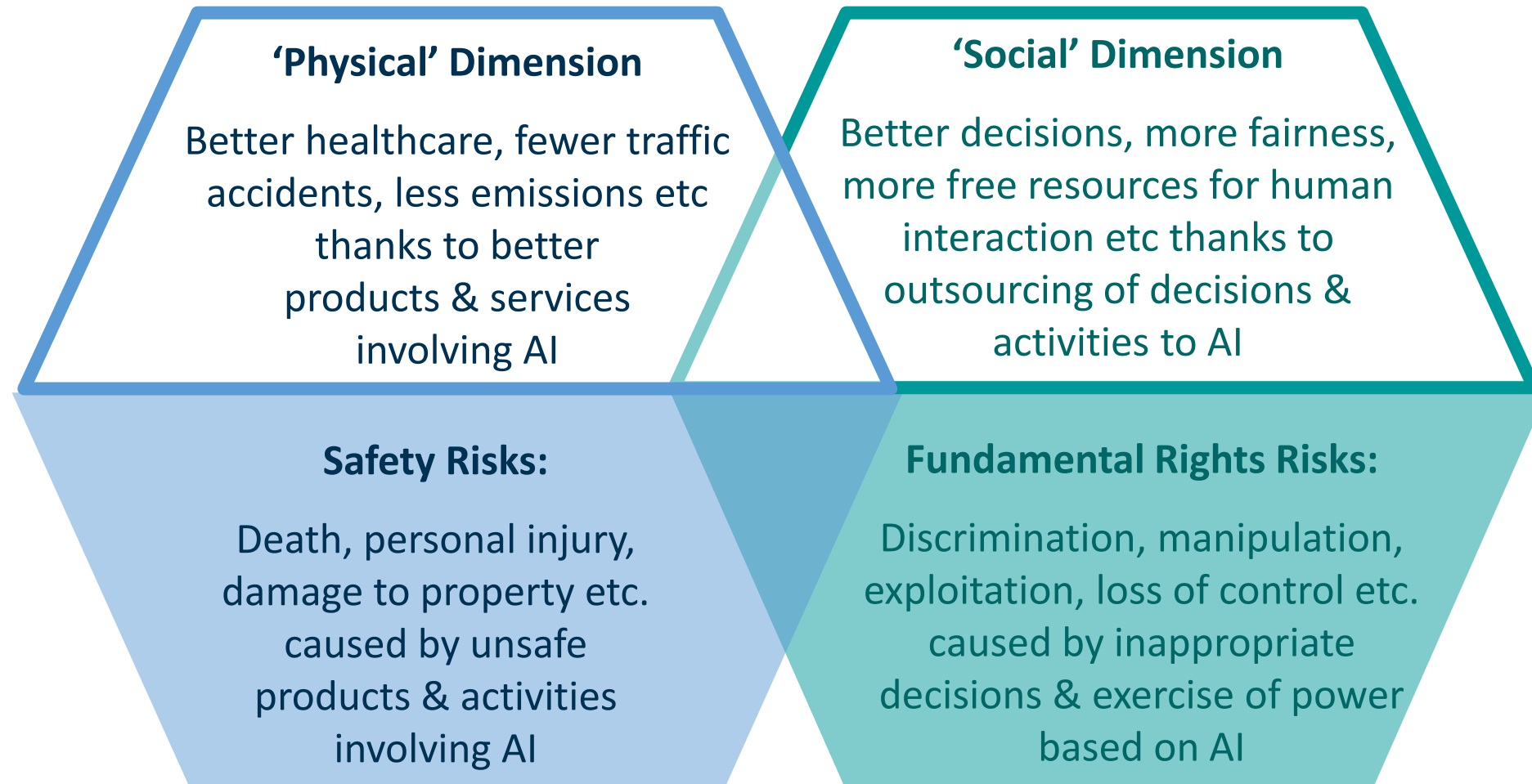
Permitted but subject to information/transparency obligations

