EUROPEAN
COMMISSION

Brussels, 4.2.2025
C(2025) 884 final

ANNEX

ANNEX

to the

Communication to the Commission

Approval of the content of the draft Communication from the Commission - Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)

CONTENTS

1. BACKGROUND AND OBJECTIVES

(1) Regulation (EU) 2024/1689 of the European Parliament and the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending certain regulations ('the AI Act')[1] entered into force on 1 August 2024. The AI Act lays down harmonised rules for the placing on the market, putting into service, and use of artificial intelligence ('AI') in the Union.[2] Its aim is to promote innovation in and the uptake of AI, while ensuring a high level of protection of health, safety and fundamental rights in the Union, including democracy and the rule of law.

(2) The AI Act follows a risk-based approach, classifying AI systems into four different risk categories:

(i) Unacceptable risk: AI systems posing unacceptable risks to fundamental rights and Union values are prohibited under Article 5 AI Act.

(ii) High risk: AI systems posing high risks to health, safety and fundamental rights are subject to a set of requirements and obligations. These systems are classified as 'high-risk' in accordance with Article 6 AI Act in conjunction with Annexes I and III AI Act.

(iii) Transparency risk: AI systems posing limited transparency risk are subject to transparency obligations under Article 50 AI Act.

(iv) Minimal to no risk: AI systems posing minimal to no risk are not regulated, but providers and deployers may voluntarily adhere to voluntary codes of conduct.[3]

(3) Pursuant to Article 96(1)(b) AI Act, the Commission is to adopt guidelines on the practical implementation of the practices prohibited under Article 5 AI Act. Those prohibitions apply six months after the entry into force of the AI Act, i.e. as from 2 February 2025.

(4) These Guidelines aim to increase legal clarity and to provide insights into the Commission's interpretation of the prohibitions in Article 5 AI Act with a view to ensuring their consistent, effective and uniform application. They should serve as practical guidance to assist competent authorities under the AI Act in their enforcement activities, as well as providers and deployers of AI systems in ensuring compliance with their obligations under the AI Act. They strive to interpret the prohibitions in a proportionate manner that achieves the objectives of the AI Act to protect fundamental rights and safety, while promoting innovation and providing legal certainty.

(5) These Guidelines are non-binding. Any authoritative interpretation of the AI Act may ultimately only be given by the Court of Justice of the European Union ('CJEU').

[1] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) (OJ L, 2024/1689, 12.7.2024).
[2] Article 1 AI Act.
[3] Article 95 AI Act.

(6) The drafting of these Guidelines was informed by input from a variety of stakeholders, e.g., providers and deployers of AI systems, civil society organisations, academia, public authorities, business associations, etc., collected during a broad consultation process organised by the Commission. The Member States within the AI Board and the European Parliament were also consulted. These Guidelines will be regularly reviewed in light of the experience gained from the practical implementation of Article 5 AI Act and technological and market developments.

(7) The application of Article 5 AI Act will require a case-by-case assessment, which takes due account of the specific situation at issue in an individual case. Therefore, the examples given in these Guidelines are merely indicative and are without prejudice to the need for such an assessment in each case.

2. OVERVIEW OF PROHIBITED AI PRACTICES

(8) Article 5 AI Act prohibits the placing on the EU market, putting into service, or use of certain AI systems for manipulative, exploitative, social control or surveillance practices, which by their inherent nature violate fundamental rights and Union values. Recital 28 AI Act clarifies that such practices are particularly harmful and abusive and should be prohibited because they contradict the Union values of respect for human dignity, freedom, equality, democracy, and the rule of law, as well as fundamental rights enshrined in the Charter of Fundamental Right of the European Union ('the Charter'), including the right to non-discrimination (Article 21 Charter) and equality (Article 20), data protection (Article 8 Charter) and private and family life (Article 7 Charter), and the rights of the child (Article 24 Charter). The prohibitions in Article 5 AI Act also aim to uphold the right to freedom of expression and information (Article 11 Charter), freedom of assembly and of association (Article 12 Charter), freedom of thought, conscience and religion (Article 10 Charter), the right to an effective remedy and fair trial (Article 47 Charter), and the presumption of innocence and the right of defence (Article 48 Charter).

2.1. Prohibitions listed in Article 5 AI Act

(9) **Overview on the Prohibitions**

Provision	Prohibition	Content
Article 5(1)(a)	Harmful manipulation, and deception	AI systems that deploy subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective or with the effect of distorting behaviour, causing or reasonably likely to cause significant harm
Article 5(1)(b)	Harmful exploitation of vulnerabilities	AI systems that exploit vulnerabilities due to age, disability or a specific social or economic situation, with the objective or with the effect of distorting behaviour, causing or reasonably likely to cause significant harm

Article 5(1)(c)	Social scoring	AI systems that evaluate or classify natural persons or groups of persons based on social behaviour or personal or personality characteristics, with the social score leading to detrimental or unfavourable treatment when data comes from unrelated social contexts or such treatment is unjustified or disproportionate to the social behaviour
Article 5(1)(d)	Individual criminal offence risk assessment and prediction	AI systems that assess or predict the risk of people committing a criminal offence based solely on profiling or personality traits and characteristics; except to support a human assessment based on objective and verifiable facts directly linked to a criminal activity
Article 5(1)(e)	Untargeted scraping to develop facial recognition databases	AI systems that create or expand facial recognition databases through untargeted scraping of facial images from the internet or closed-circuit television ('CCTV') footage
Article 5(1)(f)	Emotion recognition	AI systems that infer emotions at the workplace or in education institutions; except for medical or safety reasons
Article 5(1)(g)	Biometric categorisation	AI systems that categorise people 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; except for labelling or filtering of lawfully acquired biometric datasets, including in the area of law enforcement
Article 5(1)(h)	Real-time remote biometric identification ('RBI')	AI systems for real-time remote biometric identification in publicly accessible spaces for the purposes of law enforcement; except if necessary for the targeted search of specific victims, the prevention of specific threats including terrorist attacks, or the search of suspects of specific offences (further procedural requirements, including for authorisation, outlined in Article 5(2-7) AI Act).

2.2. Legal basis of the prohibitions

(10) The AI Act is supported by two legal bases: Article 114 of the Treaty on the Functioning of the European Union ('TFEU') (the internal market legal basis) and Article 16 TFEU (the data protection legal basis). Article 16 TFEU serves as a legal basis for the specific rules on the processing of personal data in relation to the prohibition on the use of remote biometric identification ('RBI') systems for law enforcement purposes, biometric categorisation systems for law enforcement purposes, and individual risk

assessments for law enforcement purposes.[4] All other prohibitions listed in Article 5 AI Act find their legal basis in Article 114 TFEU.

2.3. Material scope: practices related to the 'placing on the market', 'putting into service' or 'use' of an AI system

(11) The practices prohibited by Article 5 AI Act relate to the placing on the market, the putting into service, or the use of specific AI systems.[5] As regards real-time remote biometric identification ('RBI') systems, the prohibition in Article 5(1)(h) AI Act only applies to their use. Article 3(1) AI Act defines what constitutes an AI system. The Guidelines on the Definition of an AI system provide the Commission's interpretation of that definition.

(12) According to Article 3(9) AI Act, the **placing on the market** of an AI system is 'the first making available of an AI system […] on the Union market'. 'Making available' is defined as the supply of the system 'for distribution or use on the Union market in the course of a commercial activity, whether in return for payment or free of charge.'[6] The making available of an AI system is covered regardless of the means of supply, such as access to the system and its service through an application programming interface ('API'), via cloud, direct downloads, as physical copies, or embedded in physical products.

> For example, a RBI system developed outside the Union by a third-country provider is placed on the Union market for the first time when it is offered in return for payment or free of charge in one or more Member States. Such placing on the market may occur by providing access to the system online through an API or other user interface.

(13) Article 3(11) AI Act defines **putting into service** as 'the supply of an AI system for first use to the deployer or for own use in the Union for its intended purpose', therefore covering both supply for first use to third parties, as well as in-house development and deployment. The intended purpose of the system is 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 of use, promotional or sales materials and statements, as well as in the technical documentation.'[7]

[4] Recital 3 AI Act. Regarding the prohibitions based on Article 16 of the TFEU, there are two relevant opt outs for Ireland and Denmark. With the discretion granted to Ireland under Protocol No. 21 on the position of the United Kingdom and Ireland in the area of freedom, security and justice (AFSJ) annexed to the TEU and TFEU, Ireland may decide not to apply the rules concerning the prohibition of real-time use of RBIs in public spaces for a law enforcement purpose as well as the procedural rules linked to that article (Article 5(2) to (6) AI Act) (see Recital 40). Denmark benefits from opt-out agreements when applying Protocol No. 22 to the TEU and TFEU and may decide not to fully apply the prohibitions based on Article 16 of the TFEU (see Recital 41).

[5] See for definitions of these terms also the Commission Notice – The 'Blue Guide' on the implementation of EU product rules 2022, 2022/C 247/01, Section 2.

[6] Article 3(10) AI Act.

[7] Article 3(12) AI Act.

For example, a provider builds a RBI system outside the Union and supplies that system to a law enforcement authority or to a private company established in a Member State to be used for the first time, thereby putting it into service.

For example, a public authority develops a scoring system in-house and deploys it to predict the risk of fraud of household allowance beneficiaries, thereby putting it into service.

(14) While the '**use'** of an AI system is not explicitly defined in the AI Act, it should be understood in a broad manner to cover the use or deployment of the system at any moment of its lifecycle after having been placed on the market or put into service. This may also cover the integration of the AI system in the services and processes of the person(s) making use of the AI system, including as part of more complex systems, processes or infrastructure. While providers of AI systems must consider the conditions of use which may be reasonably foreseen prior to placing their AI systems on the market (intended use and reasonably foreseeable misuse[8]), deployers remain responsible for taking the lawful conditions for the use of the system into account.[9] For the purposes of Article 5 AI Act, the reference to 'use' should be understood to include any misuse of an AI system ('reasonably foreseeable' or not) that may amount to a prohibited practice.[10]

For example, an AI system used by an employer to infer emotions at the workspace is prohibited, except when used for medical or safety purposes (Article 5(1)(f) AI Act). The prohibition applies to deployers regardless of whether the provider (the supplier of the system) has excluded such use in its contractual relationships with the deployer (the employer), i.e. in the terms of use.

2.4. Personal scope: responsible actors

(15) The AI Act distinguishes between different categories of operators in relation to AI systems: providers, deployers, importers, distributors, and product manufacturers. The present Guidelines will focus only on providers and deployers given the scope of the prohibited practices in Article 5 AI Act.

(16) According to Article 3(3) AI Act, **providers** are natural or legal persons, public authorities, agencies or other bodies, that develop AI systems or have them developed and place them on the Union market, or put them into service under their own name or trademark[11] (see section 2.3 above). Providers established or located outside the Union

[8] See Article 3(12) and (13) AI Act.
[9] See for definitions of these terms also the Commission Notice – The 'Blue Guide' on the implementation of EU product rules 2022, 2022/C 247/01, Section 2.8.
[10] Recital 28 AI Act.
[11] Article 3(3), (9) and (11) AI Act. In relation to high-risk AI systems, Article 25 AI Act envisages that 1. Any distributor, importer, deployer or other third-party shall be considered to be a provider of a high-risk AI system for the purposes of this Regulation and shall be subject to the obligations of the provider under Article 16, in any of the following circumstances: (a) they put their name or trademark on a high-risk AI system already placed on the market or put into service, without prejudice to contractual arrangements stipulating that the obligations are otherwise allocated; (b) they make a substantial modification to a high-risk AI system that has already been placed on the market or has already

are subject to the provisions of the AI Act if they place those systems on the market or put them into service in the Union,[12] or if the output of the AI system is used in the Union[13]. Providers must ensure their AI systems meet all relevant requirements before placing them on the market or putting them into service.

For example, a provider of a RBI system is the manufacturer of the system that markets the system in the Union under its trademark. The provider of such a system could also be a public authority that develops the system in-house and puts it into service for its own use.

(17) **Deployers** are natural or legal persons, public authorities, agencies or other bodies using AI systems under their authority, unless the use is for a personal non-professional activity.[14] 'Authority' over an AI system should be understood as assuming responsibility over the decision to deploy the system and over the manner of its actual use. Deployers fall within the scope of the AI Act, if their place of establishment or location is within the Union[15] or, if they are located in a third country, the output of the AI system is used in the Union[16].

(18) Where the deployer of an AI system is a legal person under whose authority the system is used, i.e. a law enforcement authority or a private security company, the individual employees that act within the procedures and under the control of that person should not be considered to be the deployer. A legal person also remains a deployer if it involves third parties (e.g., contractors, external staff) in the operation of the system on its behalf and under its responsibility and control.

(19) Operators may fulfil more than one role concurrently in relation to an AI system. For example, if an operator develops its own AI system that it uses afterwards, it will be considered both the provider and the deployer of that system, even if that system is also used by other deployers to whom the system has been provided in return for payment or free of charge.

(20) Continuous compliance with the AI Act is required during all phases of the AI lifecycle. This necessitates ongoing monitoring of and updates to AI systems placed on the market or put into service in the Union to ensure that an AI system remains compliant with the AI Act throughout its lifecycle and that it does not result in a practice prohibited under Article 5 AI Act. Providers and deployers of AI systems have different responsibilities depending on their roles and control over the design, the development and the actual use of the system to avoid a prohibited practice. For each of the prohibitions, these roles and responsibilities should be interpreted in a proportionate manner, taking into account

been put into service in such a way that it remains a high-risk AI system pursuant to Article 6; (c) they modify the intended purpose of an AI system, including a general-purpose AI system, which has not been classified as high-risk and has already been placed on the market or put into service in such a way that the AI system concerned becomes a high-risk AI system in accordance with Article 6.

[12] Article 2(1)(a) AI Act.
[13] Article 2(1)(c) AI Act.
[14] Article 3(4) AI Act.
[15] Article 2(1)(b) AI Act.
[16] Article 2(1)(c) AI Act.

who in the value chain is best placed to adopt specific preventive and mitigating measures and ensure compliant development and use of AI systems in line with the objectives and the approach of the AI Act.

2.5. Exclusion from the scope of the AI Act

(21) Article 2 AI Act provides for a number of general exclusions from scope which are relevant for a complete understanding of the practical application of the prohibitions listed in Article 5 AI Act.

2.5.1. National security, defence and military purposes

(22) According to Article 2(3) AI Act, the AI Act does not apply to areas outside the scope of Union law, and should not, in any event, affect the competences of the Member States concerning national security, regardless of the type of entity entrusted by the Member States with carrying out tasks in relation to those competences. The AI Act expressly excludes from its scope AI systems that are 'placed on the market, put into service, or used with or without modification exclusively for military, defence or national security purposes, regardless of the type of entity carrying out those activities.' Whether that exclusion applies therefore depends on the purposes or the uses of the AI system, not the entities carrying out the activities with that system, which may also cover private operators entrusted by the Member States with carrying out tasks in relation to those competences.

(23) According to the CJEU, the term **'national security**' refers to 'the primary interest in protecting the essential functions of the State and the fundamental interests of society and encompasses the prevention and punishment of activities capable of seriously destabilising the fundamental constitutional, political, economic or social structures of a country and, in particular, of directly threatening society, the population or the State itself, such as terrorist activities.'[17] National security does not cover, for example activities relating to road safety,[18] or the organisation or administration of justice.[19] As stated by the CJEU, 'it is for the Member States to define their essential security interests and to adopt appropriate measures to ensure their internal and external security, [….] a national measure […] taken for the purpose of protecting national security cannot render EU law inapplicable and exempt the Member States from their obligation to comply with that law.'[20]

(24) For the exclusion in Article 2(3), second subparagraph, AI Act to apply, the AI system must be placed on the market, put into service or used exclusively for military, defence or national security purposes. Recital 24 AI Act further clarifies how the notion

[17] Judgment of the Court of Justice of 6 October 2020, *La Quadrature du Net and Others*, C-511/18, C-512/18 and C-520/18,EU:C:2020:791, paragraph 135; Judgment of the Court of Justice of 5 June 2023, *Commission* v *Poland*, C-204/21, EU:C:2023:442, paragraph 318, referring to Case C-439/19, paragraph 67 and Case C-306/21, paragraph 40.
[18] Judgment of the Court of Justice of 22 June 2021, *Latvijas Republikas Saeima*, C-439/19, EU:C:2021:504, paragraph 68.
[19] Judgment of the Court of Justice of 5 June 2023, *Commission v Poland*, C-204/21, EU:C:2023:442, paragraph 319.
[20] Judgement of the Court of Justice of 6 October 2020, *Privacy International*, C-623/17, EU:C:2020:790, paragraph 44.

'**exclusively'** should be interpreted and when an AI system used for such purposes may nevertheless fall within the scope of the AI Act.

> For example, if an AI system placed on the market, put into service or used for military, defence or national security purposes is used (temporarily or permanently) for other purposes, such as for civilian or humanitarian purposes, law enforcement or public security purposes, that system will fall within the scope of the AI Act. In that case, the entity using the AI system for the other purposes should ensure compliance of the AI system with the AI Act, unless the system already complies with that act, which has to be verified before such use.

(25) Furthermore, recital 24 AI Act clarifies that AI systems placed on the market or put into service for an excluded purpose, namely military, defence or national security, and for one or more non-excluded purposes, such as civilian or law enforcement purposes (so called '**dual use'** systems), fall within the scope of the AI Act. Providers of those systems should ensure that they comply with the requirements in the AI Act.

> For example, if a company offers a RBI system for various purposes, including law enforcement and national security, that company is the provider of that 'dual use' system and must ensure its compliance with the requirements in the AI Act.

(26) However, the fact that an AI system may fall within the scope of the AI Act should not affect the ability of entities carrying out national security, defence and military activities to use that system for national security, military and defence purposes, regardless of the type of entity carrying out those activities[21].

> For example, if a national security agency or a private operator is tasked by a national intelligence agency to use real-time RBI systems for national security purposes (such as to gather intelligence), such use would be excluded from the scope of the AI Act.

(27) The clear delineation of the national security exclusion is particularly important where AI systems are placed on the market, put into service or used for law enforcement purposes that fall within the scope of the AI Act. This is relevant for the prohibitions regarding individual crime predictions and assessments and regarding the use of real-time RBI systems for law enforcement purposes laid down in Article 5(1)(d) and (h) AI Act respectively. Police and other law enforcement authorities are tasked with the prevention, detection, investigation and prosecution of criminal offences or the execution of criminal penalties, including safeguarding against and preventing threats to public security[22]. Whenever AI systems are used for such purposes, they will fall within the scope of the AI Act.

[21] Recital 24 AI Act.
[22] Article 3(46) AI Act.

(28) The activities of Europol and other Union security agencies, such as Frontex, fall within the scope of the AI Act.

2.5.2. Judicial and law enforcement cooperation with third countries

(29) According to Article 2(4) AI Act, the AI Act does not apply to public authorities in a third country or international organisations, 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, provided that such a third country or international organisation provides adequate safeguards with respect to the protection of fundamental rights and freedoms of individuals. Where relevant, this exclusion may cover the activities of private entities entrusted by the third country in question to carry out specific tasks in support of such law enforcement and judicial cooperation.[23] At the same time, for the exclusion to apply, these frameworks for cooperation or international agreements must include adequate safeguards with respect to the protection of the fundamental rights and freedoms of individuals, to be assessed by the market surveillance authorities competent for the supervision of AI systems used in the area of law enforcement and justice.[24] Recital 22 AI Act clarifies that the recipient national authorities and Union institutions, bodies, offices and agencies making use of such AI outputs in the Union remain accountable to ensure their use complies with Union law. When those international agreements are revised or new ones are concluded in the future, the contracting parties should make utmost efforts to align those agreements with the requirements of the AI Act.

2.5.3. Research & Development

(30) According to Article 2(8) AI Act, the AI Act does not apply 'to any research, testing or development activity regarding AI systems or AI models prior to their being placed on the market or put into service'. This exclusion is in line with the market-based logic of the AI Act, which applies to AI systems once they are placed on the market or put into service.

