AI & Partners

Amsterdam - London - Singapore

EU AI Act: Legal Text

March 2024

This document (the "Document B") contains relevant provisions from the European Union ("EU") artificial intelligence ("AI") Act (the "EU AI Act") for use alongside the Three-Step Process Document ("T-SP Document"). Document B serves as a reference for the latest legislative text to help organizations and wider stakeholders, such as developers, deployers, providers, authorised representatives and others, understand the TSP Document and how it can be applied to them. References drawn from the EU AI Act reflect the version of text valid as at 2 February 2024, accessible here.

I. General Provisions

Steps 1.1, 1.2, 1.3 and 2.3

2. Scope

1.This Regulation applies to

providers placing on the market or putting into service AI systems or placing on the market general-purpose AI models in the Union, irrespective of whether those providers are established or who are located within the Union or in a third country;

(b) deployers of AI systems that have their place of establishment or who are located within the Union.

(c) providers and deployers of AI systems that have their place of establishment or who are located in a third country, where the output produced by the system is used in the Union;

(ca) importers and distributors of AI systems;

(cb) product manufacturers placing on the market or putting into service an AI system together with their product and under their own name or trademark;

(cc) authorised representatives of providers, which are not established in the Union.

(cc) affected persons that are located in the Union.

2. For AI systems classified as high-risk AI systems in accordance with Articles 6(1) and 6(2) related to products covered by Union harmonisation legislation listed in Annex II, section B only Article 84 of this Regulation shall apply. Article 53 shall apply only insofar as the requirements for high-risk AI systems under this Regulation have been integrated under that Union harmonisation legislation.

"3. This Regulation shall not apply to areas outside the scope of EU law and in any event shall not affect the competences of the Member States concerning national security, regardless of the type of entity entrusted by the Member States to carry out the tasks in relation to those competences.

This Regulation shall not apply to AI systems if and insofar placed on the market, put into service, or used with or without modification of such systems exclusively for military, defence or national security purposes, regardless of the type of entity carrying out those activities.

This Regulation shall not apply to AI systems which are not placed on the market or put into service in the Union, where the output is used in the Union exclusively for military, defence or national security purposes, regardless of the type of entity carrying out those activities. "

4. This Regulation shall not apply to public authorities in a third country nor to international organisations falling within the scope of this Regulation pursuant to paragraph 1, where those authorities or organisations use AI systems in the framework of international cooperation or agreements for law enforcement and judicial cooperation with the Union or with one or more Member States, under the condition that this third country or international organisations provide adequate safeguards with respect to the protection of fundamental rights and freedoms of individuals.

5. This Regulation shall not affect the application of the provisions on the liability of intermediary service providers set out in Chapter II, Section 4 of Directive 2000/31/EC of the European Parliament and of the Council29 [as to be replaced by the corresponding provisions of the Digital Services Act].

5a. This Regulation shall not apply to AI systems and models, including their output, specifically developed and put into service for the sole purpose of scientific research and development.

5a. Union law on the protection of personal data, privacy and the confidentiality of communications applies to personal data processed in connection with the rights and obligations laid down in this Regulation. This Regulation shall not affect Regulations (EU) 2016/679 and (EU) 2018/1725 and Directives 2002/58/EC and (EU) 2016/680, without prejudice to arrangements provided for in Article 10(5) and Article 54 of this Regulation.

5b. This Regulation shall not apply to any research, testing and development activity regarding AI systems or models prior to being placed on the market or put into service; those activities shall be conducted respecting applicable Union law. The testing in real world conditions shall not be covered by this exemption.

5b. This Regulation is without prejudice to the rules laid down by other Union legal acts related to consumer protection and product safety.

5c. This Regulation shall not apply to obligations of deployers who are natural persons using AI systems in the course of a purely personal non-professional activity.

"5e. This Regulation shall not preclude Member States or the Union from maintaining or introducing laws, regulations or administrative provisions which are more favourable to workers in terms of protecting their rights in respect of the use of AI systems by

employers, or to encourage or allow the application of collective agreements which are more favourable to workers. "

5g. The obligations laid down in this Regulation shall not apply to AI systems released under free and open source licences unless they are placed on the market or put into service as high-risk AI systems or an AI system that falls under Title II and IV.

I. General Provisions

3. Definitions

(1) 'AI system' is a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments;

(1a) 'risk' means the combination of the probability of an occurrence of harm and the severity of that harm;

(2) 'provider' means a natural or legal person, public authority, agency or other body that develops an AI system or a general purpose AI model or that has an AI system or a general purpose AI model developed and places them on the market or puts the system into service under its own name or trademark, whether for payment or free of charge;

(4) 'deployer means any natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity;

(5) 'authorised representative' means any natural or legal person located or established in the Union who has received and accepted a written mandate from a provider of an AI system or a general-purpose AI model to, respectively, perform and carry out on its behalf the obligations and procedures established by this Regulation;

(6) 'importer' means any natural or legal person located or established in the Union that places on the market an AI system that bears the name or trademark of a natural or legal person established outside the Union;

(7) 'distributor' means any natural or legal person in the supply chain, other than the provider or the importer, that makes an AI system available on the Union market;

(8) 'operator' means the provider, the product manufacturer, the deployer, the authorised representative, the importer or the distributor;