Minimal or no risk

Permitted with no restrictions



Safety and fundamental rights risks





Safety Risks

Article 6

Classification rules for high-risk AI systems

1. Irrespective of whether an AI system is placed on the market or put into service independently from the products referred to in points (a) and (b), that AI system shall be considered high-risk where both of the following conditions are fulfilled:
 - (a) the AI system is intended to be used as a safety component of a product, or is itself a product, covered by the Union harmonisation legislation listed in Annex II;
 - (b) the product whose safety component is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment with a view to the placing on the market or putting into service of that product pursuant to the Union harmonisation legislation listed in Annex II.
2. In addition to the high-risk AI systems referred to in paragraph 1, AI systems referred to in Annex III shall also be considered high-risk.

Article 7

Amendments to Annex III

1. The Commission is empowered to adopt delegated acts in accordance with Article 73 to update the list in Annex III by adding high-risk AI systems where both of the following conditions are fulfilled:
 - (a) the AI systems are intended to be used in any of the areas listed in points 1 to 8 of Annex III;
 - (b) the AI systems pose a risk of harm to the health and safety, or a risk of adverse impact on fundamental rights, that is, in respect of its severity and probability of occurrence, equivalent to or greater than the risk of harm or of adverse impact posed by the high-risk AI systems already referred to in Annex III.
2. When assessing for the purposes of paragraph 1 whether an AI system poses a risk of harm to the health and safety or a risk of adverse impact on fundamental rights that is equivalent to or greater than the risk of harm posed by the high-risk AI systems



Fundamental Rights Risks



Safety Risks

ANNEX II

LIST OF UNION HARMONISATION LEGISLATION

Section A – List of Union harmonisation legislation based on the New Legislative Framework

1. Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC (OJ L 157, 9.6.2006, p. 24) [as repealed by the Machinery Regulation];
2. Directive 2009/48/EC of the European Parliament and of the Council of 18 June 2009 on the safety of toys (OJ L 170, 30.6.2009, p. 1);
3. Directive 2013/53/EU of the European Parliament and of the Council of 20 November 2013 on recreational craft and personal watercraft and repealing Directive 94/25/EC (OJ L 354, 28.12.2013, p. 90);
4. Directive 2014/33/EU of the European Parliament and of the Council of 26 February 2014 on the harmonisation of the laws of the Member States relating to lifts and safety components for lifts (OJ L 96, 29.3.2014, p. 251);
5. Directive 2014/34/EU of the European Parliament and of the Council of 26 February 2014 on the harmonisation of the laws of the Member States relating to equipment and protective systems intended for use in potentially explosive atmospheres (OJ L 96, 29.3.2014, p. 309);
6. Directive 2014/53/EU of the European Parliament and of the Council of 16 April 2014 on the harmonisation of the laws of the Member States relating to the making available on the market of radio equipment and repealing Directive 1999/5/EC (OJ L

ANNEX III
HIGH-RISK AI SYSTEMS REFERRED TO IN ARTICLE 6(2)