> For example, during the research and development (R&D) phase, AI developers have the freedom to experiment and test new functionalities which might involve techniques that could be seen as manipulative and covered by Article 5(1)(a) AI Act, if used in consumer-facing applications. The AI Act allows for such experimentation by recognising that early-stage R&D is essential for refining AI technologies and ensuring that they meet safety and ethical standards prior to their placing on the market.

(31) As clarified in recital 25 AI Act, the AI Act aims to support innovation and recognises the importance of scientific research in advancing AI technologies and contributing to scientific progress and innovation. Article 2(6) AI Act therefore provides an exclusion

[23] See Recital 22 AI Act.
[24] See Recital 22 and Article 74(8) AI Act.

for ‘AI systems or AI models, including their outputs, specifically developed and put into service for the sole purpose of scientific research and development’.

> For example, research into cognitive and behavioural responses to AI-driven subliminal or deceptive stimuli can provide valuable insights into human-AI interactions, informing safer and more effective AI applications in the future. Such research is permitted, since it is excluded from the scope of the AI Act, notwithstanding the prohibition in Article 5(1)(a) AI Act.

(32) The exclusion in Article 2(8) AIA Act is, however, without prejudice to the obligation to comply with the AI Act where an AI system is placed on the market or put into service as a result of such research and development activity.[25] Testing in real-world conditions within the meaning of the AI Act[26] is also not covered by that exclusion.

> For example, a municipality wishing to test facial recognition software using a RBI system in the streets during carnival recruits volunteers to be identified by the system in real-world conditions. Because real-world testing does not fall within the exclusion of Article 2(8) AI Act, the planned testing must be fully compliant with the requirements for RBI systems in the AI Act, unless the system is tested in an AI regulatory sandbox or in accordance with the special regime for testing in real world conditions outside the sandbox, as provided for in Articles 60 and 61 AI Act.[27]

(33) In any event, any research and development activity (including when excluded from the scope of the AI Act) should be carried out in accordance with recognised ethical and professional standards for scientific research and should be conducted in accordance with applicable Union law[28] (e.g., data protection law that remains applicable).

2.5.4. Personal non-professional activity

(34) Article 2(10) AI Act provides that the AI Act ‘does not apply to obligations of deployers who are natural persons using systems in the course of a purely personal non-professional activity’. The definition of deployer also excludes users engaged in such activities (see section 2.4. above). Any activity through which a natural person gains an economic benefit on a regular basis or is otherwise involved in a professional, business, trade, occupational or freelance activity should be considered as a ‘professional’ activity. The specification of ‘personal’ is a qualifier of non-professional, meaning that the person should act in both a personal and a non-professional capacity. The exclusion

[25] Recital 25 AI Act.
[26] According to Article 3(57) AI Act, ‘testing in real-world conditions’ means the temporary testing of an AI system for its intended purpose in real-world conditions outside 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. The AI Act provides a special regime for such testing in real-world conditions which does not qualify as placing the AI system on the market or putting it into service within the meaning of this Regulation, provided that all the conditions laid down in Articles 57 or 60 are fulfilled, including obtaining free and informed consent from the persons participating in the testing etc.; see Article 60 AI Act.
[27] The AI Act contains detailed and specific obligations for AI Regulatory Sandboxes and real-world testing. See Article 57 AI Act *et seq.*
[28] Recital 25 AI Act.

should therefore, for example, not encompass criminal activities since these should not be considered purely personal.

> For example, an individual using a facial recognition system at home (e.g., to control access and to monitor for safety the entrance to the home) would fall under the exclusion of Article 2(10) AI Act and, hence, would not be subject to the obligations for deployers under the AI Act, even in cases where it is required to transmit (parts of) the footage to law enforcement authorities.
>
> By contrast, a natural person using an AI system for professional activities such as freelancers, journalists, doctors, etc. would need to comply with the obligations for deployers of facial recognition systems under the AI Act. Any use by natural persons where they are acting on behalf or under the authority of a deployer acting in a professional capacity will also fall within the scope of the AI Act.
>
> Furthermore, criminal activities cannot be considered purely personal activities, even if no economic benefit is sought or attained. For other unlawful activities, such as non-compliance with consumer protection or data protection law and national administrative legislation, the exclusion in the AI Act applies, but the other relevant legal frameworks continue to apply).

(35) The exclusion in Article 2(10) AI Act applies only as regards the obligations of deployers when using the system for purely personal non-professional activities. The system as such remains within the scope of the AI Act as regards the obligations of providers placing the system on the market or putting it into service, other professional deployers, and other responsible actors, such as importers and distributors.

> For example, an emotion recognition system, if intended to be used by natural persons for purely personal non-professional activities, remains a high-risk AI system as classified in Article 6 AI Act and must be fully in compliance with the AI Act. At the same time the deployer that uses it for purely personal non-professional use (e.g. an autistic person) will not be covered by specific obligations for deployers under the AI Act and the use would be out of scope.

2.5.5. AI systems released under free and open source licences

(36) According to Article 2(12) AI Act, the AI Act does not apply to AI systems released under free and open-source licences[29], unless they are placed on the market or put into service as high-risk AI systems or as an AI system that falls under Article 5 (prohibited AI practices) or Article 50 AI Act (transparency obligations for certain AI systems). This means that providers of AI systems cannot benefit from this exclusion if the AI system they place on the market or put into service constitutes a prohibited practice under Articles 5 AI Act.

[29] Recital 102 AI Act describes that a release of software and data under free and open-source licence 'allows them to be openly shared and where users can freely access, use, modify and redistribute them or modified versions thereto'.

2.6. Interplay of the prohibitions with the requirements for high-risk AI systems

(37) The AI practices prohibited by Article 5 AI Act should be considered in relation to the AI systems classified as high-risk in accordance with Article 6 AI Act, in particular those listed in Annex III.[30] That is because the use of AI systems classified as high-risk may in some cases qualify as a prohibited practice in specific instances if all conditions under one or more of the prohibitions in Article 5 AI Act are fulfilled. Conversely, most AI systems that fall under an exception from a prohibition listed in Article 5 AI Act will qualify as high-risk.

> For example, emotion recognition systems, where they do not fulfil the conditions for the prohibition in Article 5(1)(f) AI Act, classify as high-risk AI systems according to Article 6(2) and Annex III, point (1)(c) AI Act. Similarly, certain AI-based scoring system, such as those used for credit-scoring or assessing risk in health and life insurance, will be considered high-risk AI systems where they do not fulfil the conditions for the prohibition listed in Article 5(1)(c) AI Act.[31] Another example are AI systems evaluating persons and determining if they are entitled to receive essential public assistance benefits and services, such as healthcare services and social security benefits that are classified as high-risk.[32] If such systems involve unacceptable social scoring and fulfil the conditions of Article 5(1)(c) AI Act, their placing on the market, putting into service and use will be prohibited in the Union.
>
> In such cases, the risk assessment and management done by the provider and the compliance with the other requirements for high-risk AI systems (e.g. data governance, transparency and human oversight), as well as the deployer's obligations for appropriate use in accordance with the instructions of use and human oversight (Article 26) and in some cases a fundamental rights impact assessment (Article 27), should help to ensure that the high-risk AI system placed on the market or deployed is lawful and does not constitute a prohibited practice.

(38) Finally, AI systems that are exceptionally not considered high-risk based on Article 6(3) AI Act, despite falling under a high-risk use case of Annex III, may still fall within the scope of the prohibitions of Article 5 AI Act. Article 6(3) AI Act only results in an AI system being considered non-high-risk; it does not exclude such AI systems from the scope of the AI Act and the prohibitions.

2.7. Application of the prohibitions to general-purpose AI systems and systems with intended purposes

[30] In this list, AI systems based on biometrics are covered, as well as AI systems used for specific purposes in certain domains such as employment, education, access to public and private services, law enforcement etc.
[31] This is expressly mentioned in Recital 58 and in Annex III AI Act.
[32] Recital 58 AI Act.

(39) The prohibitions apply to any AI system, whether with an 'intended purpose'[33] or 'general-purpose' (i.e. that can serve a variety of purposes), for direct use or for integration in other AI systems.[34] Accordingly, each operator should take measures for which they are best placed based on their role and control over the system in the value chain to ensure a responsible and safe provision and use of AI systems, balancing their risks and benefits with a view to achieving the twin objectives of the AI Act.

(40) Deployers are thus expected not to use any AI system in a manner prohibited under Article 5 AI Act, including not to bypass any safety guardrails implemented by the providers of the system. While the harm often arises from the way the AI systems are used in practice, providers also have a responsibility not to place on the market or put into service AI systems, including general-purpose AI systems, that are reasonably likely to behave or be directly used in a manner prohibited by Article 5 AI Act.[35] In this context, providers are also expected to take effective and verifiable measures to build in safeguards and prevent and mitigate such harmful behaviour and misuse to the extent they are reasonably foreseeable and the measures are feasible and proportionate depending on the specific AI system and circumstances of the case. In their contractual relationships with deployers (i.e., in the terms of use of the AI system), providers are also expected to exclude use of their AI system for prohibited practices and provide appropriate information in the instructions of use for deployers and regarding the necessary human oversight.

> For example, a general-purpose AI system used as a chatbot may deploy manipulative and deceptive techniques which is likely to cause significant harm. To prevent prohibited behaviour of the AI system and uses that are reasonably likely to manipulate, deceive and cause significant harms under Article 5(1)(a) AI Act, the provider is expected to take appropriate and proportionate measures (e.g., appropriate safe and ethical design, integration of technical and other safeguards, restrictions of use, transparency and user control, appropriate information in the instructions of use) before the AI system is placed on the market (Article 5(1)(a) AI Act) to ensure the chatbot does not cause significant harm to users or other persons or groups of persons (see also section 3.2.3.c)).

(41) In certain cases, in particular where the prohibitions are linked to a very specific purpose of the system[36], providers may have limited possibilities to integrate other preventive and mitigating measures and will have to rely on primarily providing appropriate instructions and information to the deployers and the required human oversight and restricting prohibited use of the system. Where appropriate, such measures may also

[33] Defined in Article 3(12) AI Act as 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.
[34] See Article 3(66) AI Act.
[35] This follows in particular from the reference to 'placing on the market' or 'putting into service' in all prohibitions listed in Article 5 AI Act, with the exception of the prohibition of real-time RBI systems in Article 5(1)(h) that applies only to the use.
[36] Article 5(1)(d)-(h) AI Act.

include monitoring for compliance with that restriction, depending on the means through which the AI system is supplied and the information at the provider's disposal for possible misuse. Any possible monitoring measures to detect misuse should not amount to a general monitoring of the activities of the deployers and should be in line with Union law.

For example, a general-purpose AI system that can recognise or infer emotions should not be used by deployers at the workplaces or in education institutions, unless an exception for medical or safety reasons applies. However, the provider may not be in a position to know the specific context in which the emotion recognition functionality of the system will be used and whether an exception to the prohibition in Article 5(1)(f) AI Act may apply. Such providers may nevertheless explicitly exclude such prohibited use in their terms of use and include appropriate information in the instructions of use to guide deployers. They are also expected to take appropriate measures if they become aware that the system is misused for this specific prohibited purpose by specific deployers, for example, if such misuse is reported or the provider becomes otherwise aware, which may be the case if the system is directly operated through a platform under the control of the provider and the provider performs checks.

2.8. Interplay between the prohibitions and other Union law

(42) The AI Act is a regulation that applies horizontally across all sectors without prejudice to other Union legislation, in particular on the protection of fundamental rights, consumer protection, employment, the protection of workers, and product safety[37]. The AI Act complements such legislation through its preventative and safety logic (AI systems may not be placed on the market or used in a certain way) and provides additional protection by addressing specific harmful AI practices which may not be prohibited by other laws. Furthermore, by addressing the earlier stages of the AI systems' lifecycle (i.e. the placing on the market and putting into service) and deployment (i.e. the use), the AI Act's prohibitions enable action to be taken against harmful practices involving AI at various points in the AI value chain.

(43) At the same time, the AI Act does not affect prohibitions that apply where an AI practice falls within other Union law[38]. Thus, even where an AI system is not prohibited by the AI Act, its use could still be prohibited or unlawful based on other primary or secondary Union law (e.g., because of the failure to respect fundamental rights in a given case, such as the lack of a legal basis for the processing of personal data required under data protection law, discrimination prohibited by Union law, etc.). The respect of the prohibitions in the AI Act are therefore not a sufficient condition for compliance with other Union legislation that remains applicable to providers and deployers of AI systems.

[37] Article 2 and Recital 9 AI Act.
[38] Article 5(8) AI Act.

For example, AI-enabled emotion recognition systems used in the workplace that are exempted from the prohibition in Article 5(1)(f) AI Act, because they are used for medical or safety reasons, remain subject to data protection law and Union and national law on employment and working conditions, including health and safety at work, which may foresee other restrictions and safeguards in relation to the use of such systems.[39]

(44) When specific activities related to the placing on the market or use of AI systems are also covered under other Union legislation, the AI Act aims to ensure the consistent implementation of the different provisions. Moreover, it enables effective cooperation between the competent authorities responsible for the enforcement of the AI Act and the authorities protecting fundamental rights pursuant to Article 77 AI Act and other provisions of the AI Act. More generally, in accordance with Article 4(3) TEU, the various authorities concerned are bound to cooperate sincerely when giving effect to their respective tasks under Union law.

(45) In the context of the prohibitions, the interplay between the AI Act and Union data protection law is particularly relevant, since AI systems often process information relating to identified or identifiable natural persons ('personal data').[40] Depending on the prohibition and the context, the most relevant legal acts in relation to such systems are Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation, hereinafter 'GDPR'), Directive (EU) 2016/680 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data (Law Enforcement Directive, hereinafter 'LED'), and Regulation (EU) 2018/1725 which lays down data protection rules for the EU Institutions, bodies, offices and agencies (hereinafter 'EUDPR'). In accordance with Article 2(7) AI Act, these acts remain unaffected and will continue to apply alongside the AI Act, which is consistent and complementary to the EU data protection acquis. Several aspects of these EU data protection rules have been clarified by the CJEU and the European Data Protection Board has adopted a series of guidelines (e.g., on the notion of 'profiling'[41], which is particularly relevant for the prohibition in Article 5(1)(d) AI Act, since it uses the same notion).

(46) Concerning the prohibitions/restrictions on the use of biometric categorisation systems and real-time RBI systems for law enforcement purposes, the AI Act applies as *lex specialis* to Article 10 LED, thus regulating such use and the processing of biometric data involved in an exhaustive manner[42]. In that context, the AI Act is not intended to

[39] See also Recital 9 AI Act.
[40] Article 2(7) AI Act; see also Recital 10 AI Act.
[41] See also Article 29 Data Protection Working Party, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, WP251rev.01, 6.2.2018, and endorsed by the EDPB.
[42] Recital 38 AI Act.

provide the legal basis for the processing of personal data under Article 8 of Directive (EU) 2016/680. All other provisions of that Directive apply in addition to the conditions set out in the AI Act, in particular for the use of real-time (RBI) systems for law enforcement purposes when permitted, subject to the limited exceptions in Article 5(1)(h) AI Act. More generally, the LED must also be complied with for any processing of personal data by competent law enforcement authorities (i.e., competent authorities under Article 3(7) LED) when they process the data for law enforcement purposes.

(47) In accordance with Article 2(9) AI Act, EU consumer protection and safety legislation also remain fully applicable to AI systems falling within scope of those acts.

For example,

- Social scoring practices by traders (including natural persons acting in a professional capacity in business-to-consumer relations), subject to case-by-case assessment, may also be considered 'unfair' and therefore in breach of consumer law (i.e. Directive 2005/29/EC);

- The use of an AI system to infer emotions may also have to comply with Regulation (EU) 2017/745 (Medical Device Regulation) if the AI system is used for medical diagnosis or medical treatment purposes.

(48) Furthermore, the AI Act applies in conjunction with relevant obligations for providers of intermediary services that embed AI systems or models into their services regulated by Regulation (EU) 2022/2065 ("the Digital Services Act"). Specifically, Article 2(5) AI Act indicates that the AI Act does not affect the application of the provisions on the liability of such providers as set out in Chapter II of the Digital Services Act.

(49) In addition, the prohibitions in the AI Act are without prejudice to any liability that the provider or deployer might incur for the harm caused according to applicable Union or national liability laws.[43]

(50) Finally, the prohibitions in Article 5 AI Act and the explicit exceptions to those prohibitions may not be used to circumvent or as a justification to infringe obligations under other Union legislation.

(51) As secondary Union legislation, the AI Act must be interpreted in the light of the fundamental rights and freedoms guaranteed by the EU Treaties and the Charter, as well as those protected by international conventions to which the Union is a party.[44]

[43] The conditions for liability (related to damage, liable person, fault or burden of proof, etc) will be determined by the applicable law, such as Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on liability for defective products, (Text with EEA relevance), OJ L, 2024/2853, 18.11.2024 or the applicable national liability laws (see also Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive) COM/2022/496 final).

[44] Even if the Union is not yet a party to the European Convention for the Protection of Human Rights and Fundamental Freedoms, Article 59(3) of the Charter states that in so far as the Charter contains rights which correspond to rights guaranteed by the European Convention for the Protection of Human Rights and Fundamental Freedoms, the meaning and scope of those rights shall be the same as those laid down by the said Convention. This provision shall not prevent Union law providing more extensive protection.

(52) Additional clarifications on the interplay of specific prohibitions with other Union law are provided under the relevant sections below.

2.9. Enforcement of Article 5 AI Act

2.9.1. Market Surveillance Authorities

(53) Market surveillance authorities designated by the Member States as well as the European Data Protection Supervisor (as the market surveillance authority for the EU institutions, agencies and bodies) are responsible for the enforcement of the rules in the AI Act for AI systems, including the prohibitions. Such enforcement takes place within the system of market surveillance and compliance of products established by Regulation (EU) 2019/1020[45], in line with other Union product safety legislation. The enforcement powers of market surveillance authorities in relation to AI systems are laid down in the AI Act and in Regulation (EU) 2019/1020. Those authorities can take enforcement actions in relation to the prohibitions on their own initiative or following a complaint, which every affected person or any other natural or legal person having grounds to consider such violations has the right to lodge[46]. Member States must designate their competent market surveillance authorities by 2 August 2025.

(54) The procedure in the AI Act to deal with AI systems presenting a risk at national level is particularly relevant in the context of enforcing the prohibitions[47]. Where there are cross-border implications beyond the territory of the market surveillance authority, the authority of the Member State concerned must inform the Commission and the market surveillance authorities of other Member States. All market surveillance authorities should follow a Union safeguard procedure with a decision taken by the Commission[48] determining whether the AI system constitutes a prohibited practice. That procedure aims to ensure that the prohibitions are applied uniformly across all Member States, so as to provide legal certainty to both providers and deployers of AI systems. To ensure the uniform application of the AI Act, national market surveillance authorities should also strive for a harmonized application of the prohibitions for comparable cases that do not cross the Member State's territory by drawing inspiration from these Guidelines and cooperating within the AI Board[49].

2.9.2. Penalties

(55) The AI Act follows a tiered approach in setting the penalties for non-compliance with its various provisions, depending on the seriousness of the infringement. Non-compliance with the prohibitions in Article 5 AI Act are considered to constitute the most severe infringement and they are therefore subject to the highest fine. Providers and deployers engaging in prohibited AI practices may be fined up to EUR 35 000 000

[45] See also Recital 156 AI Act.
[46] Article 85 AI Act.
[47] Article 79 AI Act.
[48] Article 81 AI Act.
[49] Article 65 and 66 AI Act.

or, if the offender is an undertaking, up to 7 % of its total worldwide annual turnover for the preceding financial year, whichever is higher.[50] Each Member State should lay down rules if and to the extent that administrative fines may be imposed on public authorities and bodies established in that Member State as providers and deployers of AI systems. EU institutions, bodies and agencies that violate the prohibitions may be subject to administrative fines of up to EUR 1 500 000.[51]

(56) It is possible that one and the same prohibited conduct constitutes a violation of two or more provisions of the AI Act (i.e. the non-labelling of deep fakes may also constitute a deceptive technique under Article 5(1)(a) AI Act). In such cases, the principle of ne bis in idem should be respected. In any event, the criteria for determining the penalty as provided for in Article 99(7) AI Act must be taken into account.