(9) 'placing on the market' means the first making available of an AI system or a general purpose AI model on the Union market;

(10) 'making available on the market' means any supply of an AI system or a general purpose AI model for distribution or use on the Union market in the course of a commercial activity, whether in return for payment or free of charge;

(11) 'putting into service' means the supply of an AI system for first use directly to the deployer or for own use in the Union for its intended purpose;

(12) 'intended purpose' 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;

(13) 'reasonably foreseeable misuse' means the use of an AI system in a way that is not in accordance with its intended purpose, but which may result from reasonably foreseeable human behaviour or interaction with other systems, including other AI systems;

(14) 'safety component of a product or system' means a component of a product or of a system which fulfils a safety function for that product or system, or the failure or malfunctioning of which endangers the health and safety of persons or property;

(15) 'instructions for use' means the information provided by the provider to inform the user of in particular an AI system's intended purpose and proper use;

(16) 'recall of an AI system' means any measure aimed at achieving the return to the provider or taking it out of service or disabling the use of an AI system made available to deployers;

(17) 'withdrawal of an AI system' means any measure aimed at preventing an AI system in the supply chain being made available on the market;

(18) 'performance of an AI system' means the ability of an AI system to achieve its intended purpose;

(19) 'notifying authority' means the national authority responsible for setting up and carrying out the necessary procedures for the assessment, designation and notification of conformity assessment bodies and for their monitoring;

(20) 'conformity assessment' means the process of demonstrating whether the requirements set out in Title III, Chapter 2 of this Regulation relating to a high-risk AI system have been fulfilled;

(21) 'conformity assessment body' means a body that performs third-party conformity assessment activities, including testing, certification and inspection;

(22) 'notified body' means a conformity assessment body notified in accordance with this Regulation and other relevant Union harmonisation legislation;

(23) 'substantial modification' means a change to the AI system after its placing on the market or putting into service which is not foreseen or planned in the initial conformity assessment by the provider and as a result of which the compliance of the AI system with the requirements set out in Title III, Chapter 2 of this Regulation is affected or results in a modification to the intended purpose for which the AI system has been assessed;

(24) 'CE marking of conformity' (CE marking) means a marking by which a provider indicates that an AI system is in conformity with the requirements set out in Title III, Chapter 2 of this Regulation and other applicable Union legislation harmonising the conditions for the marketing of products ('Union harmonisation legislation') providing for its affixing;

(25) 'post-market monitoring system' means all activities carried out by providers of AI systems to collect and review experience gained from the use of AI systems they place on the market or put into service for the purpose of identifying any need to immediately apply any necessary corrective or preventive actions;

(26) 'market surveillance authority' means the national authority carrying out the activities and taking the measures pursuant to Regulation (EU) 2019/1020;

(27) 'harmonised standard' means a European standard as defined in Article 2(1)(c) of Regulation (EU) No 1025/2012;

(28) 'common specification' means a set of technical specifications, as defined in point 4 of Article 2 of Regulation (EU) No 1025/2012 providing means to comply with certain requirements established under this Regulation;

(29) 'training data' means data used for training an AI system through fitting its learnable parameters;

(30) 'validation data' means data used for providing an evaluation of the trained AI system and for tuning its non-learnable parameters and its learning process, among other things, in order to prevent underfitting or overfitting; whereas the validation dataset is a separate dataset or part of the training dataset, either as a fixed or variable split;

(31) 'testing data' means data used for providing an independent evaluation of the AI system in order to confirm the expected performance of that system before its placing on the market or putting into service;

(32) 'input data' means data provided to or directly acquired by an AI system on the basis of which the system produces an output;

(33) 'biometric data' means personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, such as facial images or dactyloscopic data;

(33a) 'biometric identification' means the automated recognition of physical, physiological, behavioural, and psychological human features for the purpose of establishing an individual's identity by comparing biometric data of that individual to stored biometric data of individuals in a database;

(33c) 'biometric verification' means the automated verification of the identity of natural persons by comparing biometric data of an individual to previously provided biometric data (one-to-one verification, including authentication);

(33d) 'special categories of personal data' means the categories of personal data referred to in Article 9(1) of Regulation (EU) 2016/679, Article 10 of Directive (EU) 2016/680 and Article 10(1) of Regulation (EU) 2018/1725;

(33e) 'sensitive operational data' means operational data related to activities of prevention, detection, investigation and prosecution of criminal offences, the disclosure of which can jeopardise the integrity of criminal proceedings;

(34) 'emotion recognition system' means an AI system for the purpose of identifying or inferring emotions or intentions of natural persons on the basis of their biometric data;

(35) 'biometric categorisation system' means an AI system for the purpose of assigning natural persons to specific categories on the basis of their biometric data unless ancillary to another commercial service and strictly necessary for objective technical reasons;

(36) 'remote biometric identification system' means an AI system for the purpose of identifying natural persons, without their active involvement, typically at a distance through the comparison of a person's biometric data with the biometric data contained in a reference database;

(37) ''real-time' remote biometric identification system' means a remote biometric identification system whereby the capturing of biometric data, the comparison and the identification all occur without a significant delay. This comprises not only instant identification, but also limited short delays in order to avoid circumvention;