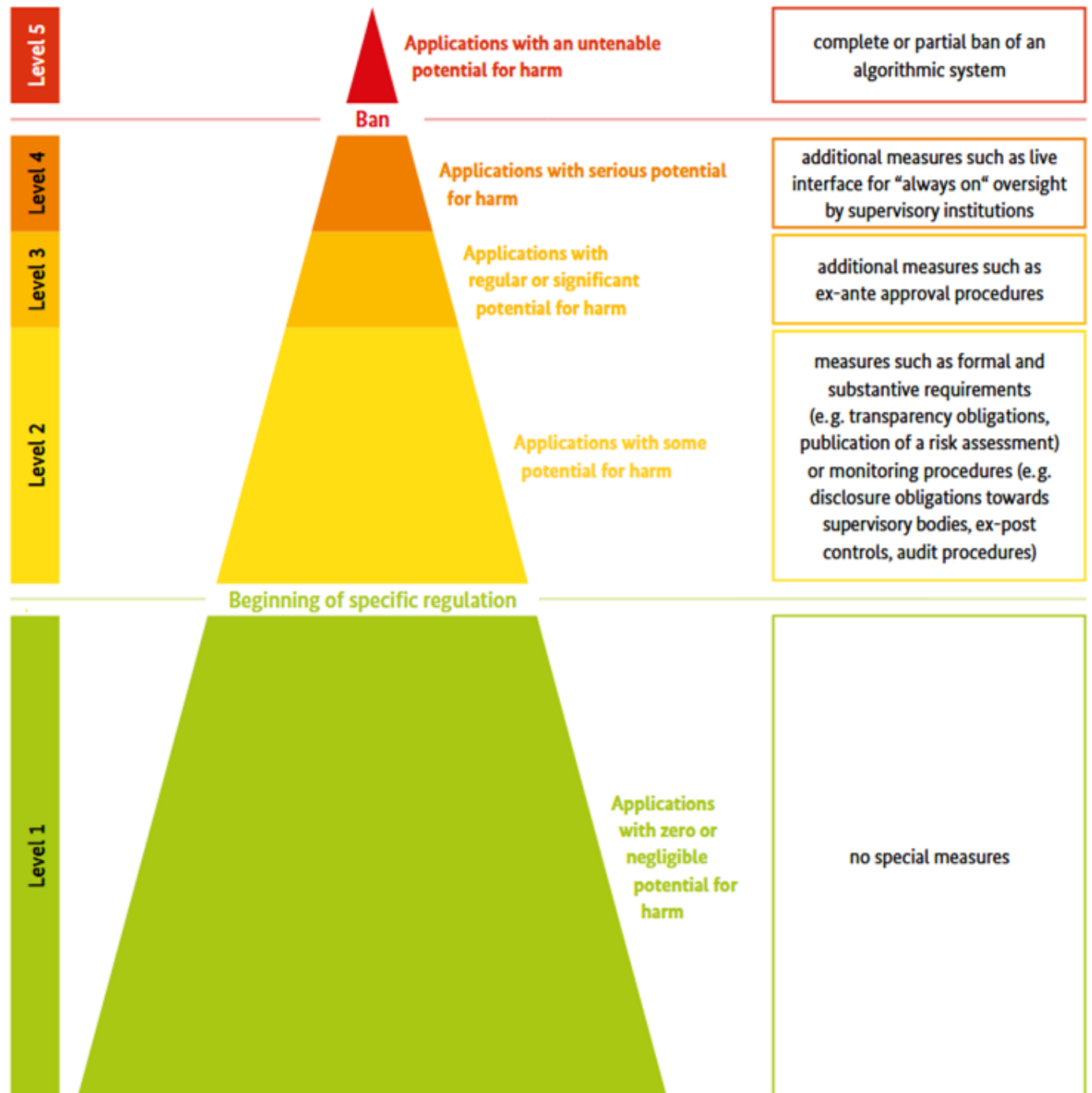
High-risk AI systems pursuant to Article 6(2) are the AI systems listed in any of the following areas:

1. Biometric identification and categorisation of natural persons:
 - (a) AI systems intended to be used for the ‘real-time’ and ‘post’ remote biometric identification of natural persons;
2. Management and operation of critical infrastructure:
 - (a) AI systems intended to be used as safety components in the management and operation of road traffic and the supply of water, gas, heating and electricity.
3. Education and vocational training:
 - (a) AI systems intended to be used for the purpose of determining access or assigning natural persons to educational and vocational training institutions;
 - (b) AI systems intended to be used for the purpose of assessing students in educational and vocational training institutions and for assessing participants in tests commonly required for admission to educational institutions.
4. Employment, workers management and access to self-employment:
 - (a) AI systems intended to be used for recruitment or selection of natural persons, notably for advertising vacancies, screening or filtering applications, evaluating candidates in the course of interviews or tests;



**Fundamental
Rights Risks**

Colours vs content



Colours vs content

- How is 'risk' defined? Does the AIA consider, to a sufficient extent, economic risks (e.g. exploitation and manipulation of consumers) and risks for the society at large, democracy, the environment, etc.?
- Who makes the risk assessment? Each provider or user, or the legislator? The legislator, and if so, to what extent is a 'sectoral' approach justified?
- Who is in charge of conformity assessment? Where is third party conformity assessment justified?
- What are the individual rights of affected person? Explainability?
-?