(57) Since violations of the prohibitions in Article 5 AI Act interfere the most with the freedoms of others and give rise to the highest fines, their scope should be interpreted narrowly.

3. ARTICLE 5(1)(A) AND (B) AI ACT – HARMFUL MANIPULATION, DECEPTION AND EXPLOITATION

(58) The first two prohibitions in Article 5(1)(a) and (b) AI Act aim to safeguard individuals and vulnerable persons from the significantly harmful effects of AI-enabled manipulation and exploitation. Those prohibitions target AI systems that deploy subliminal, purposefully manipulative or deceptive techniques that are significantly harmful and materially influence the behaviour of natural persons or group(s) of persons (Article 5(1)(a) AI Act) or exploit vulnerabilities due to age, disability, or a specific socio-economic situation (Article 5(1)(b) AI Act).

3.1. Rationale and objectives

(59) The underlying rationale of these prohibitions is to protect individual autonomy and well-being from manipulative, deceptive, and exploitative AI practices that can subvert and impair an individual's autonomy, decision-making, and free choices.[52] The prohibitions aim to protect the right to human dignity (Article 1 of the Charter), which also constitutes the basis of all fundamental rights and includes individual autonomy as an essential aspect. In particular, the prohibitions aim to prevent manipulation and exploitation through AI systems that reduce individuals to mere tools for achieving certain ends and to safeguard those that are most vulnerable and susceptible to harmful manipulation and exploitation. The prohibitions against significantly harmful manipulative, deceptive and exploitative AI practices fully align with the broader objectives of the AI Act to promote trustworthy and human-centric AI systems that are safe, transparent, fair and serve humanity and align with human agency and EU values.

[50] Article 99 AI Act.
[51] Article 100 AI Act.
[52] Recital 29 AI Act.

3.2. Main components of the prohibition in Article 5(1)(a) AI Act – harmful manipulation

Article 5(1)(a) AI Act provides:

1. The following AI 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, or the effect of materially distorting the behaviour of a person or a group of persons by appreciably impairing their ability to make an informed decision, thereby causing them to take a decision that they would not have otherwise taken in a manner that causes or is reasonably likely to cause that person, another person or group of persons significant harm;

(60) Several cumulative conditions must be fulfilled for the prohibition in Article 5(1)(a) AI Act to apply:

(i) The practice must constitute the 'placing on the market', the 'putting into service', or the 'use' of an AI system.

(ii) The AI system must deploy subliminal (beyond a person's consciousness), purposefully manipulative or deceptive techniques.

(iii) The techniques deployed by the AI system should have the objective or the effect of materially distorting the behaviour of a person or a group of persons. The distortion must appreciably impair their ability to make an informed decision, resulting in a decision that the person or the group of persons would not have otherwise made.

(iv) The distorted behaviour must cause or be reasonably likely to cause significant harm to that person, another person, or a group of persons.

(61) For the prohibition to apply, all four conditions must be simultaneously fulfilled and there must be a plausible causal link between the techniques deployed, the material distortion of the behaviour of the person, and the significant harm that has resulted or is reasonably likely to result from that behaviour.

(62) The first condition, i.e. the 'placing on the market', the 'putting into service' or the 'use' of an AI system, has already been analysed. The prohibition, therefore, applies to both providers and deployers of AI systems, each within their respective responsibilities, not to place on the market, put into service or use such systems. The next sections focus on the other three conditions.

3.2.1. Subliminal, purposefully manipulative or deceptive techniques

(63) Article 5(1)(a) AI Act prohibits three alternative types of manipulative techniques: a) subliminal techniques beyond a person's consciousness: b) purposefully manipulative techniques; and c) deceptive techniques. To fall within scope of Article 5(1)(a) AI Act, an AI system must deploy one or more of these techniques.

a) Subliminal techniques

(64) While the AI Act does not define 'subliminal techniques', Article 5(1)(a) AI Act specifies that subliminal techniques operate beyond (below or above) the threshold of conscious awareness. Because subliminal techniques and the way they operate are inherently covert, such techniques bypass a person's rational defences against manipulation and are capable of influencing decisions without the conscious awareness of the person, raising significant ethical concerns and impairing individual autonomy, agency and free choice[53].

(65) The subliminal techniques must be capable of influencing behaviour in ways in which the person remains unaware of such influence, how it works, or its effects on the person's decision-making or value- and opinion-formation. In particular, subliminal techniques may use stimuli delivered through audio, visual, or tactile media that are too brief or subtle to be noticed and that have been traditionally known and prohibited in other sectors, such as media advertising.[54] These stimuli, while not consciously perceived, may still be processed by the brain and influence behaviour.

Examples of subliminal techniques (not necessarily prohibited unless all other conditions listed in Article 5(1)(a) AI Act are fulfilled) include:

- **Visual Subliminal Messages:** an AI system may show or embed images or text flashed briefly during video playback which are technically visible, but flashed too quickly for the conscious mind to register, while still being capable of influencing attitudes or behaviours.

- **Auditory Subliminal Messages:** an AI system may deploy sounds or verbal messages at low volumes or masked by other sounds, influencing the listener without conscious awareness. These sounds are still technically within the range of hearing, but are not consciously noticed by the listener due to their subtlety or masking by other audio. - Tactile Subliminal Stimuli: an AI system may stimulate subtle physical sensations that are perceived unconsciously, capable of influencing emotional states or behaviour.

- **Subvisual and Subaudible Cueing:** an AI system may deploy stimuli that are not just subtle or masked, but are presented in a way that makes them entirely undetectable by the human senses under normal conditions, for example flashing visual stimuli (e.g. flashing images) too quickly for the human eye to detect consciously or playing sounds at volumes imperceptible to the human ear.

- **Embedded Images:** an AI system may hide images within other visual content which are not consciously perceived, but may still be processed by the brain and influence behaviour.

[53] Recital 29 AI Act.

[54] See in particular, the Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (OJ L 95, 15.4.2010, p. 1) ('AVMSD'). which strictly prohibits subliminal techniques in audiovisual commercial communications.

- **Misdirection:** an AI system may draw attention to specific stimuli or content to prevent noticing other content, often by exploiting cognitive biases and vulnerabilities in attention.

-**Temporal manipulation:** an AI system may alter the perception of time in user interactions, thus influencing their behaviour and causing impatience and dependence.

(66) The rapid development of AI and related technologies, such as big data analytics, neuro technologies, brain-computer interfaces and virtual reality, heightens the risk of sophisticated subliminal manipulation and its capability to effectively influence human behaviour in a subconscious manner.[55] AI can also extend to emerging machine-brain interfaces and advanced techniques like dream-hacking and brain spyware.

For example, a game can leverage AI-enabled neuro technologies and machine-brain interfaces that permit users to control (parts of) a game with headgear that detects brain activity. AI may be used to train the user's brain surreptitiously and without their awareness to reveal or infer from the neural data information that can be very intrusive and sensitive (e.g. personal bank information, intimate information, etc.) in a manner that can cause them significant harm. The prohibition in Article 5(1)(a) AI Act targets only cases of such significantly harmful subliminal manipulation and not machine-brain interface applications in general when designed in a safe and secure manner and respectful of privacy and individual autonomy.

b) Purposefully manipulative techniques

(67) 'Purposefully manipulative techniques' are not defined in the AI Act, but they should be understood as techniques that are designed or objectively aim to influence, alter, or control an individual's behaviour in a manner that undermines their individual autonomy and free choices. Manipulative techniques are typically designed to exploit cognitive biases, psychological vulnerabilities, or situational factors that make individuals more susceptible to influence. Because of their adaptability, AI systems are also able to respond well to a person's individual circumstances or vulnerabilities and increase the effectiveness and impact of manipulation at scale. While the manipulative capability is an important element to determine the nature of the technique, it is not necessary that the provider or deployer or the system itself deploying the manipulative techniques also intends to cause harm[56].

(68) While not all manipulative techniques operate beyond the threshold of conscious awareness, many do and there may be an overlap with subliminal techniques, since such techniques also ultimately have manipulative effects. Recital 29 AI Act clarifies that the prohibition in Article 5(1)(a) also covers techniques where individuals, even if they are aware of the influence attempt, may not be able to control or resist its manipulative

[55] See Recital 29 AI Act.

[56] See in this context Recital 28 and sections 3.2.2. and 3.2.3. of the Guidelines.

effect[57]. As a result, individuals are influenced or pushed into behaviour and decisions they would normally not have made if they were not subjected to the manipulative techniques to a point that undermines their individual autonomy or free choice.

> An example of purposefully manipulative techniques is sensory manipulation where an AI system deploys background audio or images that lead to mood alterations, for example increasing anxiety and mental distress that influence users' behaviour to the point of creating significant harm.
>
> Another example is personalised manipulation where an AI system that creates and tailors highly persuasive messages based on an individual's personal data or exploits other individual vulnerabilities influences their behaviour or choices to a point of creating significant harm.

(69) The prohibition against purposefully manipulative techniques also covers AI systems that manipulate individuals without any human intending them to do so. Article 5(1)(a) AI Act prohibits AI systems that deploy certain techniques or exhibit a specific manipulative behaviour. Therefore, it could also be the AI system that deploys such manipulative techniques, rather than the provider or the deployer that has designed or used the system in this way.

> For example, regardless of whether the provider intends it, an AI system may learn manipulative techniques because the data on which it is trained contain many instances of manipulative techniques,[58] or because reinforcement learning from human feedback can be 'gamed' through manipulative techniques.[59]
>
> By contrast, if the manipulative behaviour of the system is merely incidental, the system should not be considered deploying purposefully manipulative techniques as long as the provider has taken appropriate preventive and mitigating measures in case significant harms are reasonably likely to occur (see section 3.2.3.c) below).

c) Deceptive techniques

(70) The AI Act does not define 'deceptive techniques'. Recital 29 AI Act clarifies that these are techniques that subvert or impair a person's autonomy, decision-making or free choice in ways that the person is not consciously aware or, where it is aware, can still be deceived or is not able to control or resist them. 'Deceptive techniques' deployed by AI systems should be understood to involve presenting false or misleading information with the objective or the effect of deceiving individuals and influencing their behaviour in a manner that undermines their autonomy, decision-making and free choices.

---

[57] Recital 28 AI Act.

[58] M. Carroll et al., Characterising Manipulation from AI Systems, In Equity and Access in Algorithms,Mechanisms, and Optimization (EAAMO '23), October 30-November 1, 2023, Boston, MA, USA. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3617694.3623226 |:2303.09387.

[59] D. Amodei, et al., Concrete Problems in AI Safety, 36th Conference on Neural Information Processing Systems (NeurIPS 2022). arXiv:1606.06565; J. Skalse et al. Defining and Characterizing Reward Gaming, Advances in Neural Information Processing Systems 35 (NeurIPS 2022) C. Denison et al., Sycophancy to Subterfuge: Investigating Reward-Tampering in Large Language, 36th Conference on Neural Information Processing Systems (NeurIPS 2022), Models, arXiv:2406.10162.

(71) In this context, the interplay between the prohibition in Article 5(1)(a) AI Act and the deployer's obligations in Article 50(4) AI Act to label 'deep fakes' and certain AI-generated text publications on matters of public interest[60], as well as the provider's obligation to ensure AI systems interacting with people are designed in a way to inform people that they are interacting with AI and not a human[61], should be clarified. Such visible disclosure constitutes a mitigating measure that should also be enabled through design features embedded in the AI system provided by the provider, including technical measures enabling the detection of AI-generated and manipulated content[62]. The visible labelling of 'deep fakes' and chatbots reduces the risk of deception that is likely to arise once the AI-generated content is disseminated to the public and reduces the risk of harmful distorting effects on the individual's opinion- and belief-formation and behaviour.

(72) By contrast, the prohibition in Article 5(1)(a) AI Act has a much more limited scope. It may, for example, cover cases where a chatbot or deceptive AI-generated content presents false or misleading information in ways that aim to or have the effect of deceiving individuals and distorting their behaviour that would not have happened if they were not exposed to the interaction with the AI system or the deceptive AI generated content, in particular if this has not been visibly disclosed[63].

(73) As with purposefully manipulative techniques, the prohibition of deceptive techniques may also cover AI systems that deceive individuals without any human intending them to do so (see section 3.2.1.b)above). For example, regardless of whether their providers intend such an outcome, AI systems may learn deceptive techniques simply because this increases their performance for the task for which they were developed, for example by reinforced learning[64].

> An example of deceptive techniques that may be deployed by AI is an AI chatbot that impersonates a friend of a person or a relative with synthetic voice and tries to pretend it is the person causing scams and significant harms.
>
> Another example is an AI system that learns to identify when it is under evaluation and temporarily halts any undesired behaviour, only to resume such behaviour once the evaluation period is over.[65] Such deceptive behaviour is particularly dangerous, since it

[60] Article 50(4) AI Act.
[61] Article 50(1) AI Act.
[62] Article 50(2) AI Act.
[63] While in principle the transparency obligations in Article 50 AI Act aim to minimise the manipulative effects of deep fakes and chatbots, there might be instances and contexts where despite the information notices these deceptive techniques may still have significant effects on individuals and distort their behaviour to a point that undermine persons' individual autonomy and informed decision-making, so they should not be misused for disinformation and manipulation purposes and might still be covered in some cases by the prohibition in Article 5(1)a) if all other conditions of the ban are fulfilled (including the significant harms).
[64] F. Ward, F. Toni, F. Belardinelli, T. Everitt, Honesty Is the Best Policy: Defining and Mitigating AI Deception (neurips.cc); Advances in Neural Information Processing Systems 36 (NeurIPS 2023); P. Park. et al. AI deception: A survey of examples, risks, and potential solutions [2406.10162] Patterns, Volume 5, Issue 5, 100988.
[65] J. Lehman, J. Clune, D. Misevic, C. Adami, L. Altenberg, J. Beaulieu, et al. The surprising creativity of digital evolution: A collection of anecdotes from the evolutionary computation and artificial life research communities. Artificial life, 26(2):274–306, 2020.

defies any external human oversight over the system and may be prohibited if it is reasonably likely to cause significant harms.

By contrast, a generative AI system that incidentally presents false or misleading information and hallucinates[66] may not be considered to deploy deceptive techniques within the meaning of Article 5(1)(a) AI Act, taking into account the limitations and the state of the art of generative AI. In particular, this may be the case where the provider of the system has properly informed users about the system's limitations and integrated appropriate safeguards into the system to minimise such outcomes and provided that the system is not intended for, nor deployed in, sensitive contexts (e.g. health, education, elections) where serious harmful consequences are likely to occur (see also considerations in section 3.2.3.c) below).

d) Combination of techniques

(74) Article 5(1)(a) AI Act applies to subliminal, purposefully manipulative, or deceptive techniques, or to combinations of such techniques that can have a compound impact. As stated above, purposefully manipulative techniques may be also subliminal in nature, if they operate beyond the threshold of conscious awareness.

(75) Furthermore, when purposefully manipulative and deceptive techniques are applied in combination, this may significantly influence the behaviour of individuals, leading them to make decisions based on unconscious manipulations and false beliefs. This combination may create a feedback loop where individuals are less likely to question or critically evaluate the information received, since the manipulative elements have already primed their cognitive biases and emotional responses.

3.2.2. With the objective or the effect of materially distorting the behaviour of a person or a group of persons

(76) A third condition for the prohibition in Article 5(1)(a) AI Act to apply is that the deployed subliminal, purposefully manipulative, or deceptive technique must have 'the objective, or the effect of materially distorting the behaviour of a person or a group of persons'. This implies a substantial impact on the behaviour where a person's autonomy and free choices are undermined, rather than a minor influence. However, intent is not a necessary requirement, since Article 5(1)(a) AI Act also covers practices that may only have the 'effect' of causing material distortion. There should be a plausible/reasonably likely causal link between the potential material distortion of the behaviour and the subliminal, purposefully manipulative or deceptive technique deployed by the AI system.

[66] 'Hallucination' is a term used to describe a technical flaw in generative AI systems when they generate unwanted information that is fabricated or factually incorrect without this being intended by their developers. See more Ji Ziwei et al., Survey of Hallucination in Natural Language Generation | ACM Computing Surveys, 55, Issue 12, Article No.: 248, Pages 1 – 38.

a) The concept of 'material distortion of the behaviour'

(77) The concept of 'material distortion of the behaviour' of a person or a group of persons is central to Article 5(1)(a) AI Act. It involves the deployment of subliminal, purposefully manipulative or deceptive techniques that are capable of influencing people's behaviour in a manner that appreciably impairs their ability to make an informed decision, thereby causing them to behave in a way or to take a decision that they would otherwise not have taken.

(78) 'Appreciable impairment' refers to a substantially reduced ability to make informed and autonomous decisions, thereby causing individuals to behave in a way or to take a decision that they would otherwise not have taken. It goes beyond minor or negligible impacts and involves a significant distortion or hindrance in decision-making and free choice, including in relation to opinion- and belief-formation. This suggests that 'material distortion' involves a degree of coercion, manipulation, or deception that goes beyond lawful persuasion, which falls outside the scope of the prohibition (see section 3.5.1. below).

(79) An 'informed decision' requires an understanding and knowledge of the relevant information, including the available options, the risks and benefits of each choice, the possible effects of the AI system on their behaviour, and, as appropriate, other contextual information that is important for the decision-making or the behaviour of the person.

(80) For the interpretation of the concept of 'material distortion of behaviour', Union consumer protection law, in particular, Directive 2005/29/EC (Unfair Commercial Practices Directive or 'UCPD'), may constitute a valid source of inspiration. The UCPD prohibits various unfair, misleading, and aggressive commercial practices (Articles 5 to 9 UCPD) capable of causing consumers to make transactional decisions that they would otherwise not have made. According to the CJEU and the Commission guidance on the UCPD[67], there is no need to prove that a consumer's economic behaviour has been distorted, it suffices to establish that a commercial practice is 'likely' (i.e., capable) of impacting an average consumer's transactional decision.[68] The CJEU has also underscored that even accurate information may be misleading if presented in a way that distorts the consumer's decision-making process.[69] National enforcement authorities are tasked to investigate the specific facts and circumstances of each case (*in concreto)* and to evaluate the potential impact of the practice on the average consumer's decision-making process (*in abstract*).[70] For that purpose, they must take

[67] See also Commission Guidance on the interpretation and application of Directive 2005/29/EC of the European Parliament and of the Council concerning unfair business-to-consumer commercial practices in the internal market, (OJ C 526, 29.12.2021, p. 1)

[68] Judgment of the Court of Justice of Judgment of the Court (Fifth Chamber) of 26 October 2016. *Canal Digital Danmark A/S*. EU:C:2016:800 , Case C-611/14, para 73.

[69] Judgment of the Court of Justice of 19 December 2013, *Trento Sviluppo and Centrale Adriatica*, C-281/12, EU:C:2013:859.

[70] Commission Notice – Guidance on the interpretation and application of Directive 2005/29/EC of the European Parliament and of the Council concerning unfair business-to-consumer commercial practices in the internal market (OJ C 526, 29.12.2021, p. 1).

the point of view of the 'average' consumer, which is the benchmark developed by the CJEU, now integrated in the UCPD[71].