(38) ''post' remote biometric identification system' means a remote biometric identification system other than a 'real-time' remote biometric identification system;

(39) 'publicly accessible space' means any publicly or privately owned physical place accessible to an undetermined number of natural persons, regardless of whether certain conditions for access may apply, and regardless of the potential capacity restrictions;

(40) 'law enforcement authority' means:

(a) any public authority competent for the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security; or

(b) any other body or entity entrusted by Member State law to exercise public authority and public powers for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security;

(41) 'law enforcement' means activities carried out by law enforcement authorities or on their behalf for the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security;

(42) 'Artificial Intelligence Office' means the Commission's function of contributing to the implementation, monitoring and supervision of AI systems, general purpose AI models and AI governance. References in this Regulation to the Artificial Intelligence office shall be understood as references to the Commission;

(43) 'national competent authority' means any of the following: the notifying authority and the market surveillance authority. As regards AI systems put into service or used by EU institutions, agencies, offices and bodies, any reference to national competent authorities or market surveillance authorities in this Regulation shall be understood as referring to the European Data Protection Supervisor;

(44) 'serious incident' means any incident or malfunctioning of an AI system that directly or indirectly leads to any of the following:

(a) the death of a person or serious damage to a person's health;

(b) a serious and irreversible disruption of the management and operation of critical infrastructure;

(ba) breach of obligations under Union law intended to protect fundamental rights;

(bb) serious damage to property or the environment.

(44a) 'personal data' means personal data as defined in Article 4, point (1) of Regulation (EU) 2016/679;

(44c) 'non-personal data' means data other than personal data as defined in point (1) of Article 4 of Regulation (EU) 2016/679;
(be) 'profiling' means any form of automated processing of personal data as defined in point (4) of Article 4 of Regulation (EU) 2016/679; or in the case of law enforcement authorities – in point 4 of Article 3 of Directive (EU) 2016/680 or, in the case of Union institutions, bodies, offices or agencies, in point 5 Article 3 of Regulation (EU) 2018/1725;

(bf) 'real world testing plan' means a document that describes the objectives, methodology, geographical, population and temporal scope, monitoring, organisation and conduct of testing in real world conditions;

(44 eb)'sandbox plan' means a document agreed between the participating provider and the competent authority describing the objectives, conditions, timeframe, methodology and requirements for the activities carried out within the sandbox;

(bg) 'AI regulatory sandbox' means a concrete and controlled framework set up by a competent authority which offers providers or prospective providers of AI systems the possibility to develop, train, validate and test, where appropriate in real world conditions, an innovative AI system, pursuant to a sandbox plan for a limited time under regulatory supervision;

(bh) 'AI literacy' refers to skills, knowledge and understanding that allows providers, users and affected persons, taking into account their respective rights and obligations in the context of this Regulation, to make an informed deployment of AI systems, as well as to gain awareness about the opportunities and risks of AI and possible harm it can cause;

(bi) 'testing in real world conditions' means the temporary testing of an AI system for its intended purpose in real world conditions outside of a laboratory or otherwise simulated environment with a view to gathering reliable and robust data and to assessing and verifying the conformity of the AI system with the requirements of this Regulation; testing in real world conditions shall not be considered as placing the AI system on the market or putting it into service within the meaning of this Regulation, provided that all conditions under Article 53 or Article 54a are fulfilled;

(bj) 'subject' for the purpose of real world testing means a natural person who participates in testing in real world conditions;
(bk) 'informed consent' means a subject's freely given, specific, unambiguous and voluntary expression of his or her willingness to participate in a particular testing in real world conditions, after having been informed of all aspects of the testing that are relevant to the subject's decision to participate;

(bl) "deep fake" means AI generated or manipulated image, audio or video content that resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful;

(44e) 'widespread infringement' means any act or omission contrary to Union law that protects the interest of individuals:

(a) which has harmed or is likely to harm the collective interests of individuals residing in at least two Member States other than the Member State, in which:

(i) the act or omission originated or took place;

(ii) the provider concerned, or, where applicable, its authorised representative is established; or

(iii) the deployer is established, when the infringement is committed by the deployer;

(b) which protects the interests of individuals, that have caused, cause or are likely to cause harm to the collective interests of individuals and that have common features, including the same unlawful practice, the same interest being infringed and that are occurring concurrently, committed by the same operator, in at least three Member States;

(44h) 'critical infrastructure' means an asset, a facility, equipment, a network or a system, or a part of thereof, which is necessary for the provision of an essential service within the meaning of Article 2(4) of Directive (EU) 2022/2557;

(44b) 'general purpose AI model' means an AI model, including when trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable to competently perform a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications. This does not cover AI models that are used before release on the market for research, development and prototyping activities;

(44c) 'high-impact capabilities' in general purpose AI models means capabilities that match or exceed the capabilities recorded in the most advanced general purpose AI models;

(44d) 'systemic risk at Union level' means a risk that is specific to the high-impact capabilities of general-purpose AI models, having a significant impact on the internal market due to its reach, and with actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain;

(44e) 'general purpose AI system' means an AI system which is based on a general purpose AI model, that has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems;

(44f) 'floating-point operation' means any mathematical operation or assignment involving floating-point numbers, which are a subset of the real numbers typically represented on computers by an integer of fixed precision scaled by an integer exponent of a fixed base;

(44g) 'downstream provider' means a provider of an AI system, including a generalpurpose AI system, which integrates an AI model, regardless of whether the model is provided by themselves and vertically integrated or provided by another entity based on contractual relations.