Prohibited AI Practices



Article 5

1. The following artificial intelligence practices shall be prohibited:
 - (a) the placing on the market, putting into service or use of an AI system that deploys subliminal techniques beyond a person's consciousness in order to materially distort a person's behaviour in a manner that causes or is likely to cause that person or another person physical or psychological harm;
 - (b) the placing on the market, putting into service or use of an AI system that exploits any of the vulnerabilities of a specific group of persons due to their age, physical or mental disability, in order to materially distort the behaviour of a person pertaining to that group in a manner that causes or is likely to cause that person or another person physical or psychological harm;
 - (c) the placing on the market, putting into service or use of AI systems by public authorities or on their behalf for the evaluation or classification of the trustworthiness of natural persons over a certain period of time based on their social behaviour or known or predicted personal or personality characteristics, with the social score leading to either or both of the following:
 - (i) detrimental or unfavourable treatment of certain natural persons or whole groups thereof in social contexts which are unrelated to the contexts in which the data was originally generated or collected;
 - (ii) detrimental or unfavourable treatment of certain natural persons or whole groups thereof that is unjustified or disproportionate to their social behaviour or its gravity;

Why restriction to 'physical or psychological harm'?
What about economic decisions, voting behaviour, ...?

Is the restriction to 'public authorities' adequate? What about gatekeeper services?

Should maybe 'discrimination' also have been mentioned? And practices prohibited under other law?

Why only some group-specific vulnerabilities? Is not exploitation of very individual vulnerabilities at least as dangerous? And why the restriction to physical or psychological harm?

Prohibited AI Practices



- (d) the use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement, unless and in as far as such use is strictly necessary for one of the following objectives:
 - (i) the targeted search for specific potential victims of crime, including missing children;
 - (ii) the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or of a terrorist attack;
 - (iii) the detection, localisation, identification or prosecution of a perpetrator or suspect of a criminal offence referred to in Article 2(2) of Council Framework Decision 2002/584/JHA⁶² and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least three years, as determined by the law of that Member State.
- 2. The use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement for any of the objectives referred to in paragraph 1 point d) shall take into account the following elements:
 - (a) the nature of the situation giving rise to the possible use, in particular the seriousness, probability and scale of the harm caused in the absence of the use of the system;
 - (b) the consequences of the use of the system for the rights and freedoms of all persons concerned in particular the seriousness, probability and scale of those

Why the restriction to ‘real time’ practices?

And is law enforcement the only problematic purpose?

Use of real time remote biometric identification is not really ‘prohibited’ but rather heavily regulated and seems somewhat an alien element in Article 5

High-risk AI systems



Risk management system

Data and data governance

Technical documentation

Record-keeping (logging)

Transparency and provision of information to users

Human oversight

Accuracy, robustness and cybersecurity

Chapter II: Requirements for high-risk AI systems



Chapter III Obligations of providers and users



Chapters IV and V: Notifying authorities and bodies, standards, conformity assessment, certification, registration

High-risk AI systems



Article 14 Human oversight

No explainability requirements vis-à-vis the affected party, only vis-à-vis the user (= business operator)

No substantive fairness requirements or rights of the affected party

1. High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use.
2. Human oversight shall aim at preventing or minimising the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, in particular when such risks persist notwithstanding the application of other requirements set out in this Chapter.
3. Human oversight shall be ensured through either one or all of the following measures:
 - (a) identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service;
 - (b) identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the user.
4. The measures referred to in paragraph 3 shall enable the individuals to whom human oversight is assigned to do the following, as appropriate to the circumstances:
 - (a) fully understand the capacities and limitations of the high-risk AI system and be able to duly monitor its operation, so that signs of anomalies, dysfunctions and unexpected performance can be detected and addressed as soon as possible;

High-risk AI systems



Article 43 Conformity assessment



Third-party conformity assessment
mandatory only for biometric
techniques (and only under certain
conditions)

Or where third party conformity
assessment is required anyway under
product safety law etc

Is this appropriate?