(81) In the context of the prohibition in Article 5(1)(a) AI Act, market surveillance authorities must also investigate each case's specific facts and circumstances, assessing whether the subliminal, purposefully manipulative or deceptive technique deployed by the AI system is likely to appreciably impair the decision-making, individual autonomy and free choice of an 'average' individual within a targeted group when the system affects a group of persons in a manner that is reasonably likely to cause significant harm. Such an interpretation seems justified given that the AI Act intends to complement the UCPD[72] and must be applied in a consistent manner. At the same time, given that Article 5(1)(a) AI Act also refers to the possibility to distort the 'behaviour of a natural person' and, if the perspective of the 'average' individual proves difficult or ineffective to assess in certain contexts (for example due to very tailored or 'personalised' manipulation or harmful effects on specific vulnerable groups), specific cases may be examined also from the perspective of specific individuals by assessing to what extent an AI system deploying subliminal, purposefully manipulate or deceptive techniques is capable of undermining their individual autonomy in concrete cases and significant harm has occurred or is likely to occur.

b) Scenario 1: Prohibited AI systems 'with the objective to' materially distort behaviour

(82) Article 5(1)(a) AI Act applies to AI systems deploying the above-mentioned techniques and having as a first scenario 'the objective to materially distort the behaviour of a person or a group of persons'. Such an objective may be pursued by the provider or the deployer of the AI system, or by the system itself within the implicit objectives it may pursue[73]. This objective should be distinguished from the 'intended purpose' of the AI system (Article 3(12) AI Act). Even if intended by the provider, the manipulative objective is in most cases not the purpose of the use for which the system is offered and it is often neither transparent, nor specified as such in the information supplied by the provider (e.g. in the instructions for use, promotional or sales materials and statements, as well as in the technical documentation).

For example, a chatbot that may be used in different contexts is designed to use subliminal messaging techniques, such as flashing brief visual cues and embedding

[71] See Recitals 18 and 19 UCPD. 'Average consumer' is a person who is reasonably well informed and reasonably observant and circumspect, considering social, cultural and linguistic factors. The average consumer test is not a statistical test (i.e., it does not require to prove that a certain percentage of consumers would have been materially distorted/appreciably impaired by a business practice). The test is based on the principle of proportionality. The UCPD adopted this notion to strike the right balance between the need to protect consumers and the promotion of free trade in an openly competitive market. Courts and authorities will have to exercise their own faculty of judgment to determine the typical reaction of the average consumer in a given case. In the UCPD Guidance, the Commission advised them to make use of behavioural insights and other data. Case C-646/22, *Compass Banca*, clarifies that the definition of the average consumer does not exclude the possibility that an individual's decision-making capacity may be impaired by constraints, such as cognitive biases. Judgment of the Court (Fifth Chamber) of 14 November 2024.*Compass Banca SpA v Autorità Garante della Concorrenza e del Mercato* (AGCM), Case C-646/22, EU:C:2024:957.

[72] Recital 29 AI Act.

[73] See Article 3(1) AI Act which states that the AI system may pursue implicit or explicit objectives when performing its functions which may include as well implicit manipulative or deceptive objectives even if the system has not been explicitly programmed in this way.

inaudible auditory signals or to exploit emotional dependency or specific vulnerabilities of users in advertisements. These techniques are deployed 'with the objective to' materially distort users' behaviour, since they are objectively a design feature that aims to influence consumers' purchasing decisions without their conscious awareness, to push people to take significantly harmful financial decisions.

Deploying AI systems to impersonate other persons could also be considered as an AI system deployed 'with the objective' to deceive and materially distort the behaviour of persons if the person is effectively deceived, thus substantially affecting their ability to make informed decisions about the identity of the person.

If in both examples the other conditions in Article 5(1)(a) AI Act are fulfilled, in particular regarding the significant harm, those systems are likely to fall within the scope of the prohibition, but this will require a case-by-case assessment.

c) Scenario 2: Prohibited AI systems 'with the effect of' materially distorting behaviour

(83) The intent of the provider or deployer to materially distort the behaviour of a person or group of persons is a sufficient, but not a necessary, condition for the prohibition in Article 5(1)(a) AI Act to apply. That prohibition also applies where no such intent is present, but the effect of the technique(s) deployed by an AI system is capable of materially distorting the behaviour of a person or a group of persons to a point that undermines their individual autonomy and free choices.

(84) A plausible/reasonably likely causal link between the subliminal, purposefully manipulative or deceptive technique deployed by the AI system and its effects on the behaviour is, however, always necessary for the prohibition to apply. In consistency with consumer protection law, these effects do not need to have fully materialised, but there must be sufficient indication that they are likely or capable to materialise and undermine individual autonomy based on an objective assessment of all circumstances of the case and existing scientific knowledge and methods, as well as available information about the impact of the system on the individuals' behaviour in real life. In this context, the fact that a system is capable of triggering behaviours that appreciably impair individuals' ability to make an informed decision and undermine their free choices suffice to fulfil that condition and is not dependent on considerations relating to the 'timing' for the harm to materialize (e.g. in case of addition-like behaviours) as long as it is reasonably likely to occur.

For example, an AI-powered well-being chatbot is intended by the provider to support and steer users in maintaining a healthy lifestyle and provide tailored advice for psychological and physical exercises. However, if the chatbot exploits individual vulnerabilities to adopt unhealthy habits or to engage in dangerous activities (e.g. engage in excessive sports without rest or drinking water) where it can reasonably be expected that certain users will follow that advice, which they would otherwise not have

done, and suffer significant harm (e.g. a heart attack, or other serious health problem), that AI system would fall under the prohibition in Article 5(1)(a) AI Act, even if the provider might not have intended this behaviour and harmful consequences for the persons.

The mere fact that the chatbot is capable of appreciably impairing individual autonomy and materially distorting the behaviour of certain users in a significantly harmful way and that the provider has not taken appropriate preventive and mitigating measures to avoid those significantly harmful effects suffices for the prohibition to apply (see more for relevant considerations of the reasonably likeliness of the harm in section 3.2.3. and out of scope section 3.5.).

3.2.3. (Reasonably likely to) cause significant harm

(85) Finally, for the prohibition in Article 5(1)(a) AI Act to apply, the distortion of the behaviour of a person or group of persons must cause or be reasonably likely to cause that person, another person, or group of persons significant harm. In this context, important concepts that require clarification are the types of harms covered by the prohibition, the threshold of significance of the harm, and its reasonable likelihood and causal link between the harm and the manipulative or deceptive technique and the person's behaviour.

a) Types of harms

(86) The AI Act addresses various types of harmful effects associated with manipulative and deceptive AI systems, each with distinct implications for individual persons and groups of persons that may be affected[74]. The main types of harms relevant for Article 5(1)(a) AI Act include physical, psychological, financial, and economic harms[75] that may be compound with broader societal harms in certain cases[76].

(87) Physical harm encompasses any injury or damage to a person's life, health and material damage to property. Physical harm to a person's life and health have, in many cases, immediate, serious, and irreversible consequences. In line with its product safety logic, the AI Act aims to prohibit AI-enabled manipulation and deception resulting in significant physical harm.

For example, an AI chatbot promotes self-harm to users or incentivises them to commit suicide or harm other persons or groups of persons by promoting terrorist content or incentivising violence against certain persons or groups of persons (i.e., minorities).

(88) Psychological harm is particularly relevant in the context of AI systems deploying manipulative techniques that exploit cognitive and emotional vulnerabilities and

[74] See Article 5(1)(a) AI Act.
[75] Recital 29 AI Act.
[76] See Recital 28 AI Act, which explains the prohibitions can also cause broader societal harms and contradict Union values of respect for human dignity, freedom, equality, democracy and the rule of law and fundamental rights enshrined in the Charter. See also Article 1 AI Act that aims to protect democracy and the rule of law as EU values.

influence an individual's behaviours in ways that can cause significant harm. Psychological harms encompass adverse effects on a person's mental health and psychological and emotional well-being. Such harms are particularly significant because they may accumulate over time and may not be immediately apparent, but may produce long-lasting and severe consequences. However, it is more difficult to measure them, which requires a case-by-case assessment, in particular to determine their severity, taking into account all relevant circumstances of the case.

For example, an AI companionship application designed to emulate human speech patterns, behaviours and emotions uses anthropomorphic features and emotional cues to influence users' feelings, dispositions, and opinions, making those users emotionally dependent on the service incentivising addiction-like behaviour and potentially causing significant harms, such as suicidal behaviours and risks of harming other persons.[77]

(89) Financial and economic harm may encompass a range of adverse effects, including financial loss, financial exclusion, economic instability.

For example, a chatbot that offers fraudulent products that cause significant financial harms.

(90) In assessing the harms caused by AI systems when applying Article 5(1)(a) AI Act, it is important to highlight that the harms are often not isolated, but manifest themselves in combination, leading to compounded and multifaceted negative impacts. Understanding the combination of harm is crucial to effectively assess their significance (see also see section 3.2.3. b) below), whereby physical, psychological, financial and economic harm may be combined and exacerbate the overall impact on individuals and communities and may even have broader adverse impacts.

For example, - An AI system that causes physical harm may also lead to psychological trauma, stress, and anxiety and vice versus. For example, addictive design of AI systems used in products and other AI-enabled applications may lead to psychological harm by fostering addictive behaviours, anxiety, and depression. The psychological distress may subsequently result in physical harm, such as insomnia and other stress-related health issues and physical problems. - AI-driven harassment may lead to both psychological distress and physical manifestations of stress, such as insomnia, deteriorated physical health or weakened immune system.

[77] Renwen Zhang, Han Li, Han Meng, Jinyuan Zhan, Hongyuan Gan, and Yi-Chieh Lee. 2024. The Dark Side of AI Companionship: A Taxonomy of Harmful Algorithmic Behaviors in Human-AI Relationships. 1, 1 (November 2024), 28 pages.

- Psychological harm resulting from the use of AI may also lead to physical harm, including death. For example, AI systems used online may facilitate gender-based violence through harassment, stalking, cyberbullying and sexual extortion.

- Individual psychological harms for example due to AI-enabled generation of 'deep fakes' impersonating actual persons to deceive and undermine the decision-making, individual autonomy and free choices of individuals can also be combined with significant harms for groups of persons (e.g., sharing the same ethnic or racial origin or gender as the victims depicted on the deep fakes).

b) Threshold for significance of the harm

(91) The prohibition in Article 5(1)(a) AI Act only applies if the harm caused by the subliminal, manipulative and deceptive techniques is '**significant**'. The AI Act does not provide a definition for the concept of 'significant harm', but it should be understood as implying **significant adverse impacts** on physical, psychological health or financial and economic interests of persons and groups of persons[78]. The determination of 'significant harm' is fact -specific, requiring careful consideration of each case's individual circumstances and a case-by-case assessment, but the individual effects should be always material and significant in each case.

(92) In other Union laws, the concept of 'significant harm' is also used as a nuanced and context-dependent concept guided by high-level protection and preventive action goals.[79] By analogy, the following key considerations can be derived and could be taken into account when assessing what constitutes significant harm within the meaning of Article 5(1)(a) AI Act:

- **The severity of harm** refers to the degree of harm that has resulted or is reasonably likely to result from using an AI system with objective and observable effects for the significant harms. It is particularly important to consider in this context the AI system's interdependencies, the combination of various types of harms, and the adverse effects on individuals or groups of persons.

- **Context and cumulative effects[80]:** The specific context, including the existing state, and the cumulative effects of multiple actions, play an important role in assessing the severity of the harm.

- **Scale and intensity:** The extent of the harm and the intensity of adverse effects are critical in evaluating whether the harm is significant. Whether the harm impacts a large number of people is also relevant for assessing its significance.

- **Affected persons' vulnerability:** Certain groups, such as children, the elderly, or persons with disabilities may be more susceptible to harm from specific AI systems.

[78] Recital 29 AI Act.
[79] See judgments of the Court of Justice of 7 September 2004, *Waddenvereniging and Vogelbeschermingsvereniging*, C-127/02, EU:C:2004:482 and of 11 April 2013, *Sweetman and Others*, C-258/11, EU:C:2013:220.
[80] See Recital 29 AI Act.

What may be considered less significant harm for persons in general might be considered significant and unacceptable for such vulnerable groups, especially children.

- **Duration and reversibility:** Long-lasting or irreversible harm typically meets the threshold for significant harm. Short-term and reversible effects might be considered less significant, unless they occur frequently.

(93) The objective of the AI Act to ensure 'a high level of protection', in conjunction with Article 191(2) TFEU, suggests a comprehensive approach to protection when assessing the significance of the harm. This means considering both immediate and direct harms and systemic, indirect adverse impacts associated with AI systems deploying subliminal, purposefully manipulative or deceptive techniques that are intended to or capable of impairing individual autonomy, decision-making and free choices of persons and groups of persons.

For example, significant physical harm that is reasonably likely to be caused by an AI system includes injuries or fatalities or a sufficiently serious impact on individuals' health or the destruction of property. AI systems that suggest to an individual to commit criminal acts such as sexual abuse and exploitation, extreme violent or terrorist content or incentivise individuals to commit crimes, self-harm or harm to other persons should be considered to reach such a threshold.

By contrast, minor physical harms may include less severe injuries, such as bruises or temporary discomfort, which do not have significant or lasting consequences and will therefore not reach the threshold of significance within the meaning of Article 5(1)(a) AI Act. Whether the physical harm specifically concerns vulnerable groups, such as children, should be assessed, as should the scale of the harm and whether it is compounded with other types of harms, such as psychological, financial etc. This will require a case-by-case assessment, taking into account the circumstances and the criteria presented above to guide that assessment.

There are numerous cases where the threshold of significant harm will likely not be reached even if the systems may be deploying subliminal, purposefully manipulative, or deceptive techniques (see for examples section 3.5. below).

c) Causal link and threshold for reasonable likelihood of the harm

(94) The concept of 'reasonably likely' is used in Article 5(1)(a) AI Act to determine whether there is a plausible/reasonably likely causal link between the manipulative or deceptive technique capable of distorting the person's behaviour in a manner that undermines their free choices and the potential significant harm. This concept allows the application of the prohibition not only in cases where the harm has occurred, but also where it is reasonably likely to occur in line with the safety logic of the AI Act. In this context, it is particularly relevant to assess whether the provider or deployer of the

AI system could have reasonably foreseen the significant harm that is reasonably likely to result from the subliminal, purposefully manipulative or the deceptive techniques deployed and whether they implemented appropriate preventive and mitigating measures to avoid or mitigate the risk of such significant harms. This implies a judgement of reasonableness on an objective basis and according to universally accepted criteria (e.g. technical and scientific), including a criterion of rationality in establishing plausible causality between the AI practice and the significant harm that may arise. The opacity or transparency of the AI system and its functioning may affect the conclusion regarding this causal link and, hence, the application of the prohibition.

(95) To avoid providing or using AI systems that are likely to be prohibited, providers and deployers of AI systems that deploy such manipulative or deceptive techniques are encouraged to take appropriate measures such as:

1. **Transparency and individual autonomy:** provide transparency in how the AI system operates, clear disclosures about its capabilities and limitations, and relevant information to ensure informed decisions; respect individual autonomy and avoid engaging in exploitative or deceptive practices that are likely to appreciably impair an individual's autonomy, decision-making and free choices in potentially harmful ways; integrate appropriate user control and safeguard measures to ensure that the system is not deceptive and operates within the boundaries of lawful persuasion that is outside the scope of the prohibition (see section 3.5.1).
2. **Compliance with relevant applicable legislation:** in many cases compliance with relevant applicable legislation will mitigate the risks of harm and indicate that the practice does not constitute a purposefully manipulative or deceptive practice and that mitigating measures have been put in place to prevent likely significant harms (see section 3.4. and 3.5.1.).
3. **State of the art practices and industry standards:** adherence to professional due diligence practices and industry standards for the responsible development and use of safe and ethical AI systems and measures to mitigate the harms can help to pre-empt and mitigate unintended significant harms.

(96) By contrast, harms and distortion of the behaviour of individuals which result from factors external to the AI system and which are not within the control and reasonably foreseeable by the provider or the deployer to pre-empt and mitigate risks would not be relevant for the assessment whether there is a plausible causal/reasonably likely link between the distorted behaviour of the persons interacting with the system and the significant harm[81].

For example, the provider of an AI system may assess and try to mitigate potential harmful manipulative effects in the design of the system and the interactions with humans through the design, prior testing, and other proportionate mitigating measures,

[81] See Recital 29 AI Act.

but it may not be in a position to foresee if a person may get depressed or change their behaviour due to other external factors in their personal life that are not known and beyond their interactions with the system.

(97) Other examples that fall outside the scope of the prohibition as not fulfilling all conditions (e.g. in case of lawful persuasion) are provided in section 3.5. below.

3.3. Main components of the prohibition in Article 5(1)(b) AI Act – harmful exploitation of vulnerabilities

Article 5(1)(b) AI Act provides:

1. The following AI practices shall be prohibited:

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

(98) Several cumulative conditions must be fulfilled for the prohibition in Article 5(1)(b) AI Act to apply:

(i) The practice must constitute the 'placing on the market', the 'putting into service', or the 'use' of an AI system.

(ii) The AI system must exploit vulnerabilities due to age, disability, or socio-economic situations.

(iii) The exploitation enabled by the AI system must have the objective, or the effect of materially distorting the behaviour of a person or a group of persons.

(iv) The distorted behaviour must cause or be reasonably likely to cause significant harm to that person, another person, or a group of persons.

(99) For the prohibition to apply all four conditions must be simultaneously fulfilled and there must be a plausible causal link between the exploitation, the material distortion of the behaviour of the person, and the significant harm that has resulted, or is reasonably likely to result, from that behaviour.

(100) The first condition, i.e. the 'placing on the market', the 'putting into service' or the 'use' of an AI system, has been already analysed in section 2.3., while the third and fourth conditions have been examined in sections 3.2.2. and 3.2.3. in relation to the prohibition in Article 5(1)(a) AI Act. The next sections will focus on the additional specific conditions listed above, i.e. those that relate to the exploitation of the vulnerabilities and the specific harm.

3.3.1. Exploitation of vulnerabilities due to age, disability, or a specific socio-economic situation

(101) To fall within the scope of the prohibition in Article 5(1)(b) AI Act, the AI system must exploit vulnerabilities inherent to certain individuals or groups of persons due to their age, disability or a specific socio-economic situation, making them particularly susceptible to manipulative and exploitative practices.

(102) The AI Act does not define the concept of 'vulnerabilities'. That concept may be understood to encompass a broad spectrum of categories, including cognitive, emotional, physical, and other forms of susceptibility that can affect the ability of an individual or a group of persons to make informed decisions or otherwise influence their behaviour. While Article 5(1)(b) AI Act refers to 'any' vulnerability, it limits the relevant persons covered by the prohibition to those defined by their age, disability, or socio-economic situations, who in principle have more limited capacity to recognise or resist the AI manipulative or exploitative practices and are in need of enhanced protection.[82] It follows from the wording of Article 5(1)(b) AI Act that this susceptibility must be the result of the person belonging to one of the groups ('due to').

(103) 'Exploitation' should be understood as objectively making use of such vulnerabilities in a manner that is harmful for the exploited (groups of) persons or other persons and should be clearly distinguished from lawful practices that are not affected by the prohibition (see section 3.5.2 out of scope). Exploitation of the vulnerabilities of persons belonging to those clearly defined groups may be cumulative (reference to 'any') which in combination may also constitute an aggravating factor that is likely to increase the harm. Exploitation of vulnerabilities of persons and groups of persons belonging to vulnerable groups other than those defined by age, disability or specific socio-economic situation are outside the scope of Article 5(1)(b) AI Act.

a) Age

(104) Age is a primary vulnerability category covered by the prohibition in Article 5(1)(b) AI Act, including both young and older people. That prohibition aims to prevent AI systems from exploiting cognitive and other limitations that children and older people may have, and to protect them from harmful undue influence, manipulation and exploitation. This aligns with the objectives of the AI Act[83] and other Union and national legal frameworks and policies aimed at ensuring child safety[84].