II. Prohibited Artificial Intelligence Practices

5. Prohibited Artificial Intelligence Practices

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 or purposefully manipulative or deceptive techniques, with the objective to or the effect of materially distorting a person's or a group of persons' behaviour by appreciably impairing the person's ability to make an informed decision, thereby causing the person to take a decision that that person would not have otherwise taken in a manner that causes or is likely to cause that person, another person or group of persons significant harm;

(b) the placing on the market, putting into service or use of an AI system that exploits any of the vulnerabilities of a person or a specific group of persons due to their age, disability or a specific social or economic situation, with the objective to or the effect of materially distorting the behaviour of that person or a person pertaining to that group in a manner that causes or is reasonably likely to cause that person or another person significant harm;

(ba) the placing on the market or putting into service for this specific purpose, or use of biometric categorisation systems that categorise individually natural persons based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation. This prohibition does not cover any labelling or filtering of lawfully acquired biometric datasets, such as images, based on biometric data or categorizing of biometric data in the area of law enforcement;

(c) the placing on the market, putting into service or use of AI systems for the evaluation or classification of natural persons or groups thereof over a certain period of time based on their social behaviour or known, inferred 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 that are unrelated to the contexts in which the data was originally generated or collected;

(ii) detrimental or unfavourable treatment of certain natural persons or groups thereof that is unjustified or disproportionate to their social behaviour or its gravity;

(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 victims of abduction, trafficking in human beings and sexual exploitation of human beings as well as search for missing persons;

(ii) the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or a genuine and present or genuine and foreseeable threat of a terrorist attack;

(iii) the localisation or identification of a person suspected of having committed a criminal offence, for the purposes of conducting a criminal investigation, prosecution or executing a criminal penalty for offences, referred to in Annex IIa and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least four years. This paragraph is without prejudice to the provisions in Article 9 of the GDPR for the processing of biometric data for purposes other than law enforcement.

(da) the placing on the market, putting into service for this specific purpose, or use of an AI system for making risk assessments of natural persons in order to assess or predict the risk of a natural person to commit a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics. This prohibition shall not apply to AI systems used to support the human assessment of the involvement of a person in a criminal activity, which is already based on objective and verifiable facts directly linked to a criminal activity;

(db) the placing on the market, putting into service for this specific purpose, or use of AI systems that create or expand facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage;

(dc) the placing on the market, putting into service for this specific purpose, or use of AI systems to infer emotions of a natural person in the areas of workplace and education institutions except in cases where the use of the AI system is intended to be put in place or into the market for medical or safety reasons.

1a. This Article shall not affect the prohibitions that apply where an artificial intelligence practice infringes other Union law.

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 only be deployed for the purposes under paragraph 1, point (d) to confirm the specifically targeted individual's identity and it 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 consequences.

In addition, 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 comply with necessary and proportionate safeguards and conditions in relation to the use in accordance with national legislations authorizing the use thereof, in particular as regards the temporal, geographic and personal limitations. The use of the 'real-time' remote biometric identification system in publicly accessible spaces shall only be authorised if the law enforcement authority has completed a fundamental rights impact assessment as provided for in Article 29a and has registered the system in the database according to Article 51. However, in duly justified cases of urgency, the use of the system may be commenced without the registration, provided that the registration is completed without undue delay.

3. As regards paragraphs 1, point (d) and 2, each use for the purpose of law enforcement of a 'real-time' remote biometric identification system in publicly accessible spaces shall be subject to a prior authorisation granted by a judicial authority or an independent administrative authority whose decision is binding of the Member State in which the use is to take place, issued upon a reasoned request and in accordance with the detailed rules of national law referred to in paragraph 4. However, in a duly justified situation of urgency, the use of the system may be commenced without an authorisation provided that such authorisation shall be requested without undue delay, at the latest within 24 hours. If such authorisation is rejected, its use shall be stopped with immediate effect and all the data, as well as the results and outputs of this use shall be immediately discarded and deleted.

The competent judicial authority or an independent administrative authority whose decision is binding shall only grant the authorisation where it is satisfied, based on objective evidence or clear indications presented to it, that the use of the 'real-time' remote biometric identification system at issue is necessary for and proportionate to achieving one of the objectives specified in paragraph 1, point (d), as identified in the request and, in particular, remains limited to what is strictly necessary concerning the period of time as well as geographic and personal scope. In deciding on the request, the competent judicial authority or an independent administrative authority whose decision is binding shall takeinto account the elements referred to in paragraph 2. It shall be ensured that no decision that produces an adverse legal effect on a person may be taken by the judicial authority or an independent administrative authority whose decision is binding solely based on the output of the remote biometric identification system.

3a. Without prejudice to paragraph 3, each use of a 'real-time' remote biometric identification system in publicly accessible spaces for law enforcement purposes shall be notified to the relevant market surveillance authority and the national data protection authority in accordance with the national rules referred to in paragraph 4. The notification shall as a minimum contain the information specified under paragraph 5 and shall not include sensitive operational data.