1. For high-risk AI systems listed in point 1 of Annex III, where, in demonstrating the compliance of a high-risk AI system with the requirements set out in Chapter 2 of this Title, the provider has applied harmonised standards referred to in Article 40, or, where applicable, common specifications referred to in Article 41, the provider shall follow one of the following procedures:
 - (a) the conformity assessment procedure based on internal control referred to in Annex VI;
 - (b) the conformity assessment procedure based on assessment of the quality management system and assessment of the technical documentation, with the involvement of a notified body, referred to in Annex VII.

Where, in demonstrating the compliance of a high-risk AI system with the requirements set out in Chapter 2 of this Title, the provider has not applied or has applied only in part harmonised standards referred to in Article 40, or where such harmonised standards do not exist and common specifications referred to in Article 41 are not available, the provider shall follow the conformity assessment procedure set out in Annex VII.

For the purpose of the conformity assessment procedure referred to in Annex VII, the provider may choose any of the notified bodies. However, when the system is intended to be put into service by law enforcement, immigration or asylum authorities as well as EU institutions, bodies or agencies, the market surveillance authority referred to in Article 63(5) or (6), as applicable, shall act as a notified body.

2. For high-risk AI systems referred to in points 2 to 8 of Annex III, providers shall follow the conformity assessment procedure based on internal control as referred to in Annex VI, which does not provide for the involvement of a notified body. For high-risk AI systems referred to in point 5(b) of Annex III, placed on the market or put into service by credit institutions regulated by Directive 2013/36/EU, the conformity assessment shall be carried out as part of the procedure referred to in Articles 97 to 101 of that Directive.

Transparency-risk AI systems



Article 52

Transparency obligations for certain AI systems

Why are emotion recognition systems not included in Annex III?

1. Providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use. This obligation shall not apply to AI systems authorised by law to detect, prevent, investigate and prosecute criminal offences, unless those systems are available for the public to report a criminal offence.

2. Users of an emotion recognition system or a biometric categorisation system shall inform of the operation of the system the natural persons exposed thereto. This obligation shall not apply to AI systems used for biometric categorisation, which are permitted by law to detect, prevent and investigate criminal offences.

Why are there no further restrictions on use of biometric categorisation (as contrasted with *identification*)?

3. Users of an AI system that generates or manipulates image, audio or video content that appreciably resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful ('deep fake'), shall disclose that the content has been artificially generated or manipulated.

However, the first subparagraph shall not apply where the use is authorised by law to detect, prevent, investigate and prosecute criminal offences or it is necessary for the exercise of the right to freedom of expression and the right to freedom of the arts and sciences guaranteed in the Charter of Fundamental Rights of the EU, and subject to appropriate safeguards for the rights and freedoms of third parties.

4. Paragraphs 1, 2 and 3 shall not affect the requirements and obligations set out in Title III of this Regulation.

Summary

- Many points raised by the DEK have been taken into account and are generally reflected in the AIA (structure of regulation, broad notion of AI; risk-based approach, sectoral supervisory authorities & DPA in charge, ...)
- However, upon closer inspection, the fact that the European Commission uses a similar ‘criticality pyramid’ with similar colours should not distract from the fact that there are very important differences (sectoral/horizontal approach, internal/third-party assessment, little focus on consumer rights and social implications of AI, no explainability requirements vis-à-vis the affected person, no individual rights or private enforcement, ...)
- First reaction largely positive, but a lot remains to be discussed ...