(105) **Children[85]** that is, persons below the age of 18 years, are particularly susceptible to manipulation due to their developmental stage, which limits their ability to assess and understand what is real and the intentions behind AI-driven interactions critically. Children, due to their cognitive and socio-emotional immaturity, are also particularly

[82] See in particular Articles 24, 25 and 26 of the Charter. See also United Nations Educational, Scientific and Cultural Organization (UNESCO) Recommendation on the Ethics of Artificial Intelligence (2021) which emphasises inclusivity and fairness in AI development and deployment. It calls for special attention to vulnerable groups, including children, older people, and people with disabilities.
[83] Recital 48 AI Act highlights that children have specific rights as enshrined in Article 24 of the Charter and in the United Nations Convention on the Rights of the Child, further developed in the UNCRC General Comment No 25 as regards the digital environment, both of which require consideration of the children's vulnerabilities and provision of such protection and care as necessary for their well-being.
[84] See, the new European strategy for a better internet for kids (BIK+), COM/2022/212 final.
[85] Union law generally considers a child to be any person under 18, aligning with the United Nations Convention on the Rights of the Child (UNCRC).

vulnerable to forming attachments to AI agents and applications, and are therefore more susceptible to manipulation, exploitation, and addictive behaviour.

For example,

- An AI-powered toy designed to interact with children keeps them interested in interactions with the toy by encouraging them to complete increasingly risky challenges, such as climbing furniture, exploring high shelves, or handling sharp objects, in exchange for digital rewards and virtual praise, pushing them towards dangerous behaviours that are likely to cause them significant physical harm. Such a system exploits children's vulnerabilities by abusing their natural curiosity and desire for rewards.

- A game uses AI to analyse children's individual behaviour and preferences on the basis of which it creates personalised and unpredictable rewards through addictive reinforcement schedules and dopamine-like loops to encourage excessive play and compulsive usage. The game is designed to be highly addictive, exploiting the vulnerabilities inherent to children, including their limited ability to understand long-term consequences, susceptibility to pressure, lack of self-control, and inclination towards instant gratification. The consequences of this AI-enabled exploitation can be severe and long-lasting for children, including potentially additive behaviour, physical health problems due to lack of exercise and sleep, deteriorated eyesight, problems with concentration and reduced cognitive capacities, poor academic performance, and social difficulties. It can significantly impact a child's development and well-being, with potential longer-term consequences that may also extend into adulthood.

In both examples, the prohibition in Article 5(1)(b) AI Act targets only such exploitation and addiction-like practices that seriously harm children and not AI-enabled toys, games, learning applications or other digital applications in general that can bring benefits and are not affected if they do not fulfil all conditions for that prohibition. See also section 3.5. out of scope.

(106) Likewise, **older people** [86] might suffer from reduced cognitive capacities (even if not suffering from dementia) and might struggle with the complexities of modern AI technologies, making them in those cases vulnerable to scams or coercive tactics.

For example,

- AI systems used to target older people with deceptive personalised offers or scams, exploiting their reduced cognitive capacity aiming to influence them to make decisions they would not have taken otherwise that are likely to cause them significant financial harm.

- A robot aimed to assist older persons may exploit their vulnerable situation and force them to do certain activities against their free choice, which can significantly worsen their mental health and cause them serious psychological harms.

In both examples, the prohibition in Article 5(1)(b) AI Act targets only such exploitative practices that are likely to seriously harm older persons and not AI-enabled personal assistants, health applications and assistive robots in general that can bring benefits and are not affected if they do not fulfil all conditions for that prohibition. See also section 3.5. out of scope.

b) Disability

(107) The second category of vulnerabilities which the prohibition in Article 5(1)(b) AI Act aims to protect are those due to disability. The objective is to prevent AI systems from exploiting cognitive and other limitations and weaknesses in persons with disabilities and to protect them from harmful undue influence, manipulation, and exploitation.

(108) Disability[87] encompasses a wide range of long-term physical, mental, intellectual, and sensory impairments which in interaction with other barriers hinder full and effective participation of individuals in the society on an equal basis with others. AI systems that exploit such vulnerabilities may be particularly harmful for persons with disabilities which can be more easily influenced or exploited due to their impairment compared to other persons.

For example,

- A therapeutic chatbot aimed to provide mental health support and coping strategies to persons with mental disabilities can exploit their limited intellectual capacities to influence them to buy expensive medical products or nudge them to behave in ways that are harmful to them or other persons.
- AI systems can identify women and young girls with disabilities online with sexually abusive content and targets them with more effective grooming practices, thus exploiting their impairments and vulnerabilities that make them more susceptible to manipulation and abuse and less capable of protecting themselves.

By contrast, AI applications that are not designed in an accessible manner should not be regarded to exploit vulnerabilities of persons with disabilities since they do not specifically target those vulnerabilities, but are simply inaccessible to the persons with disabilities.

[87] Recital 29 AIA explains that 'disability' should be understood within the meaning of Article 3(1) Directive (EU) 2019/882 of the European Parliament and of the Council of 17 April 2019 on the accessibility requirements for products and services (Text with EEA relevance), PE/81/2018/REV/1, OJ L 151, 7.6.2019, p. 70–115.

c) Specific socio-economic situation

(109) The third category of vulnerabilities which the prohibition in Article 5(1)(b) AI Act seeks to protect are those due to a specific socio-economic situation that is likely to make the persons concerned more vulnerable to exploitation. 'Specific' should not be interpreted in this context as a unique individual characteristic, but rather a legal status or membership to a specific vulnerable social or economic group. Recital 29 AI Act contains a non-exhaustive list of examples of such situations, such as persons living in extreme poverty and ethnic or religious minorities. The category aims to cover, in principle, relatively stable and long-term characteristics, but transient circumstances, such as temporary unemployment, over-indebtedness or migration status, may also be covered as a specific socio-economic situation. However, situations such as grievances or loneliness that may be experienced by any person are not covered, since they are not specific from a socio-economic perspective (their exploitation may be covered though under Article 5(1)(a) AI Act).

(110) Persons in disadvantaged socio-economic situations are usually more vulnerable and have fewer resources and lower digital literacy than the general population, which makes it harder for them to discern or counteract exploitative AI practices. Article 5(1)(b) AI Act aims to ensure that AI technologies do not perpetuate or exacerbate existing financial and other social inequalities and injustices by exploiting the vulnerabilities of those people.

> For example, an AI-predictive algorithm can be used to target with advertisements for predatory financial products people who live in low-income post-codes and are in a dire financial situation, thus exploiting their susceptibility to such advertisements because of possible despair and causing them significant financial harm.
>
> By contrast, an AI system that is inadvertently biased and disproportionately impacts socio-disadvantaged persons (indirect discrimination) due to biased training data should not automatically be considered to exploit persons' socio-economic vulnerabilities, since they are not specifically targeted as in the case of direct discrimination when such targeting is a deliberate feature of the system's design of the algorithm or when such discriminatory impact is due to targeting other proxy characteristics (e.g. postal codes) that closely correlate with the protected characteristics. At the same time, providers or deployers of AI systems that are aware that their systems unlawfully discriminate against persons or groups of persons in specific socio-economic situation should also be considered to exploit their vulnerabilities, if they are aware of the reasonably likely significant harm that they are likely to suffer and they have not taken appropriate corrective measures (see section 3.2.3. c) above).

(111) In the context of a specific socio-economic situation, it is crucial to consider the relevance of proxies linked to grounds of discrimination protected under Union equality law, such as racial origin, ethnicity, nationality or religion.

For example, socio-economic status and ethnic origin might intersect, meaning that AI systems exploiting socio-economic data might disproportionately affect ethnic minorities or persons from specific racial origin. This can exacerbate existing disparities and contribute to systemic discrimination or even exclusion of individuals from these groups.

However, Article 5(1)(b) AI Act does not apply to AI systems that target consumers based on a wide ranging variables that do not tangentially correlate with vulnerable groups in specific socio-economic situations, such as what brand and model of telephone a person has, in how big city they live, how much and where they travel etc. Even if these characteristics may reflect socio-economic situation of individuals in general, they are not determinative of individuals in a specific socio-economic situation, whose vulnerabilities the prohibition aims to safeguard against exploitation.

(112) Other people in unique social contexts may be, for instance, migrants or refugees, who often lack stable legal status and socio-economic stability and may be particularly susceptible to exploitation by AI systems.

For example, a chatbot is intended to interact in a personalised manner with users, some of whom happen to be migrants. The chatbot identifies and makes use of the vulnerabilities and discontent of migrants, who are in principle in a vulnerable and instable specific socio-economic situation, and pushes them towards extremist views in response to their queries, including violence against (certain groups of) the population in the country.

3.3.2. With the objective or the effect of materially distorting behaviour

(113) The third condition for the prohibition in Article 5(1)(b) AI Act to apply is that the exploitation of the vulnerabilities examined above must have either a) 'the objective' or b) 'the effect of materially distorting the behaviour of a person or a group of persons'. This implies a substantial impact, rather than a minor or trivial one, but does not necessarily require intent, since Article 5(1)(b) AI Act covers practices that may only have the 'effect' of causing material distortions. Article 5(1)(a) and (b) AI Act make use of the same concepts and should therefore be interpreted in the same way. The explanations provided in section 3.2.2. are therefore equally relevant for Article 5(1)(b) AI Act. The only noteworthy difference is the need in Article 5(1)(a) AI Act for the exploitative practice to 'appreciably impair the ability to make an informed decision', which is not present in Article 5(1)(b) AI Act, since the specific vulnerabilities of children and other vulnerable persons reduce their capacity to make such informed decisions and force them into adopting behaviour against which they cannot protect themselves as other adults might do.

3.3.3. (Reasonably likely to) cause significant harm

(114) Finally, for the prohibition in Article 5(1)(b) AI Act to apply, the distortion of the behaviour of the vulnerable person or group of persons must cause or be reasonably likely to cause that person or another person significant harm. Article 5(1)(a) and (b) AI Act make use of the same concepts and should therefore be interpreted in the same way. The explanations provided in section 3.2.3. in relation to the types of harms, the threshold for significance of the harm, and the causal link and its reasonableness are thus equally relevant for the interpretation of Article 5(1)(b) AI Act.

(115) As explained in section 3.2.3., significant harm encompasses a range of significant adverse impacts, including physical, psychological, financial, and economic harms that must be reasonably likely to occur for the prohibition in Article 5(1)(b) AI Act to apply. For vulnerable groups — children, older persons, persons with disabilities, and socio-economically disadvantaged populations — these harms may be particularly severe and multifaceted due to their heightened susceptibility to exploitation. What may be considered an acceptable risk of harm for adults often represents an unacceptable harm for children and these other vulnerable groups. A precautionary approach is therefore particularly warranted in case of uncertainty and potential for significant harms.

(116) For instance, **children** are highly impressionable and may not possess the cognitive maturity to evaluate persuasive content critically or resist certain exploitative practices that aim to keep them dependent on the AI-enabled services. This could, in turn, contribute to the shaping of their values, beliefs and steer behaviours in potentially harmful ways. The significant harm here is both physical and psychological, exacerbated by the children's inability to discern and resist the exploitation and the harmful effects to their development and well-being that may have a long-term impact.

For instance,

- AI systems used for the generation of child sexual abuse material (or manipulating existing material depicting real children to create further novel content featuring them) and the development of strategies for grooming and sexually extorting children are likely to cause serious harms and abuses of the affected children and often result in long-term physical, psychological and social consequences for survivors[88].

- AI systems may target the vulnerabilities of young users and use addictive reinforced schedules with the objective to keep them dependent on the service are particularly harmful for young persons and girls. They may cause serious psychological and physical harms, including anxiety and depression, body dissatisfaction, eating disorders and mental health problems, including in some cases self-harm and suicidal behaviour.[89] This may also have long term harmful consequences for child

[88] Commission Staff Working Document: Impact Assessment Report accompanying the document Proposal for a Directive of the European Parliament and the Council on combating child sexual abuse and sexual exploitation and child sexual abuse material, SWD/2024/33 final. See also for statistics Internet Watch Foundation 2024 report which contains detailed statistics on AI generated CSAM, available at: https://www.iwf.org.uk/about-us/why-we-exist/our-research/how-ai-is-being-abused-to-create-child-sexual-abuse-imagery,

[89] Elizabeth J. et al, A meta-analysis of the association between adolescent social media use and depressive symptoms, Journal of Affective Disorders, Volume 275, 1 October 2020, Pages 165-174.

development, including impaired cognitive development and learning and reduced social skills and displacement of experiences such as physical play, sleep, and face-to-face social interactions that are essential for the emotional and physical well-being of the child[90].

- An AI system that is designed in an anthropomorphic way and simulates human-like emotional responses in its interactions with children can exploit children's vulnerabilities in a manner that fosters unhealthy emotional attachment, manipulates engagement time and distorts children's understanding of authentic human relationships. This may hamper their normal social and emotional development and relations with other human beings and socio-emotional skills like empathy, emotional regulation, and social understanding and adaptability[91]. As a result, this may lead to psychological harms such as increased anxiety and dependency of the children on the service and longer-term harms to the child's well-being.

Such deliberate addictive and exploitative design features of the AI-enabled services that can lead to combined significant harms, as described above, should be distinguished from other legitimate behaviour of the providers and deployers to pursue user engagement that respects individual autonomy and the safety of children and do not lead to significant harms which are outside the scope of Article 5(1)(b) AI Act (see section out of scope 5.3).

(117) Similarly, **older people** may face cognitive decline and reduced digital literacy, making them prime targets for AI-driven scams or manipulative marketing. The harm in this case is often financial and psychological, compounded by the frustration and isolation many older people experience, which may be exploited to amplify the manipulative impact.

For instance, an AI system that exploits the reduced cognitive vulnerabilities of older people by particularly targeting expensive medical treatments, unnecessary insurance policies or deceptive investments schemes to older persons may lead to significant loss of savings, increased debt, and emotional distress for older people.

Certain AI-enabled differential pricing practices in key services such as insurance that exploit the specific socio-economic situation and provide higher prices to lower income consumers can lead to a significant financial burden to pay more for the same coverage, leaving them vulnerable to shocks.[92]

[90] Siebers, T., Beyens, I., Pouwels, J. L. & Valkenburg, P. M. Social Media and Distraction: An Experience Sampling Study among Adolescents. Media Psychology 25, 343–366 (2022).
[91] Laestadius, L., Bishop, A., Gonzalez, M., Illenčík, D. & Campos-Castillo, C. Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. New Media & Society 146144482211420 (2022) doi:10.1177/14614448221142007; Neugnot-Cerioli, M. & Laurenty, O. M. The Future of Child Development in the AI Era. Cross-Disciplinary Perspectives Between AI and Child Development Experts. Preprint at https://doi.org/10.48550/ARXIV.2405.19275 (2024).
[92] 2023 EIOPA Consumer Trends Report, page 16, last paragraph.

(118) **Persons with disabilities** also represent a vulnerable group that exploitative and manipulative AI systems may significantly harm.

> For instance, an AI system that uses emotion recognition to support mentally disabled individuals in their daily life may also manipulate them into making harmful decisions, like purchasing products promising unrealistic mental health benefits. This is likely to worsen their mental health condition and financially exploit them through the purchase of ineffective and expensive products, which is likely to cause them significant psychological and financial harms.

(119) **Socio-economically disadvantaged individuals** are particularly susceptible to AI systems exploiting their financial desperation and precarious social situation and are often less informed and digitally literate.

> For instance, an AI chatbot could target specific socio-economically disadvantaged groups inciting them to commit acts of violence or injuries of other persons, by identifying their heightened susceptibility to certain types of content, fear-based narratives, or exploitative offers. The system's targeted approach exacerbates the existing vulnerabilities of these socio-economically disadvantaged individuals, deepening their challenges. In certain cases this may lead to increased anxiety, depression, feelings of helplessness, social isolation, or self-harm and radicalisation to a point that reaches the threshold of significant harm under Article 5(1)(b) AI Act.

(120) Unlike Article 5(1)(a) AI Act, Article 5(1)(b) AI Act does not explicitly refer to group harms, while recital 29 AI Act refers for both prohibitions to harms suffered by both specific persons and groups of individuals. The two prohibitions should thus be interpreted in a consistent manner aligned also with the safety logic of the AI Act and the objective of the prohibition in Article 5(1)(b) to protect all individuals belonging to the specific vulnerable groups due to age, disability and specific socio-economic situation. Harms that can be externalised and affect other persons, even if not directly affected by the system, should therefore also be taken into account in the assessment of the significance of the harm under Article 5(1)(b) AI Act.

> For instance,
>
> - The AI-enabled exploitation of children's vulnerabilities may have long-term societal impacts, including increased prevalence of mental health concerns, healthcare costs, and reduced productivity due to chronic health issues.
>
> - An AI system exploiting the financial vulnerabilities of economically disadvantaged people may lead to financial exclusion and create a downward spiral of socio-economic hardship for those disadvantaged groups. Such exploitation may cause societal harms with broader negative impacts on societal structures and values, including the perpetuation and exacerbation of discrimination and social inequality and the exclusion of those groups.

- A chatbot targeting certain vulnerable socio-economic groups with misinformation or hate speech may lead to social polarisation and radicalisation, possibly igniting violence and even injuries and deaths of other persons.

(121) These examples of exploitative AI practices should be distinguished from numerous other AI systems which do not exploit the vulnerabilities of children, persons with disabilities, or persons in specific socio-economic situations and are not reasonably likely to cause significant harms, but aim to benefit those persons when properly designed and used (see also section 3.5. out of scope).

For instance,

- AI systems that support children in their learning and in games;
- AI systems that help older persons in their daily life and improve their health and medical treatment, such as personal assistants or assistive robots, or improve their digital skills;
- AI systems that support the economic and other integration of socio-disadvantaged persons in the society, improve their skills, etc.;
- AI systems and devices that support visually or hearing impaired persons or provide adapted and personalised learning;
- AI systems that generate accessible solutions removing barriers for the use by persons with disabilities of products and services;
- AI-enabled prosthetics etc. that help disabled persons in their daily life and enable their integration and full participation in society.

3.4. Interplay between the prohibitions in Article 5(1)(a) and (b) AI Act

(122) The interplay between the prohibitions in Article 5(1)(a) and (b) AI Act requires the delineation of the specific contexts that each provision covers to ensure that they are applied in a complementary manner.

(123) The primary focus of the prohibition in Article 5(1)(a) AI Act is placed on the nature of the techniques, specifically those that operate below the threshold of conscious awareness or other purposefully manipulative or deceptive techniques. The key elements here are the primarily covert nature of the influence and its impact on the individual affected by the system that undermines their cognitive autonomy to make informed and autonomous decisions.

(124) By contrast, the primary focus of the prohibition in Article 5(1)(b) AI Act is the protection of particularly vulnerable persons due to their age, disability, or a specific socio-economic situation, which are in principle more susceptible to AI exploitation due to inherent or situational factors and, therefore, require additional protection against exploitation. The key elements here are the characteristics of the affected vulnerable

persons and the fact that their specific vulnerabilities are being exploited by the AI system.

> For instance, if an AI system uses rapid image flashes to influence purchasing decisions, it may fall under Article 5(1)(a) AI Act due to the subliminal nature of the manipulation. Conversely, an AI system that targets older persons with insurance offers by exploiting their reduced cognitive capacity may fall under Article 5(1)(b) AI Act.

(125) In scenarios where both provisions may seem applicable, the primary criterion for differentiation should be the dominant aspect of the exploitation. If the exploitation applies regardless of the specific vulnerabilities of the persons concerned, Article 5(1)(a) AI Act should take precedence, while taking into account the particular effects of the manipulative or deceptive technique on the vulnerable persons' behaviour and the specific harms that those persons are likely to experience. If the AI-enabled manipulation and exploitation is targeted at a specific vulnerable group of persons due to their age, disability, or specific socio-economic situation or aimed to exploit their vulnerabilities, then Article 5(1)(b) AI Act should be applied instead. Exploitation of vulnerabilities of other groups may be covered as part of Article 5(1)(a) AI Act if the purposefully manipulative practice leverage on specific vulnerabilities and weaknesses of those persons.