4. A Member State may decide to provide for the possibility to fully or partially authorise the use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement within the limits and under the conditions listed in paragraphs 1, point (d), 2 and 3. Member States concerned shall lay down in their national law the necessary detailed rules for the request, issuance and exercise of, as well as supervision and reporting relating to, the authorisations referred to in paragraph 3. Those rules shall also specify in respect of which of the objectives listed in paragraph 1, point (d), including which of the criminal offences referred to in point (iii) thereof, the competent authorities may be authorised to use those systems for the purpose of law enforcement. Member States shall notify those rules to the Commission at the latest 30 days following the adoption thereof. Member States may introduce, in accordance with Union law, more restrictive laws on the use of remote biometric identification systems.

"5. National market surveillance authorities and the national data protection authorities of Member States that have been notified of the use of 'real-time' remote biometric identification systems in publicly accessible spaces for law enforcement purposes pursuant to paragraph 3a shall submit to the Commission annual reports on such use. For that purpose, the Commission shall provide Member States and national market surveillance and data protection authorities with a template, including information on the number of the

decisions taken by competent judicial authorities or an independent administrative authority whose decision is binding upon requests for authorisations in accordance with paragraph 3 and their result."

6. The Commission shall publish annual reports on the use of 'real-time' remote biometric identification systems in publicly accessible spaces for law enforcement purposes based on aggregated data in Member States based on the annual reports referred to in paragraph 5, which shall not include sensitive operational data of the related law enforcement activities.

III. High-Risk AI System

Step 2.1

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 the AI system is itself a product, covered by the Union harmonisation legislation listed in Annex II;

(b) the product whose safety component pursuant to point (a) 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.

2a. By derogation from paragraph 2 AI systems shall not be considered as high risk if they do not pose a significant risk of harm, to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making. This shall be the case if one or more of the following criteria are fulfilled:

(a) the AI system is intended to perform a narrow procedural task;

(b) the AI system is intended to improve the result of a previously completed human activity;

(c) the AI system is intended to detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment, without proper human review; or

(d) the AI system is intended to perform a preparatory task to an assessment relevant for the purpose of the use cases listed in Annex III.

Notwithstanding first subparagraph of this paragraph, an AI system shall always be considered high-risk if the AI system performs profiling of natural persons.

2b. A provider who considers that an AI system referred to in Annex III is not high-risk shall document its assessment before that system is placed on the market or put into service. Such provider shall be subject to the registration obligation set out in Article 51(1a). Upon request of national competent authorities, the provider shall provide the documentation of the assessment.

2c. The Commission shall, after consulting the AI Board, and no later than 18 months after the entry into force of this Regulation, provide guidelines specifying the practical implementation of this article completed by a comprehensive list of practical examples of high risk and non-high risk use cases on AI systems in accordance with the conditions set out in Article 82a.

2d. The Commission is empowered to adopt delegated acts in accordance with Article 73 to amend the criteria laid down in points (a) to (d) of the first subparagraph of paragraph 2a.

The Commission may adopt delegated acts adding new criteria to those laid down in points (a) to (d) of the first subparagraph of paragraph 2a, or modifying them, only where there is concrete and reliable evidence of the existence of AI systems that fall under the scope of Annex III but that do not pose a significant risk of harm to the health, safety and fundamental rights.

The Commission shall adopt delegated acts deleting any of the criteria laid down in the first subparagraph of paragraph 2a where there is concrete and reliable evidence that this is necessary for the purpose of maintaining the level of protection of health, safety and fundamental rights in the Union.

Any amendment to the criteria laid down in points (a) to (d) set out in the first subparagraph of paragraph 2a shall not decrease the overall level of protection of health, safety and fundamental rights in the Union.

When adopting the delegated acts, the Commission shall ensure consistency with the delegated acts adopted pursuant to Article 7(1) and shall take account of market and technological developments.

IV. Transparency Obligations for Providers and Deployers of Certain AI Systems

52. Transparency obligations for providers and users of certain AI systems and GPAI models

1. Providers shall ensure that AI systems intended to directly interact with natural persons are designed and developed in such a way that the concerned natural persons are informed that they are interacting with an AI system, unless this is obvious from the point of view of a natural person who is reasonably well-informed, observant and circumspect, taking into account 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, subject to appropriate safeguards for the rights and freedoms of third parties unless those systems are available for the public to report a criminal offence.

1a. Providers of AI systems, including GPAI systems, generating synthetic audio, image, video or text content, shall ensure the outputs of the AI system are marked in a machinereadable format and detectable as artificially generated or manipulated. Providers shall ensure their technical solutions are effective, interoperable, robust and reliable as far as this is technically feasible, taking into account specificities and limitations of different types of content, costs of implementation and the generally acknowledged state-of-the-art, as may be reflected in relevant technical standards. This obligation shall not apply to the extent the AI systems perform an assistive function for standard editing or do not substantially alter the input data provided by the deployer or the semantics thereof, or where authorised by law to detect, prevent, investigate and prosecute criminal offences.

2. Deployers of an emotion recognition system or a biometric categorisation system shall inform of the operation of the system the natural persons exposed thereto and process the personal data in accordance with Regulation (EU) 2016/679, Regulation (EU) 2016/1725 and Directive (EU) 2016/280, as applicable. This obligation shall not apply to AI systems used for biometric categorization and emotion recognition, which are permitted by law to detect, prevent and investigate criminal offences, subject to appropriate safeguards for the rights and freedoms of third parties, and in compliance with Union law.

3. Deployers of an AI system that generates or manipulates image, audio or video content constituting a deep fake, shall disclose that the content has been artificially generated or manipulated. This obligation shall not apply where the use is authorised by law to detect, prevent, investigate and prosecute criminal offence. Where the content forms part of an evidently artistic, creative, satirical, fictional analogous work or programme, the transparency obligations set out in this paragraph are limited to disclosure of the existence of such generated or manipulated content in an appropriate manner that does not hamper the display or enjoyment of the work.

Deployers of an AI system that generates or manipulates text which is published with the purpose of informing the public on matters of public interest shall disclose that the text has been artificially generated or manipulated. This obligation shall not apply where the use is authorised by law to detect, prevent, investigate and prosecute criminal offences or where the AI-generated content has undergone a process of human review or editorial control and where a natural or legal person holds editorial responsibility for the publication of the content.

3a. The information referred to in paragraphs 1 to 3 shall be provided to the concerned natural persons in a clear and distinguishable manner at the latest at the time of the first interaction or exposure. The information shall respect the applicable accessibility requirements.

4. Paragraphs 1, 2 and 3 shall not affect the requirements and obligations set out in Title III of this Regulation and shall be without prejudice to other transparency obligations for users of AI systems laid down in Union or national law.

4a. The AI Office shall encourage and facilitate the drawing up of codes of practice at Union level to facilitate the effective implementation of the obligations regarding the detection and labelling of artificially generated or manipulated content. The Commission is empowered to adopt implementing acts to approve these codes of practice in accordance with the procedure laid down in Article 52e paragraphs 6-8. If it deems the code is not adequate, the Commission is empowered to adopt an implementing act specifying the common rules for the implementation of those obligations in accordance with the examination procedure laid down in Article 73 paragraph 2.

Step 2.2

IVA. General-Purpose AI Models

1. Classification Rules

52a. Classification of general-purpose AI models as general purpose AI models with systemic risk

1. A general purpose AI model shall be classified as general-purpose AI model with systemic risk if it meets any of the following criteria:

(a) it has high impact capabilities evaluated on the basis of appropriate technical tools and methodologies, including indicators and benchmarks;

(b) based on a decision of the Commission, ex officio or following a qualified alert by the scientific panel that a general purpose AI model has capabilities or impact equivalent to those of point (a).

2. A general purpose AI model shall be presumed to have high impact capabilities pursuant to point a) of paragraph 1 when the cumulative amount of compute used for its training measured in floating point operations (FLOPs) is greater than 10^25.

3. The Commission shall adopt delegated acts in accordance with Article 73(2) to amend the thresholds listed in the paragraphs above, as well as to supplement benchmarks and indicators in light of evolving technological developments, such as algorithmic improvements or increased hardware efficiency, when necessary, for these thresholds to reflect the state of the art.

VIII. Post-Market Monitoring, Information Sharing, Market Surveillance

3. Enforcement

65a. Procedure for dealing with AI systems classified by the provider as a not high-risk in application of Annex III

1. Where a market surveillance authority has sufficient reasons to consider that an AI system classified by the provider as non-high-risk in application of Annex III is high-risk, they market surveillance authority shall carry out an evaluation of the AI system concerned in respect of its classification as a high-risk AI system based on the conditions set out in Annex III and the Commission guidelines.

2. Where, in the course of that evaluation, the market surveillance authority finds that the AI system concerned is high-risk, it shall without undue delay require the relevant provider to take all necessary actions to bring the AI system into compliance with the requirements and obligations laid down in this Regulation as well as take appropriate corrective action within a period it may prescribe.

3. Where the market surveillance authority considers that the use of the AI system concerned is not restricted to its national territory, it shall inform the Commission and the other Member States without undue delay of the results of the evaluation and of the actions which it has required the provider to take.

4. The provider shall ensure that all necessary action is taken to bring the AI system into compliance with the requirements and obligations laid down in this Regulation. Where the provider of an AI system concerned does not bring the AI system into compliance with the requirements and obligations of this Regulation within the period referred to in paragraph 2, the provider shall be subject to fines in accordance with Article 71.

5. The provider shall ensure that all appropriate corrective action is taken in respect of all the AI systems concerned that it has made available on the market throughout the Union.

6. Where the provider of the AI system concerned does not take adequate corrective action within the period referred to in paragraph 2, then the provisions of Article 65 paragraphs 5 to 9 apply.

7. Where, in the course of that evaluation pursuant to paragraph 1, the market surveillance authority establishes that the AI system was misclassified by the provider as not high-risk to circumvent the application of requirements in Title III, Chapter 2, the provider shall be subject to fines in accordance with Article 71.

8. In exercising their power to monitor the application of this article and in accordance with Article 11 of Regulation (EU) 2019/1020, market surveillance authorities may perform appropriate checks, taking into account in particular information stored in the EU database referred to in Article 60.