3.5. Out of scope

(126) For the prohibitions in Article 5(1)(a) and (b) AI Act to apply, all conditions listed in the relevant provisions must be fulfilled, as examined above. All other AI systems that do not fulfil these conditions are outside the scope of those prohibitions, with some examples described below.

3.5.1. Lawful persuasion

(127) Distinguishing manipulation from persuasion is crucial to delineate the scope of the prohibition in Article 5(1)(a) AI Act, which does not apply to lawful persuasion practices. While both manipulation and persuasion influence individuals' decisions and behaviours, they differ significantly in methods and ethical implications.

(128) Manipulation involves, in most cases, covert techniques undermining autonomy, leading individuals to make decisions they might not have otherwise made if they were fully aware of the influences at play. These techniques often exploit psychological weaknesses or cognitive biases. By contrast, persuasion operates within the bounds of transparency and respect for individual autonomy. It involves presenting arguments or information in a way that appeals to reason and emotions, but explains the AI system's objectives and functioning, provide relevant and accurate information to ensure informed decision-making and supports the individual's ability to evaluate the information and make free and autonomous choices.

For example, an AI system using personalised recommendations based on transparent algorithms and user preferences and controls engages in persuasion. By contrast, a system that uses subliminal clues (e.g. imperceptible images) to influence users towards specific choices without their knowledge and understanding constitutes manipulation.

(129) The objective and impact of these techniques also differ. Manipulation often aims at benefitting the manipulator at the expense of the individual's autonomy and well-being. By contrast, persuasion aims to inform and convince, aligning interests and benefits for both parties. Ethical persuasion respects an individual's autonomy to make informed choices and avoids exploiting vulnerabilities.

For example, an AI system which operates in a transparent manner and analyses customers' emotions to improve customer interactions and provide support with the knowledge of the users engages in persuasion and aligns with their interests. By contrast, an emotion recognition system used for targeted advertising that infers emotions of consumers in a hidden manner to offer products of higher prices at a specific moment when the user is more likely to buy them engages in manipulation and is to the detriment of the consumers.

(130) Consent also plays an important role in certain cases. In persuasive interactions, individuals are aware of the influence attempt and can freely and autonomously choose it. In manipulative interactions, the lack of awareness of the techniques or their impact negates the freedom of choice and informed and autonomous decision-making.

For example, an AI system that aims to help users learn a foreign language better and faster through the deployment of subliminal techniques is not manipulative if it operates in a transparent manner and respects individual autonomy and user's free and informed choice to consent to the use of the system or not.

(131) Compliance with legal and regulatory frameworks also plays an important role in measuring manipulation as compared to lawful persuasion. AI practices that comply with applicable laws that uphold transparency, fairness, and individuals' rights and autonomy are therefore more likely not to be prohibited under the AI Act.

For example, compliance with data protection laws, such as the GDPR, which mandates transparency obligations in data processing, namely that the information to be provided to the data subjects should avoid deceptive or manipulative language[93]. In some instances, consent may be required for personal data processing to be lawful, as for certain online personalised advertisements based on off-service users' data in social networks[94]. That consent must be, amongst others, free and informed. AI systems that meet these legal standards are more likely to engage in lawful persuasion. Conversely,

[93] European Data Protection Board Guidelines, https://www.edpb.europa.eu/system/files/2023-02/edpb_03-2022_guidelines_on_deceptive_design_patterns_in_social_media_platform_interfaces_v2_en_0.pdf, para. 18.

[94] Judgment of the Court of Justice of 4 July 2023, *Meta Platforms and Others*, C-252/21, ECLI:EU:C:2023:537 (hereinafter referred to as the '*Meta Platforms* judgment').

systems that circumvent these requirements to influence behaviour are likely engaging in manipulation.

(132) In particular, recital 29 AI Act clarifies that the prohibitions in Article 5(1)(a) and (b) AI Act do not affect lawful practices in the context of medical treatment under certain conditions.

For example, AI-enabled subliminal techniques may be used in the psychological treatment of a mental disease or physical rehabilitation when carried out in accordance with the applicable law and medical standards, including obtaining the explicit consent of the individual or its legal representatives as a condition for use.

(133) Furthermore, recital 29 AI Act clarifies that common and legitimate commercial practices, such as advertising, should not be regarded 'in themselves' or by their very nature as harmful manipulative, deceptive or exploitative AI-enabled practices.

For example,

- Advertising techniques that use AI to personalise content based on user preferences are not inherently manipulative if they do not deploy subliminal, purposefully manipulative or deceptive techniques that subvert individual autonomy or exploit vulnerabilities in harmful ways as prohibited under Article 5(1)(a) and (b) AI Act. Compliance with the relevant obligations under the GDPR, consumer protection law and Regulation (EU) 2022/2065 ('the DSA') help to mitigate such risks.

- The generation of child sexual abuse material to train and improve the effectiveness of AI models and classifiers to detect child sexual material online are common legitimate practices that are not exploitative of children's vulnerabilities and are, on the contrary, essential to improve child safety online.

- AI systems used for providing banking services, such as mortgages and loans, that use the age or the specific socio- economic situation of the client as an input, in compliance with Union legislation on financial service, consumer protection, data protection and non-discrimination, do not qualify as the exploitation of vulnerabilities within the meaning of Article 5(1)(b) AI Act when they are designed to protect and support people identified as vulnerable due to their age, disability or specific socio-economic circumstances and are beneficial for those groups, contributing also to fairer and more sustainable financial services for those groups.

- AI systems that detect drowsiness and fatigue in drivers and alert them to rest in compliance with safety laws are beneficial and do not qualify as exploitation of vulnerabilities within the meaning of Article 5(1)(b) AI Act.

3.5.2. Manipulative, deceptive and exploitative AI systems that are not likely to cause significant harm

(134) An essential condition for the prohibitions in Article 5(1)(a) and (b) AI Act to apply is that the AI-enabled manipulation and exploitation of vulnerabilities should cause or be reasonably likely to cause significant harm. All manipulative, deceptive and exploitative AI applications that are not reasonably likely to cause significant harms are in principle outside the scope of the prohibitions without prejudice to other Union law that still applies (see section 3.6. below).

Examples of AI systems that are not likely to cause significant harm include:

- An AI companionship system is designed in an anthropomorphic way and with affective computing to make the system more appealing and effectively makes users more engaged, but is not engaging in other manipulative or deceptive practices in a manner that is reasonably likely to cause them serious psychological, physical or other harms, unhealthy attachment and dependency.

- A therapeutic chatbot uses subliminal techniques to steer users towards a healthier lifestyle and to quit bad habits, such as smoking. Even if the users who follow the chatbot's advice and subliminal therapy experience some physical discomfort and psychological stress due to the effort made to quit smoking, the AI-enabled chatbot cannot be considered likely to cause significant harm. Such temporary discomfort is unavoidable and outweighed by the long-term benefits for users' health. There are no hidden attempts to influence decision-making beyond promoting healthy habits.

- An online music platform uses an emotion recognition system to infer users' emotions and automatically recommends them songs in line with their moods, while avoiding excessive exposure to depressive songs. Since users are just listening to music and are not otherwise harmed or led to depression and anxiety, the system is not reasonably likely to cause significant harm.

- AI-enabled manipulative and deceptive techniques used in security training and other learning simulations that mimic phishing attempts to educate users on cybersecurity threats. These systems may deploy purposefully manipulative techniques (e.g., exploiting cognitive biases) without users' awareness that distort the behaviour, but this is done temporarily for beneficial training and awareness raising purposes and without causing significant harms.

3.6. Interplay with other Union law

(135) The prohibitions in Article 5(1)(a) and (b) AI Act are without prejudice to and complement other Union law. The same practice that falls within the prohibition of Article 5(1)(a) or (b) AI Act may also constitute an infringement of other Union law legislation and be subject to enforcement under both the AI Act and those other acts. This is important because different provisions in those acts aim to protect different interests and have different objectives, scopes, and addressees. This ensures a comprehensive regulatory approach that protects persons and groups of persons from

harmful AI exploitation and manipulation and ensures safe and trustworthy AI-enabled services and products in the Union.

(136) The prohibitions in Article 5(1)(a) and (b) AI Act align closely with the objectives of EU consumer protection law, in particular the UCPD which protects consumers from business practices that are misleading or aggressive, including when they are AI-driven. Both the AI Act and the UCPD aim to proactively prevent consumer harm from AI-driven business practices that are manipulative, misleading, or aggressive. At the same time, the prohibitions in Article 5(1)(a) and (b) AI Act are broader in scope, since they protect not only consumers, but any natural person and their behaviour in a variety of contexts beyond commercial settings. The harms covered by the AI Act are also broader beyond economic harms, although the AI Act sets a threshold of significant harm that is not present in consumer protection law.

(137) The prohibitions are also consistent with Union data protection law, including the principles on lawful, fair, and transparent data processing, that aims to protect data subjects' personal data and ultimately preserve their fundamental rights and autonomy. The availability of more (personal) data and the increased possibilities to process this data with AI systems increase the risk of harmful manipulative, deceptive or exploitative practices, such as those falling within the scope of Article 5(1)(a) and (b) AI Act. In this context, compliance with the data protection rules for transparency, data minimisation, fairness and lawfulness, for example for personalised profiling and advertising, based on off-service users' data[95] may contribute to avoid harmful personalised manipulation and exploitation.

(138) The interplay with Union non-discrimination law is also relevant for the prohibition in Article 5(1)(b) AI Act,[96] given that vulnerabilities due to age and disability are also protected grounds on which people have the right not to be discriminated, while socio-economic situation intersects with a variety of other grounds, such as race and ethnic origin. The prohibitions in the AI Act do not affect prohibitions based on other grounds or discriminatory practices that do not entail significant harms and that are already prohibited by Union non-discrimination law.

(139) The prohibitions in Article 5(1)(a) and (b) AI Act are also complementary to Regulation (EU) 2022/2065 (the Digital Services Act ('DSA') which regulates online intermediary services, such as online platforms and search engines, and ensures transparency and accountability in the provision of those services. Notably, Article 25(1) DSA prohibits

[95] Particularly relevant in this respect is the Judgment of the Court (Grand Chamber) of 4 July 2023, Case C-252/21 *Meta Platforms Inc and Others v Bundeskartellamt*. Although the CJEU finds, inter alia, that the processing of off-service users' personal data for direct marketing purposes by a large social network platform may be regarded as carried out for a legitimate interest of the controller, this cannot be done without consent from a user as a legal basis due to the interests and fundamental rights of such a user, which under the circumstances of that case, in particular the extensive processing, override the interest of that operator in such personalised advertising through which social platforms finance their activities (see the *Meta Platforms* judgment, paragraphs 115 to 118).

[96] E.g., Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin OJ L 180, 19.7.2000, p. 22–26; Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation OJ L 303, 2.12.2000, p. 16–22; Directive 2006/54/EC of the European Parliament and of the Council of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (recast), OJ L 204, 26.7.2006, p. 23–36; Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services, OJ L 373, 21.12.2004, p. 37–43.

dark patterns within the user interface to ensure that providers of online platforms do not mislead or coerce users into actions that may not align with their genuine intentions. Such dark patterns should be understood to constitute an example of manipulative or deceptive techniques within the meaning of Article 5(1)(a) AI Act, when they are likely to cause significant harms.

(140) The DSA also sets out obligations for providers of online platforms to ensure transparency in advertising (Articles 26 and 38 for very large online platforms or very large search engines), on the use of recommender systems (Article 27) and on the protection of minors (Article 28 DSA). Moreover, if an online platform or search engine is classified as a very large online platform or very large search engine, the provider of that designated service has additional obligations to assess and mitigate systemic risks stemming from the design or functioning of its service and its related systems, including algorithmic systems (Articles 34 and 35 DSA). When conducting risks assessments, providers of very large online platforms and of very large online search engines should consider how their recommender systems, advertising, content moderation and any other relevant algorithmic systems influence such systemic risks. Such risk assessments should also analyse how systemic risks are influenced by, among other things, the intentional manipulation and automated exploitation of the service (c.f. Article 34(2) DSA and recital 83 DSA). Nevertheless, the scope of Article 5(1)(a) or (b) AI Act covers a broad variety of other scenarios (e.g., chatbots, AI-enabled services and products) that may be offered or used by other actors than providers of intermediary services.

(141) The prohibition of manipulative AI techniques pursuant to Article 5(1)(a) AI Act also supports the objectives of Directive 2010/13/EU (the AVMSD)[97] by preventing harmful AI-driven advertisements[98] and other AI-enabled manipulative and exploitative practices that may be significantly harmful in the media sector.

(142) The AI Act also complements Regulation (EU) 2024/900 (the Political advertising regulation)[99] which provides harmonised rules, including transparency and related due diligence obligations, for the provision of political advertising and related services; and on the use of targeting and ad-delivery techniques in the context of online political advertising. This Regulation prohibits profiling based on special categories of personal data in the context of online political advertising and targeting of persons at least one year under the voting age established by national rules. Furthermore, targeting and ad-delivery techniques in the context of online political advertising can only be done if based on personal data collected from the data subjects and with their explicit consent.

[97] Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive ('AVMSD') amended by Directive (EU) 2018/1808) that, inter alia, aim to improve the protection for children and tackle hate speech more effectively.
[98] Article 9 of AVMSD.
[99] Regulation (EU) 2024/900 of the European Parliament and of the Council of 13 March 2024 on the transparency and targeting of political advertising, PE/90/2023/REV/1, OJ L, 2024/900, 20.3.2024.

Additional transparency requirements also apply, i.e. disclosure of political advertisement, describing the use of such techniques and main parameters and additional information on the logic involved, including about the use of AI systems. Targeted political advertising based on processing of personal data in compliance with that Regulation[100] will help to ensure that profiling of voters and targeting and ad-delivery of political ads operate within the boundaries of lawful persuasion.

(143) The AI Act prohibition of harmful exploitative and deceptive AI practices is also complementary to other applicable Union legislation that sets general transparency rules on advertising and consumer protection and due conduct of operators (e.g. Directive 2014/65/EU MIFID, Directive (EU) 2016/97 on Insurance Distribution[101], Directive (EU) 2023/2225 on Consumer Credit Agreements, Directive (EU) 2002/65 Distance Marketing, Directive 2006/114/EC on misleading and comparative advertising and Directive (EU) 2011/83 on consumer rights sets general consumer protection standards). In this regard, the European Insurance and Occupational Pensions Authority (EIOPA) has already issued a supervisory statement on some unfair exploitative practices in relation to differential pricing that could also fall under the scope of the AI Act when enabled by AI systems[102].

(144) The prohibitions in Article 5(1)(a) and (b) AI Act are also without prejudice to and complement EU product safety legislation (e.g., for medical devices, toys, machinery), which plays a crucial role in ensuring the safety of products that integrate AI systems. This entails compliance with ex ante safety requirements for regulated products and their proactive monitoring to ensure that they do not pose safety risks leading to physical and mental harms. The manufacturer of those products embedding AI systems should therefore take into account these prohibitions in their risk assessments and safety mitigating measures to the extent this fits with the logic and the scope of the relevant Union harmonised safety legislation. Union safety legislation is also complementary to the AI Act prohibitions and can also intervene and address safety risks that do not pose significant harm. In particular, Regulation (EU) 2023/988 (the General Product Safety Regulation)[103] acts as a safety net and requires all consumer products not covered by specific requirements in other sectoral Union product safety legislation (including products embedding AI systems not classified as high-risk pursuant to Article 6 and subjected to the requirements in the AI Act) to be safe under normal or reasonably foreseeable conditions of use, in particular addressing risks to physical and mental health risks for consumers.

[100] Once applicable as of October 2025.
[101] Directive (EU) 2016/97 of the European Parliament and of the Council of 20 January 2016 on insurance distribution (recast), OJ L 26, 2.2.2016, p. 19–59. E.g., Article 17(1) of the Insurance Distribution Directive for insurance distributors to act honestly, fairly and professionally in accordance with the best interests of their customers.
[102] https://www.eiopa.europa.eu/document/download/1e9a8fb2-e688-4bf5-a347-ee0a1ec3aab3_en?filename=EIOPA-BoS-23-076-Supervisory-Statement-on-differential-pricing-practices_0.pdf.
[103] Regulation (EU) 2023/988 of the European Parliament and of the Council of 10 May 2023 on general product safety, amending Regulation (EU) No 1025/2012 of the European Parliament and of the Council and Directive (EU) 2020/1828 of the European Parliament and the Council, and repealing Directive 2001/95/EC of the European Parliament and of the Council and Council Directive 87/357/EEC (Text with EEA relevance).

(145) Finally, the interplay with criminal law is critical. The prohibitions in Article 5(1)(a) and (b) AI Act aim to prevent harmful behaviour that may constitute or lead to criminal offences, such as fraud, forgery, scams, coercion, or the generation and dissemination of illegal content, such as terrorist content, child sexual abuse material, hate speech and sexually explicit deepfakes.[104] Importantly, as internal market legislation, the prohibitions in Article 5(1)(a) and (b) AI Act cover not only the use, but also the placing on the market of the AI system, thus preventing harm early on by limiting access to such prohibited systems that can facilitate and obscure criminal activities. Furthermore, the prohibitions in Article 5(1)(a) and (b) AI Act could also cover other harmful practices that are not qualified as criminal offences under Union or national law.

4. ARTICLE 5(1)(C) AI ACT - SOCIAL SCORING

(146) While AI-enabled scoring can bring benefits to steer good behaviour, improve safety, efficiency or quality of services, there are certain 'social scoring' practices that treat or harm people unfairly and amount to social control and surveillance. The prohibition in Article 5(1)(c) AI Act targets such unacceptable AI-enabled 'social scoring' practices that assess or classify individuals or groups based on their social behaviour or personal characteristics and lead to detrimental or unfavourable treatment, in particular where the data comes from multiple unrelated social contexts or the treatment is disproportionate to the gravity of the social behaviour. The 'social scoring' prohibition has a broad scope of application in both public and private contexts and is not limited to a specific sector or field[105].

(147) At the same time, the prohibition is not intended to affect lawful practices that evaluate people for specific purposes that are legitimate and in compliance with Union and national law,[106] in particular where those laws specify the types of data relevant for the specific evaluation purposes and ensure that any resulting detrimental or unfavourable treatment of persons is justified and proportionate (see section 4.3. out of scope).

4.1. Rationale and objectives

(148) AI systems enabling 'social scoring' practices may lead to discriminatory and unfair outcomes for certain individuals and groups, including their exclusion from society, as well as social control and surveillance practices that are incompatible with Union values. The prohibition of 'social scoring' aims to protect, in particular, the right to human dignity and other fundamental rights, including the right to non-discrimination and equality, to data protection, and to private and family life, as well as relevant social and economic rights, as applicable. It also aims to safeguard and promote the Union

[104] Directive (EU) 2024/1385 of the European Parliament and of the Council of 14 May 2024 on combating violence against women and domestic violence, PE/33/2024/REV/1, OJ L, 2024/1385, 24.5.2024.
[105] The prohibition of social scoring differs from the prohibition in Article 5(1)(d) AI Act, which is more specialised in relation to the evaluation/scoring practice which is applicable only to the risk assessment and prediction of the likelihood of a person committing criminal offences by prohibiting AI systems solely based on profiling or assessment of personality traits and characteristics (see section 5).
[106] Recital 31 AI Act.

values of democracy, equality (including equal access to public and private services), and justice.[107]

4.2. Main concepts and components of the 'social scoring' prohibition

> ***Article 5(1)(c) AI Act provides:***
>
> The following AI practices shall be prohibited:
>
> (c) the placing on the market, the putting into service or the use of AI systems for the evaluation or classification of natural persons or groups of persons 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 groups of persons 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 of persons that is unjustified or disproportionate to their social behaviour or its gravity;

(149) Several cumulative conditions must be fulfilled for the prohibition in Article 5(1)(c) AI Act to apply:

(i) The practice must constitute the 'placing on the market', the 'putting into service' or the 'use' of an AI system.