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. Biometrics, insofar as their use is permitted under relevant Union or national law:

(a) Remote biometric identification systems. This shall not include AI systems intended to be used for biometric verification whose sole purpose is to confirm that a specific natural person is the person he or she claims to be;

(aa) AI systems intended to be used for biometric categorisation, according to sensitive or protected attributes or characteristics based on the inference of those attributes or characteristics;

(ab) AI systems intended to be used for emotion recognition.

2. Critical infrastructure:

(a) AI systems intended to be used as safety components in the management and operation of critical digital infrastructure, road traffic and the supply of water, gas, heating and electricity.

3. Education and vocational training:

(a) AI systems intended to be used to determine access or admission or to assign natural persons to educational and vocational training institutions at all levels;

(b) AI systems intended to be used to evaluate learning outcomes, including when those outcomes are used to steer the learning process of natural persons in educational and vocational training institutions at all levels;

(ba) AI systems intended to be used for the purpose of assessing the appropriate level of education that individual will receive or will be able to access, in the context of/within education and vocational training institution;

(bb) AI systems intended to be used for monitoring and detecting prohibited behaviour of students during tests in the context of/within education and vocational training 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 to place targeted job advertisements, to analyse and filter job applications, and to evaluate candidates;

(b) AI intended to be used to make decisions affecting terms of the work related relationships, promotion and termination of work-related contractual relationships, to allocate tasks based on individual behaviour or personal traits or characteristics and to monitor and evaluate performance and behaviour of persons in such relationships.

5. Access to and enjoyment of essential private services and essential public services and benefits:

(a) AI systems intended to be used by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for essential public assistance benefits and services, including healthcare services, as well as to grant, reduce, revoke, or reclaim such benefits and services;

(b) AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems used for the purpose of detecting financial fraud;

(c) AI systems intended to evaluate and classify emergency calls by natural persons or to be used to dispatch, or to establish priority in the dispatching of emergency first response services, including by police, firefighters and medical aid, as well as of emergency healthcare patient triage systems;

(ca) AI systems intended to be used for risk assessment and pricing in relation to natural persons in the case of life and health insurance.

6. Law enforcement, insofar as their use is permitted under relevant Union or national law:

(a) AI systems intended to be used by or on behalf of law enforcement authorities, or by Union institutions, agencies, offices or bodies in support of law enforcement authorities or on their behalf to assess the risk of a natural person to become a victim of criminal offences;

(b) AI systems intended to be used by or on behalf of law enforcement authorities or by Union institutions, bodies and agencies in support of Law enforcement authorities as polygraphs and similar tools;

(d) AI systems intended to be used by or on behalf of law enforcement authorities, or by Union institutions, agencies, offices or bodies in support of law enforcement authorities to evaluate the reliability of evidence in the course of investigation or prosecution of criminal offences;

(e) AI systems intended to be used by law enforcement authorities or on their behalf or by Union institutions, agencies, offices or bodies in support of law enforcement authorities for assessing the risk of a natural person of offending or re-offending not solely based on profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 or to assess personality traits and characteristics or past criminal behaviour of natural persons or groups;

(f) AI systems intended to be used by or on behalf of law enforcement authorities or by Union agencies institutions, agencies, offices or bodies in support of law enforcement authorities for profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 in the course of detection, investigation or prosecution of criminal offences.

7. Migration, asylum and border control management, insofar as their use is permitted under relevant Union or national law:

(a) AI systems intended to be used by competent public authorities as polygraphs and similar tools;

Recitals

(1) The purpose of this Regulation is to improve the functioning of the internal market by laying down a uniform legal framework in particular for the development, placing on the market, putting into service and the use of artificial intelligence systems in the Union in conformity with Union values, to promote the uptake of human centric and trustworthy artificial intelligence while ensuring a high level of protection of health, safety, fundamental rights enshrined in the Charter, including democracy and rule of law and environmental protection, against harmful effects of artificial intelligence systems in the Union and to support innovation. This regulation ensures the free movement of AI-based goods and services cross-border, thus preventing Member States from imposing restrictions on the development, marketing and use of Artificial Intelligence systems (AI systems), unless explicitly authorised by this Regulation.

(5) A Union legal framework laying down harmonised rules on artificial intelligence is therefore needed to foster the development, use and uptake of artificial intelligence in the internal market that at the same time meets a high level of protection of public interests, such as health and safety and the protection of fundamental rights, including democracy, rule of law and environmental protection as recognised and protected by Union law. To achieve that objective, rules regulating the placing on the market, putting into service and use of certain AI systems should be laid down, thus ensuring the smooth functioning of the internal market and allowing those systems to benefit from the principle of free movement of goods and services. These rules should be clear and robust in protecting fundamental rights, supportive of new innovative solutions, enabling to a European ecosystem of public and private actors creating AI systems in line with Union values and unlocking the potential of the digital transformation across all regions of the Union. By laying down those rules as well as measures in support of innovation with a particular focus on SMEs including startups, this Regulation supports the objective of promoting the European human-centric approach to AI and being a global leader in the development of secure, trustworthy and ethical artificial intelligence as stated by the European Council4 , and it ensures the protection of ethical principles, as specifically requested by the the European Parliament5 .