(ii) The AI system must be intended or used for the evaluation or classification of natural persons or groups of persons over a certain period of time based on:
   (a) their social behaviour; or
   (b) known, inferred or predicted personal or personality characteristics.

(iii) The social score created with the assistance of the AI system must lead or be capable of leading to the detrimental or unfavourable treatment of persons or groups in one or more of the following scenarios:
   (a) in social contexts unrelated to those in which the data was originally generated or collected; and/or
   (b) treatment that is unjustified or disproportionate to their social behaviour or its gravity.

(150) For the prohibition in Article 5(1)(c) AI Act to apply, all three conditions must be simultaneously fulfilled. The first condition, i.e. the placing on the market, the putting into service or the use of the AI system, has been already analysed in section 2.3. The prohibition therefore applies to both providers and deployers of AI systems, each within their respective responsibilities, not to place on the market, put into service or use such AI systems. The remaining criteria for the prohibition on 'social scoring' are further described and analysed below.

[107] Recital 31 AI Act.

4.2.1. 'Social scoring': evaluation or classification based on social behaviour or personal or personality characteristics over a certain period of time

a) Evaluation or classification of natural persons or group of persons

(151) The second condition for the prohibition in Article 5(1)(c) AI Act to apply is that the AI system is intended or used for the **evaluation or classification** of natural persons or groups of persons and assigns them scores based on their social behaviour or their personal or personality characteristics. The score produced by the system may take various forms, such as a mathematical number (for example, from 0 to 1), a ranking, or a label.

(152) The scope of the prohibition is broad covering evaluation and classification practices in both the public and the private sector (see section 4.2.3.). At the same time the evaluation or classification concerns only natural persons or groups of natural persons, thereby excluding in principle legal entities (see section 4.3. out of scope).

(153) While **'evaluation'** suggests the involvement of some form of an **assessment or judgement** about a person or group of persons, a simple **classification** of persons or groups of persons based on characteristics, such as their age, sex, and height, need not necessarily lead to evaluation[108]. The scope of 'classification' is therefore broader than 'evaluation' and can also cover other types of classifications or categorisations of natural persons or groups of persons based on criteria that do not necessarily involve a particular assessment or judgement about those persons or groups of persons and their characteristics or behaviour.

(154) The term '**evaluation'** also relates to the concept of 'profiling', which is regulated by Union data protection legislation[109] and constitutes a specific form of evaluation. While no direct reference is made in Article 5(1)(c) AI Act to that concept or that legislation,[110] they may also be relevant for the prohibition contained in that provision, as well as for other prohibitions in the AI Act,[111] when the evaluation occurs in an automated fashion by an AI system based on personal data. **Profiling** means the use of information about an individual (or group of individuals) and evaluating their characteristics or behaviour patterns in order to place them into a certain category or group, in particular **to analyse and/or make predictions** about, for example, their ability to perform a task; interests; or likely behaviour'.[112] Profiling of natural persons under EU data protection law, when conducted through AI systems, may therefore also be covered by Article 5(1)(c) AI Act.

[108] Article 29 Working Party, *Guidelines on Automated individual decision making and Profiling for the purposes of Regulation 2016/679,* WP251rev.01, 6.2.2018, p. 7.
[109] See Article 4(4) and Article 22 GDPR and Article 11 LED. See also Article 29 Working Party, *Guidelines on Automated individual decision making and Profiling for the purposes of Regulation 2016/679,* WP251rev.01, 6.2.2018, p. 7.
[110] Art 3(52) AI Act does contain a definition of 'profiling' which cross-refers to the definition in Art 4(4) GDPR.
[111] In particular, the prohibition of individual crime risk prediction in Article 5(1)(d) AI Act, which does refer to 'profiling', and in certain instances emotion recognition and biometric categorisation in Article 5(1)(f) and (g) AI Act.
[112] Article 29 Working Party, *Guidelines on Automated individual decision making and Profiling for the purposes of Regulation 2016/679,* WP251rev.01, 6.2.2018, p. 7.

For example, in the *SCHUFA I* judgment, the CJEU examined a **creditworthiness scoring system** used in Germany.[113] In that case, the 'score' generated by the computer programme was a 'probability value' concerning the ability of a person to meet payment commitments, which was qualified by the CJEU as 'profiling'. More specifically, the system established 'a prognosis on the probability of a future behaviour of a person ('score'), such as the repayment of a loan, based on certain characteristics of that person. The establishment of scores ('scoring') is based on the assumption that, by assigning a person to a group of other persons with comparable characteristics who have behaved in a certain way, similar behaviour can be predicted.[114] According to the CJEU, this activity met the definition of "profiling" within the meaning of Article 4(4) GDPR.[115] This form of profiling may also be considered to constitute an evaluation of persons based on their personal characteristics within the meaning of Article 5(1)(c) AI Act, which will be prohibited if done with AI systems and provided that the other conditions for the application of that provision are fulfilled.

b) Over a certain period of time

(155) The prohibition in Article 5(1)(c) AI Act requires that the evaluation or classification is based on data that spans over '**a certain period of time'**. This suggests that the assessment should not be limited to a one-time or an at once rating or grading with data or behaviour from a very specific individual context. At the same time, it is important that this condition is assessed, taking all circumstances of the case into account to avoid circumvention of the scope of the prohibition.

For example, an authority for migration and asylum implements a partly automated surveillance system at refugee camps built on a range of surveillance infrastructure, including cameras and motion sensors. If the analysed data spans a period of time and specific individuals are evaluated (such as migrants) for example to ascertain whether they are at risk of trying to abscond, then this would qualify as 'over a certain period of time' and the prohibition in Article 5(1)(c) AI Act may apply if all other conditions are fulfilled.

c) Based on their social behaviour or known, inferred or predicted personal or personality characteristics

(156) The practices of 'evaluation' and 'classification' prohibited under Article 5(1)(c) AI Act must be based on the AI-enabled processing of data (often extensive) in relation to either i) the social behaviour of individuals or groups of persons or ii) their known, inferred or predicted personal and personality characteristics, or both. The data may be

[113] Judgment of the Court of Justice of 7 December 2023, *SCHUFA Holding (Scoring)*, C-634/21, EU:C:2023:957 (hereinafter referred to as the '*SCHUFA I* judgment') , e.g. paragraph 47.

[114] *Ibidem,* paragraph 14 (own emphasis).

[115] *Ibidem,* paragraph 47.

directly provided by the persons or indirectly collected, i.e. through surveillance, obtained from third parties or through inferences from other information.

(157) Regarding the first scenario, '**social behaviour'** is a broad term that can generally include actions, behaviour, habits, interactions within society, etc., and usually covers behaviour related data points from multiple sources[116]. This could include behaviour of individuals and groups of individuals in social and private contexts, such as participation in cultural events, volunteering, etc., but also social behaviour in business contexts, for example the payment of debts, behaviour when using certain services, as well as relations with public and private entities, government, police, and the law (for example, whether a person obeys traffic rules). Social behaviour data from multiple contexts and data points may be collected in a centralised way by the same entity, but is most often collected in a distributed way and combined from different sources, which may involve increased monitoring and the tracking of individuals (so called 'dataveillance').

(158) The second scenario is where the scoring is based on **personal or personality characteristics**, which may or may not involve specific social behavioural aspects. 'Personal characteristics' may include a variety of information relating to a person, for example sex, sexual orientation or sexual characteristics, gender, gender identity, race, ethnicity, family situation, address, income, household members, profession, employment or other legal status, performance at work, economic situation, financial liquidity, health, personal preferences, interests, reliability, behaviour, location or movement, level of debt, type of car etc.[117] 'Personality characteristics' should be in principle interpreted as synonymous with personal characteristics, but may also imply the creation of specific profiles of individuals as personalities. Personality characteristics may be also based on a number of factors and imply a judgement, which may be made by the individuals themselves, other persons, or generated by AI systems. In the AI Act, personality characteristics are sometimes referred to as personality traits and characteristics;[118] those concepts should be interpreted consistently.

(159) 'Known, inferred or predicted' personal or personality characteristics are different types of information and personal data that need to be distinguished. '**Known characteristics'** are based on information which has been provided to the AI system as an input, and which is in most cases verifiable information. By contrast, **'inferred characteristics**' are based on information which has been inferred from other information, with the inference usually made by an AI system. '**Predicted characteristics**' are those which are estimated based on patterns with less than 100% accuracy. The concepts of 'inferred' (or derived) data are also used in the context of profiling in Union data protection law and may therefore be a source of inspiration for

[116] See Recital 31 AI Act.
[117] See Recital 42 AI Act which lists some examples of such characteristics.
[118] Article 5(1)(d) AI Act.

interpreting those concepts used in Article 5(1)(c) AI Act.[119] The use of these different types of data may have different implications for the accuracy and the fairness of the scoring practices and therefore may be taken into account, in particular where the processing is opaque or relies on data points whose accuracy is more difficult to be verified.

4.2.2. The social score must lead to detrimental or unfavourable treatment in unrelated social contexts and/or unjustified or disproportionate treatment to the gravity of the social behaviour

a) Causal link between the social score and the treatment

(160) For the prohibition in Article 5(1)(c) AI Act to apply, the social score created by or with the assistance of an AI system must **lead to a detrimental or unfavourable treatment** for the evaluated person or group of persons. In other words, the treatment must be the consequence of the score, and the score the cause of the treatment. Such a plausible causal link may also exist in cases where the harmful consequences have not yet materialised but the AI system is intended to or capable of producing such an adverse outcome. This is particularly relevant given that the prohibited practice in Article 5(1)(c) AI Act also covers the 'placing on the market' of such AI systems.

(161) Article 5(1)(c) AI Act does not require the evaluation or classification performed by the AI system to be the sole cause of the detrimental or unfavourable treatment. It therefore also covers AI-enabled scoring practices that may also be subject to or combined with other human assessments. At the same time, the AI output must play a sufficiently important role in producing the social score. For example, in the case where a public authority deploys an AI system to assess the trustworthiness of persons and combines its output with a human assessment of additional facts, this AI-enabled social scoring practice will fall within the scope of the prohibition only if the AI-generated score plays a sufficiently important role in the final decision, provided the other conditions for detrimental or unfavourable treatment are fulfilled as described below (see section 4.2.2. b).

(162) A score may lead to detrimental or unfavourable treatment even if it is produced by an organisation(s) different from the one that uses the score[120]. For example, a public authority may obtain a score for a natural person's creditworthiness assessment produced by another company specialised in creditworthiness and risk assessments, which are based on information about the individuals and their behaviour from a variety of sources.

[119] See Article 29 Working Party, *Guidelines on Automated individual decision making and Profiling for the purposes of Regulation 2016/679*, WP251rev.01, 6.2.2018, p. 7 et seq.

[120] This interpretation is consistent with the CJEU judgment in the *SCHUFA I* judgment, where the CJEU held in the context of automated decision-making that a 'score' (evaluation constituting profiling) produced by an entity other than the one taking the final decision can constitute an automated decision under Article 22 of the GDPR. See *SCHUFA I* judgment, paragraphs 42 to 51 and 60 to 62.

b) Detrimental or unfavourable treatment in unrelated social contexts and/or unjustified or disproportionate treatment

(163) The final condition for the prohibition in Article 5(1)(c) AI Act to apply is that the use of the social score must result (or be capable of resulting) in detrimental or unfavourable treatment either:

i. in social context(s) unrelated to the contexts in which the data was originally generated or collected, or
ii. unjustified or disproportionate to the social behaviour or its gravity.

(164) These conditions are alternative and may apply also in combination. An analysis on a case-by-case basis is necessary to assess if at least one of them is fulfilled, since many AI-enabled scoring and evaluation practices may not fulfil them and therefore be outside the scope of the prohibition. In particular, this may not be the case where the AI-enabled scoring practices are for a specific legitimate evaluation purpose and comply with applicable Union and national laws that specify the data considered as relevant for the purposes of evaluation and ensure that the detrimental or unfavourable treatment is justified and proportionate to the social behaviour (see section 4.3. out of scope).

(165) '**Unfavourable treatment'** means that as a result from the scoring, the person or group of persons must be treated less favourably compared to others without necessarily requiring a particular harm or damage (for example, in the case of scoring practices where people are singled out for additional inspections in case of fraud suspicious). By contrast, '**detrimental'** treatment requires the person or group of persons to suffer certain harm and detriment from the treatment. Unfavourable or detrimental treatment may also be discriminatory and prohibited under EU non-discrimination law or imply the exclusion of certain individuals or groups[121], but that is not a necessary condition for the prohibition to apply. Article 5(1)(c) AI Act could therefore cover unfair treatment beyond EU non-discrimination law that applies only to certain protected groups (e.g., age, ethnic and racial origin, sex, religion).

Scenario 1: Detrimental or unfavourable treatment in unrelated social contexts

(166) In the first scenario described under Article 5(1)(c)(i) AI Act, the detrimental or unfavourable treatment resulting from the score must take place **in social context(s) unrelated** to the contexts in which the data was originally generated or collected. This implies not only that the persons may be treated in an unfavourable or detrimental manner due to the social score, but also that the data about their social behaviour or their known, inferred or predicted personal or personality characteristics are generated or collected in social contexts unrelated to the one in which the scoring takes place. The data collected or generated from these unrelated contexts must be subsequently used by the AI system for the scoring of the persons without an apparent connection for the

[121] Recital 31 AI Act.

purpose of the evaluation or classification or in a manner that leads to generalised surveillance of the persons or the groups of persons. In most cases, this happens against the reasonable expectations of the persons and in violation of Union data protection law and possibly other applicable rules that specify the types of data and sources considered relevant and necessary for the evaluation or classification. Whether this condition is fulfilled will require a case-by-case assessment, taking into account the purpose of the evaluation and the contexts from which the data has been collected and generated.

Examples of detrimental or unfavourable treatment in unrelated social contexts prohibited under Article 5(1)(c) i) AI Act

- National tax authorities use an AI predictive tool on all taxpayers' tax returns in a country to select tax returns for closer inspection. The AI tool uses relevant variables, such as yearly income, assets (real estate property, cars etc.), data on family members of beneficiaries, but also unrelated data, such as taxpayers' social habits or internet connections, to single out specific individuals for inspections.

- A social welfare agency uses an AI system to estimate the probability of fraud by beneficiaries of household allowances that relies on characteristics collected or inferred from social contexts with no apparent connection or relevance for the assessment of fraud, such as having a spouse of a certain nationality or ethnic origin, having an internet connection, behaviour on social platforms, or performance at the workplace, etc. [122] By contrast, data that is relevant for the allocation of the benefits and lawfully collected could be used to determine the risk of fraud, since public authorities pursue a legitimate aim in verifying if social benefits are correctly allocated.

- A public labour agency uses an AI system to score unemployed individuals based on an interview and an AI-based assessment for determining whether an individual should benefit from state support for employment. That score is based on relevant personal characteristics, such as age and education, but also variables collected or inferred from data and contexts with no apparent connection to the purpose of evaluation, such as marital status, health data for chronic diseases, addiction, etc.[123]

These unacceptable scoring practices may be distinguished from lawful practices that evaluate persons for specific purpose in compliance with Union and national law, in particular when such laws, in compliance with EU law, specify the data considered as relevant and necessary for the purposes of evaluation (see section 4.3. out of scope).

[122] For a comparison of a similar national systems of benefits and the social scoring, see D. Hadwick & S. Lan, 'Lessons to be learned from the Dutch childcare allowance scandal: A comparative review of algorithmic governance by tax administrations in the Netherlands, France and Germany' (2021) World Tax Journal, Vol. 13, Issue 4.Familiales (CNAF).

[123] A similar system was used in Poland for a system 'Profiling the Unemployed', which was abandoned after it was deemed unconstitutional. See Szymielewicz, *Profiling the unemployed in Poland.: Social and Political Implications of Algorithmic Decision Making*, Fundacja Panoptykon, 2015, p. 18.

Scenario 2: Unfavourable or detrimental treatment disproportionate to the social behaviour

(167) Another alternative scenario under Article 5(1)(c)(ii) AI Act where an AI scoring system may be prohibited is if the treatment resulting from the score is unjustified or disproportionate to the gravity of the social behaviour. The severity of the impact and the interference with the fundamental rights of the person concerned resulting from the social scoring compared to the gravity of the social behaviour of the person should determine whether such treatment is disproportionate for the legitimate aim pursued, taking into account the general principle of proportionality. This requires a case-by-case assessment, which should consider all relevant circumstances of the case, as well as general ethical considerations and principles for fairness and social justice related to the assessment of the social behaviour and the proportionality of the detrimental treatment. The treatment may also be 'unjustified', such as lacking a legitimate aim. Sectoral Union or national legislation setting specific criteria and procedures that regulate such potential detrimental or unfavourable treatment may also be relevant as part of this assessment.

Examples of unjustified or disproportionate treatment compared to the social behaviour prohibited under Article 5(1)(c) ii) AI Act

- A public agency uses an AI system to profile families for early detection of children at risk based on criteria such as parental mental health and unemployment, but also information on parents' social behaviour derived from multiple contexts. Based on the resulting score, families are singled out for inspection and children considered 'at risk' are taken from their families, including in cases of minor transgressions by the parents, such as occasionally missing doctors' appointments or receiving traffic fines.

- A municipality uses an AI system to score trustworthiness of residents based on multiple data points related to their social behaviour in a variety of contexts. The generated score for residents considered 'less trustworthy' is used for blacklisting, i.e. withdrawal of public benefits, other serious punitive measures, and increased control or surveillance. Among the factors considered in the assessment are insufficient volunteering and minor misbehaviour, such as not returning books to the library on time, leaving rubbish on the street outside the day of collection, and a delay in the payment of local taxes.

These unacceptable social scoring practices may be distinguished from lawful practices that evaluate persons for a legitimate specific purpose in compliance with Union and national law, in particular where those laws ensure that detrimental or unfavourable treatment is justified and proportionate to the social behaviour (see section 4.3. out of scope).

(168) Both alternatives under Article 5(1)(c)(i) and (ii) AI Act may also be fulfilled simultaneously.

Examples of unjustified or disproportionate treatment under Article 5(1)(c)) i) and ii) AI Act

- A tax authority uses an AI system to detect child benefit fraud by profiling and assigning beneficiaries suspected of fraud to categories such as 'deliberate intent/gross negligence' using criteria such as low income, dual nationality, social behaviour, etc. Based on the risk score, a beneficiary's file is inspected and, in many cases, their childcare benefit ceased, they receive notice to repay the received benefits, and no longer qualify for standard debt collection arrangements. Such scoring causes many families to be heavily indebted and leads to unjust, discriminatory and detrimental treatment of individuals and groups of individuals[124], driving many families into severe financial hardship.

- A public authority uses an AI system to control fraud in the student housing grant process that considers among the indicators the internet connections, family status or level of education of beneficiaries as distinguishing factors for fraud risk, which do not seem relevant, nor justified.

- A government introduces a comprehensive AI-based system that monitors and rates citizens based on their behaviour in various aspects of life, such as social interactions, online activities, purchasing habits, and punctuality in paying bills. People with a lower score face restricted access to public services, higher interest rates on loans, and difficulty in traveling, renting apartments, and even finding jobs. The system leads to excessive surveillance of individuals and detrimental treatment in contexts unrelated to the social behaviour used to determine the social score (e.g. job opportunities are influenced by social media activity), while also imposing excessive penalties for minor infractions (e.g. social and financial disadvantages for relatively minor offences).

These unacceptable social scoring practices may be distinguished from lawful practices evaluating persons for legitimate specific purposes that do not fulfil these conditions and are in compliance with Union and national law, in particular when those laws ensure that detrimental or unfavourable treatment is justified and proportionate and data from related social contexts is used (see section 4.3. out of scope).

(169) The prohibition under Article 5(1)(c) AI Act may also cover cases where awards or preferential treatment are given to certain individuals or groups of persons, since this implies less favourable treatment of other individuals (e.g. in cases of support employment programmes, (de-)prioritisation for housing or resettlement).