(6) The notion of AI system in this Regulation should be clearly defined and closely aligned with the work of international organisations working on artificial intelligence to ensure legal certainty, facilitate international convergence and wide acceptance, while providing the flexibility to accommodate the rapid technological developments in this field. Moreover, it should be based on key characteristics of artificial intelligence systems, that distinguish it from simpler traditional software systems or programming approaches and should not cover systems that are based on the rules defined solely by natural persons to automatically execute operations. A key characteristic of AI systems is their capability to infer. This inference refers to the process of obtaining the outputs, such as predictions, content, recommendations, or decisions, which can influence physical and virtual environments and to a capability of AI systems to derive models and/or algorithms from inputs/data. The techniques that enable inference while building an AI system include machine learning approaches that learn from data how to achieve certain objectives; and logic- and knowledge-based approaches that infer from encoded knowledge or symbolic representation of the task to be solved. The capacity of an AI system to infer goes beyond basic data processing, enable learning, reasoning or modelling. The term “machine-based” refers to the fact that AI systems run on machines. The reference

to explicit or implicit objectives underscores that AI systems can operate according to explicit defined objectives or to implicit objectives. The objectives of the AI system may be different from the intended purpose of the AI system in a specific context. For the purposes of this Regulation, environments should be understood as the contexts in which the AI systems operate, whereas outputs generated by the AI system, reflect different functions performed by AI systems and include predictions, content, recommendations or decisions. AI systems are designed to operate with varying levels of autonomy, meaning that they have some degree of independence of actions from human involvement and of capabilities to operate without human intervention. The adaptiveness that an AI system could exhibit after deployment, refers to self-learning capabilities, allowing the system to change while in use. AI systems can be used on a stand-alone basis or as a component of a product, irrespective of whether the system is physically integrated into the product (embedded) or serve the functionality of the product without being integrated therein (non-embedded).

(60c) Large generative AI models are a typical example for a general-purpose AI model, given that they allow for flexible generation of content (such as in the form of text, audio, images or video) that can readily accommodate a wide range of distinctive tasks.

(60i) Software and data, including models, released under a free and open-source licence that allows them to be openly shared and where users can freely access, use, modify and redistribute them or modified versions thereof, can contribute to research and innovation in the market and can provide significant growth opportunities for the Union economy. General purpose AI models released under free and open-source licences should be considered to ensure high levels of transparency and openness if their parameters, including the weights, the information on the model architecture, and the information on model usage are made publicly available. The licence should be considered free and open-source also when it allows users to run, copy, distribute, study, change and improve software and data, including models under the condition that the original provider of the model is credited, the identical or comparable terms of distribution are respected.

(60i+1) Free and open-source AI components covers the software and data, including models and general purpose AI models, tools, services or processes of an AI system. Free and opensource AI components can be provided through different channels, including their development on open repositories. For the purpose of this Regulation, AI components that are provided against a price or otherwise monetised, including through the provision of technical support or other services, including through a software platform, related to the AI component, or the use of personal data for reasons other than exclusively for improving the security, compatibility or interoperability of the software, with the exception of transactions between micro enterprises, should not benefit from the exceptions provided to free and open source AI components. The fact of making AI components available through open repositories should not, in itself, constitute a monetisation.

(70a) A variety of AI systems can generate large quantities of synthetic content that becomes increasingly hard for humans to distinguish from human-generated and authentic content. The wide availability and increasing capabilities of those systems have a significant impact on the integrity and trust in the information ecosystem, raising new risks of misinformation and manipulation at scale, fraud, impersonation and consumer deception. In the light of those impacts, the fast technological pace and the need for new methods and techniques to trace origin of information, it is appropriate to require providers of those systems to embed technical solutions that enable marking in a machine readable format and detection that the output has been generated or manipulated by an AI system and not a human. Such techniques and methods should be sufficiently reliable, interoperable, effective and robust as far as this is technically feasible, taking into account available techniques or a combination of such techniques, such as watermarks, metadata identifications, cryptographic methods for proving provenance and authenticity of content, logging methods, fingerprints or other techniques, as may be appropriate. When implementing this obligation, providers should also take into account the specificities and the limitations of the different types of content and the relevant technological and market developments in the field, as reflected in the generally acknowledged state-of-the-art. Such techniques and methods can be implemented at the level of the system or at the level of the model, including general-purpose AI models generating content, thereby facilitating fulfilment of this obligation by the downstream provider of the AI system. To remain proportionate, it is appropriate to envisage that this marking obligation should not cover AI systems performing primarily an assistive function for standard editing or AI systems not substantially altering the input data provided by the deployer or the semantics thereof.

Disclaimer

Document B is provided for informational purposes only. The data and provisions within the Document B are not endorsements or guarantees. The author does not assume any liability for the timeliness, accuracy, completeness or quality of the information provided. Liability claims against the author relating to material or non-material damage caused by the use or non-use of the information provided or by the use of incorrect or incomplete information are generally excluded, insofar as there is no demonstrable fault of an intentional or negligent nature on the part of the author. All content on the Document B is non-binding and is subject to change. The author expressly reserves the right to change, supplement or delete parts of the pages or the entire Document B without separate notice or to suspend publication temporarily or permanently. Contact data and/or

names published within the Document B or comparable details such as postal addresses, telephone and fax numbers as well as e-mail addresses may not be used by third parties for the purpose of sending information that has not been expressly requested. Users should conduct independent research and seek professional advice where necessary.