4.2.3. Regardless of whether provided or used by public or private persons

[124] For a similar example of the Dutch Childcare Benefits scandal in the Netherlands, see *Belastingdienst treft 232 gezinnen met onevenredig harde actie*, 27.11.2019, (in Dutch). A Dutch court decided in 2020 that 'Systeem Risico Indicatie (SyRi) was unlawful. See also. Geen powerplay maar fair play. Onevenredig harde aanpak van 232 gezinnen met kinderopvangtoeslag, 2017, p. 32.

(170) As already noted, Article 5(1)(c) AI Act prohibits unacceptable AI-enabled social scoring practices regardless of whether the AI system or the score are provided or used by public or private persons. While scoring in the public sector may have very significant consequences for people due to an imbalance of power and a dependence on public services, similarly harmful consequences may also occur in the private sector, where scoring practices are also increasingly undertaken by companies and other entities.

For example,

- An insurance company collects spending and other financial information from a bank which is unrelated to the determination of eligibility of candidates for life insurance and which is used to determine the price of the premium to be paid for such insurance. An AI system analyses this information and recommends, on that basis, whether to refuse a contract or set higher life insurance premiums for a particular individual or a group of customers.
- A private credit agency uses an AI system to determine the creditworthiness of people and decide whether an individual should obtain a loan for housing based on unrelated personal characteristics.

These unacceptable social scoring practices may be distinguished from lawful practices evaluating persons for specific legitimate purposes that do not fulfil these conditions and are in compliance with Union and national law, in particular when those laws ensure that detrimental or unfavourable treatment is justified and proportionate and data from related social contexts is used (see section 4.3. out of scope).

(171) In the case of checks by competent market surveillance authorities, it is on the provider and the deployer of the AI system, each within their responsibilities, to demonstrate that the AI practice is legitimate and justified, including by providing transparency of the functioning of the AI system, and information about the types of data and data sources, ensuring that only data related to the social context in which the score is used are processed for the purpose of the evaluation or classification and those data were lawfully collected, the system is performing as intended, and any resulting detrimental or unfavourable treatment is justified and proportionate to the social behaviour. Compliance with applicable legislation and appropriate and proportionate safeguards built in the system and applied during its operation will help to avoid the prohibition from applying, while enabling the use of AI systems for the evaluation or classification of persons for legitimate and beneficial purposes (e.g. to improve the effectiveness of processes, quality of service, safety, etc.) (see section 4.3. out of scope).

(172) Compliance with the requirements for high-risk AI systems (e.g., in the area of essential public services and benefits, credit-scoring and creditworthiness assessment, migration etc.) may also help to ensure that AI systems used for evaluation and classification purposes in those high-risk areas do not constitute unacceptable social scoring practices

that providers and deployers should consider when implementing their respective obligations (e.g., on risk management, transparency, data governance, fundamental rights impact assessment, human oversight, monitoring, etc.).

4.3. Out of scope

(173) The prohibition in Article 5(1)(c) AI Act only applies to the scoring of natural persons or groups of persons, thus excluding in principle scoring of legal entities where the evaluation is not based on personal or personality characteristics or social behaviour of individuals, even if in some cases individuals may be indirectly impacted by the score (e.g. all citizens in a municipality in case of allocation of budget). However, if legal entities have been evaluated based on an overall score that aggregates the evaluation or classification of a group of natural persons based on their social behaviour or personal or personality characteristics and this score directly affects those persons (e.g., all employees in a company, students in a specific school whose behaviour has been evaluated), the practice may fall within the scope of Article 5(1)(c) AI Act if all other conditions are fulfilled. This will depend on a case-by-case assessment.

(174) AI-based social scoring as a 'probabilistic value' and prognosis should also be distinguished from individual ratings by users which assess the quality of a service (such as a driver in an online car-sharing platform or a host in an online platform for accommodation). Such ratings are the mere aggregation of individual human scores that do not necessarily involve AI, unless the data are combined with other information and analysed by the AI system for evaluating or classifying individuals fulfilling all conditions in Article 5(1)(c) AI Act.

(175) Furthermore, the scoring of natural persons is not at all times prohibited, but only in the limited cases where all of the conditions of Article 5(1)(c) AI Act are cumulatively fulfilled, as analysed above. Recital 31 AI Act, in particular mentions that the prohibition 'should not affect lawful evaluation practices of natural persons that are carried out for a specific purpose in accordance with Union and national law'. For example, credit scoring and risk scoring and underwriting are essential aspects of the services of financial and insurance businesses. Such practices, as well as other legitimate practices (i.e. to improve the quality and efficiency of services, to ensure more efficient claims handling, to perform specific employee evaluations, fraud prevention and detection, law enforcement or scoring of users' behaviour on online platforms), are not *per se* prohibited, if lawful and undertaken in line with the AI Act and other applicable Union law and national law, which must comply with Union law.

(176) In other words, AI systems which evaluate or classify individuals for the purposes of generating a social score in a lawful manner and for a specific purpose in the related context as that in which the personal data used for the score were collected are not

prohibited, provided that any detrimental or unfavourable treatment from using the score is justified and proportionate to the gravity of the social behaviour.[125]

(177) Compliance with sectoral Union legislation, such as in the field of credit-scoring, anti-money laundering, etc., that specifies the type of data that can be used as relevant and necessary for the specific legitimate purpose of evaluation and ensures that the treatment is justified and proportionate to the social behaviour may thus ensure that the AI practice falls outside the scope of the prohibition in Article 5(1)(c) AI Act.

Examples of legitimate scoring practices in line with Union and national law that are outside the scope of Article 5(1)(c) AI Act:

- Financial credit scoring systems used by creditors or credit information agencies to assess a customer's financial creditworthiness or outstanding debts, providing a credit score or determining their creditworthiness assessment, which are based on the customer's income and expenses and other financial and economic circumstances, are out of scope of Article 5(1)(c) AI Act if they are relevant for the legitimate purpose of the credit scoring and if they comply with consumer protection laws[126] specifying the type of data and the necessary safeguards to ensure the fair treatment of consumers in creditworthiness assessments.

- Companies have a legitimate interest to evaluate customers for financial fraud and those practices are not affected by the prohibition, if the evaluation is based on relevant data such as transactional behaviour and metadata in the context of the services, past history and other factors from sources that are objectively relevant to determine the risk of fraud and if the detrimental treatment is justified and proportionate as a consequence of the fraudulent behaviour.

- Information collected through telematic devices that show that a driver is speeding or not maintaining safe driving practices used by an insurer that offers telematics-based tariffs in relation to a policyholder's high-risk driving behaviour may be used to increase the premium of that policyholder due to the higher risk of an accident caused by that driving behaviour, provided the increase in the premium is proportionate to the risky behaviour of the driver.

- The collection and processing of data that is relevant and necessary for the intended legitimate purpose of the AI systems (e.g., health and schizophrenic data collected from various sources to diagnose patients) is out of scope of Article 5(1)(c) AI Act, in particular because it process relevant and necessary data and typically does not entail unjustified detrimental or unfavourable treatment of certain natural persons.

- Online platforms profiling users for safety reasons on their services based on data which is relevant for the context and purpose of assessment is out of scope of Article

[125] Recital 31 AI Act.
[126] See in particular Directive (EU) 2023/2225 of 18 October 2023 on credit agreements for consumers and repealing Directive 2008/48/EC and the European Banking Authority' Guidelines on loan origination and monitoring from 29 May 2020, EBA/GL/2020/06.

5(1)(c) AI Act, when the evaluation does not result in detrimental treatment that is disproportionate to the gravity of the user's misbehaviour.

- AI-enabled targeted commercial advertising is out of scope, where it is based on relevant data (e.g. users' preferences), it is done in line with Union law on consumer protection, data protection and digital services, and it does not result in detrimental or unfavourable treatment disproportionate to the gravity of the user's social behaviour (e.g. exploitative and unfair differential pricing).

- AI systems using data collected in refugee camps (e.g., behavioural compliance) for decisions about resettlement or employment is not affected by the prohibition, given that this data is relevant for the purpose of assessment and provided that the procedures under applicable Union migration law are fulfilled to ensure the treatment is justified and proportionate.

- AI-enabled scoring by an online shopping platform which offers privileges to users with a strong purchase history and a low rate of product returns, such as a faster returns application process or returnless refunds are out of scope of Article 5(1)(c) AI Act, given that the advantages are justified and proportionate to reward positive behaviour and other users continue to have access to the standard return process.

- AI-evaluation and scoring of individuals by police and other law enforcement authorities that collect data about individuals' social behaviour from multiple contexts are out of scope of Article 5(1)(c) AI Act where those data are relevant for the specific purposes of the prevention, detection, prosecution and punishment of criminal offences, and where the detrimental treatment is justified and proportionate in accordance with substantive and procedural Union and national criminal and police law. It is also relevant to consider in this context the prohibition in Article 5(1)(d) AI Act, which imposes additional and more specific conditions for AI-enabled risk assessments and predictions of the likelihood of a person committing a criminal offence which must not be solely based on profiling or the assessment of personality traits (see section 5).

4.4. Interplay with other Union legal acts

(178) Providers and deployers should carefully assess whether other applicable Union and national legislation applies to any particular AI scoring system used in their activities, in particular if there is more specific legislation that strictly regulates the types of data that can be used as relevant and necessary for specific evaluation purposes and if there are more specific rules and procedures to ensure justified and fair treatment.

(179) AI-enabled social scoring practices by private parties acting as traders in business-to-consumer relations may also be in breach of Union consumer protection law, i.e., Directive 2005/29/EC on unfair business-to-consumer commercial practices (the 'UCPD'). The UCPD prohibits commercial practices if they are contrary to the

requirements of professional diligence and materially distort or are likely to materially distort the economic behaviour of the average consumer or average member of the group with regard to the product (Article 5 UCPD). The scoring practices may also be found misleading (Articles 6-7 UCPD) subject to case-by-case assessment of the impact of the commercial practice on the consumer's transactional decision.

(180) Social scoring, whether by public or by private parties, may also be in breach of Union data protection laws, for example as regards the legal ground for processing (lawfulness), the data protection principles (e.g. data minimisation and necessity, fairness, transparency), and any other obligations, including the rules on solely automated individual decision-making, where relevant.

(181) Where the evaluation or classification is based on one of the grounds protected from discrimination (e.g., age, religion, racial or ethnic origin, sex etc.) or results directly or indirectly in discrimination of those groups, such a practice will also be subject to Union non-discrimination law.

(182) The Consumer Credit Directive (EU) 2023/2225[127] may also be relevant in this context. Article 18(3) CCD requires that the assessment of creditworthiness is carried out based on relevant and accurate information on the consumer's income and expenses and other financial and economic circumstances which is necessary and proportionate to the nature, duration, value and risks of the credit for the consumer. That information may include evidence of income or other sources of repayment, information on financial assets and liabilities, or information on other financial commitments. The CCD explicitly prohibits special categories of personal data to be included in the information and to obtain information from social networks. The European Banking Authority's Guidelines on loan origination and monitoring[128] further specify the relevant information for the purpose of creditworthiness assessments. This specification of the type of data in these sectoral laws for specific evaluation purposes are relevant considerations to be taken into account when determining whether a practice falls within the scope of the prohibition in Article 5(1)(c) AI Act.

(183) Similarly, AI systems used for the evaluation and classification of persons for anti-money laundering and terrorism financing purposes should also comply with relevant Union legislation on these matters.

5. ARTICLE 5(1)(D) AI ACT – INDIVIDUAL RISK ASSESSMENT AND PREDICTION OF CRIMINAL OFFENCES

(184) Article 5(1)(d) AI Act prohibits AI systems assessing or predicting the risk of a natural person committing a criminal offence based solely on profiling or assessing personality traits and characteristics.

[127] Directive 2008/48/EC of the European Parliament and of the Council of 23 April 2008 on credit agreements for consumers and repealing Council Directive 87/102/EEC, *OJ L 133*, 22/05/2008, p. 66–92.

[128] European Banking Authority, Guidelines on loan origination and monitoring from 29 May 2020, EBA/GL/2020/06.

(185) The provision indicates, in its last phrase, that the prohibition does not apply if the AI system is 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 that activity. Such AI systems that fall outside the scope of the prohibition intended to be used by law enforcement authorities, or on their behalf, or by Union institutions, bodies, offices or agencies in support of law enforcement authorities, for assessing the risk of a natural person offending or re-offending not solely on the basis of profiling, or the assessment of personality traits and characteristics or past criminal behaviour are classified as 'high-risk' AI systems (Annex III, point 6, letter (d) AI Act) and must comply with all relevant requirements and obligations under the AI Act.

5.1. Rationale and objectives

(186) Recital 42 AI Act explains the background and rationale of the prohibition in Article 5(1)(d) AI Act, namely, that natural persons should be judged on their actual behaviour and not on AI-predicted behaviour based solely on their profiling, personality traits or characteristics.

5.2. Main concepts and components of the prohibition

> ***Article 5(1)(d) AI Act provides***
>
> The following AI practices shall be prohibited:
>
> d) the placing on the market, the putting into service for this specific purpose, or the use of an AI system for making risk assessments of natural persons in order to assess or predict the risk of a natural person committing 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;

(187) Several cumulative conditions must be fulfilled for the prohibition in Article 5(1)(d) AI Act to apply:

(i) The practice must constitute the 'placing on the market', 'the putting into service for this specific purpose' or the 'use' of an AI system.

(ii) The AI system must make risk assessments that assess or predict the risk of a natural person committing a criminal offence.

(iii) The risk assessment or the prediction must be based solely on either, or both, of the following:
   (a) the profiling of a natural person,
   (b) assessing a natural person's personality traits and characteristics.

(188) For the prohibition to apply all three conditions must be simultaneously fulfilled. The first condition, i.e. the placing on the market, putting into service or use of the AI system, has been already analysed in section 2.3. The prohibition, therefore, applies to

both providers and deployers of AI systems, each within their respective responsibilities not to place on the market, put into service, or use such AI systems for this specific purpose. The other two conditions for the prohibition to apply are analysed below.

5.2.1. Assessing the risk or predicting the likelihood of a person committing a crime

(189) Risk assessments to assess or predict the risk of an individual committing a criminal offence are often referred to as individual 'crime prediction' or 'crime forecasting'. While there is no generally agreed definition of 'crime prediction' or 'crime forecasting'[129], these terms refer in general to a variety of advanced AI technologies and analytical methods applied to large amount of often historical data (including socio-economic data, but also police records, etc.) which, in combination with criminology theories, are used to forecast crime as a basis to inform police and law enforcement strategies and action to combat, control, and prevent crime.[130]

(190) Crime prediction AI systems identify patterns within historical data, associating indicators with the likelihood of a crime occurring, and then generate risk scores as predictive outputs. For example, such systems may be used for planning police task forces, for monitoring high-risk situations, and for conducting controls of persons predicted as likely (re-)offenders. Such systems bring opportunities for law enforcement authorities, especially those with scarce resources, increasing efficiency, and enabling a proactive approach for detecting, deterring, and anticipating criminal offences.[131] However, such use of historical data on crimes committed to predict other persons' future behaviour may perpetuate or even reinforce biases, and may result in crucial individual circumstances being 'overlooked' when these circumstances are not part of the data set or considered in the algorithms on which the particular AI system operates. This may also undermine public trust in law enforcement and the justice system in general[132].

(191) Such risk assessments and predictions are, in principle, forward-looking and concern future criminal offences (not yet committed) or crimes that are assessed as a risk of being committed at the moment, including in cases of an attempt or preparatory activities undertaken to commit a criminal offence.[133] They can be made at any stage of the law enforcement activities, such as during prevention and detection of crimes, but also during the investigation, prosecution and execution of criminal penalties (including

[129] For example, see systems mentioned in the EU Fundamental Rights Agency handbook, such as the Criminality Awareness System (CAS) in the Netherlands and Precobs in Germany and Switzerland, Handbook, 2018, p.138. *Preventing unlawful profiling today and in the future: a guide*, Handbook, 2018, p.138.

[130] See Europol, AI and policing The benefits and challenges of artificial intelligence for law enforcement, An Observatory Report from the Europol Innovation Lab, 23 September 2024. See also F. Yang, 'Predictive Policing' in *Oxford Research Encyclopedia, Criminology and Criminal Justice*, Oxford University Press, 2019.

[131] For example, OxRec (Dutch Probation Office, 'Reclassering Nederland') Prediction of violent reoffending in prisoners and individuals on probation: a Dutch validation study (OxRec) - PMC (nih.gov)

[132] See for instance EU Fundamental Rights Agency (8 December 2022) Bias in algorithms - Artificial intelligence and discrimination | European Union Agency for Fundamental Rights.

133 See in this respect Recital 42 AI Act that refers in this respect to the 'likelihood of their offending' and the 'occurrence of actual or potential criminal offences' which are used in present, but not past tense.

when judicial authorities assess the risk of re-offending e.g. in the context of making decisions on the imposition of pre-trial detention) as well as part of the individuals' plan for re-integration into society after serving a criminal sentence[134].

(192) The prohibition in Article 5(1)(d) AI Act does not outlaw crime prediction and risk assessment practices as such. It only applies to AI systems for making risk assessments to assess or predict the risk of a natural person committing a criminal offence, where also the third condition referred to above is met. Moreover, as noted, the prohibition does not apply in the situations described in the express exclusion contained in the last phrase of Article 5(1)(d) AI Act.

5.2.2. Solely based on profiling of a natural person or on assessing their personality traits and characteristics

(193) The third condition for the prohibition in Article 5(1)(d) AI Act to apply is that the risk assessment to assess or predict the risk of a natural person committing a crime must be based solely on a) the profiling of the person or b) on the assessment of their personality traits and characteristics.

(194) The prohibition in Article 5(1)(d) AI Act applies, irrespective of whether the AI system profiles or assesses the personality traits and characteristics of only one natural person or a group of natural persons simultaneously, since the prohibition aims to protect every individual in respect of whom the risk of committing a criminal offence is being predicted or assessed.

a) Profiling of a natural person

(195) Unlike Article 5(1)(c) AI Act, Article 5(1)(d) explicitly uses the term 'profiling'. Article 3(52) AI Act defines that term by reference to its definition in Article 4(4) GDPR[135]. The concept of profiling includes the objective to 'evaluate certain personal aspects' as one of its core elements.[136] In the context of Article 5(1)(d) AI Act, the profiling is done for the purposes of assessing or predicting the risk of a person committing a crime.

(196) The concept of so-called group profiling[137] may also be relevant in this context. That concept refers to the construction and the application of a descriptive profile for a given group, for example categories of perpetrators of criminal offences (e.g., terrorists, gangsters etc.) constructed on historic data about previously committed crimes by other

[134] As an example, Art. 24(4) of EU Directive 2011/93 on combating the sexual abuse and sexual exploitation of children and child pornography, requires persons undergoing criminal proceedings or convicted of acts linked to child sexual abuse to undergo an assessment of the danger they pose of recidivism.

[135] Article 3(4) LED, which is relevant for the prohibition in Article 5(1)(d) AI Act, defines profiling in an identical manner to that in Article 4(4) GDPR as 'any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements'. The same definition is also contained in Article 3(5) of Regulation (EU) 2018/1725 on the processing of personal data by the Union institutions, bodies, offices and agencies, (OJ L 295, 21.11.2018, p. 39).

[136] See also Article 29 Data Protection Working Party, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, WP251rev.01, 6.2.2018, and endorsed by the EDPB, p. 7. See also Fundamental Right Agency, *Preventing unlawful profiling today and in the future: a guide*, Handbook, 2018, p.138.

[137] See about a group profiling, e.g., Fundamental Right Agency, *Preventing unlawful profiling today and in the future: a guide*, Handbook, 2018, p. 21.

persons. Those group profiles may be used later to assess and predict the risk of other persons committing similar offences. Whenever an AI system makes prediction and applies such a (group) profile to a specific individual, this constitutes profiling of the person and may therefore fall within the prohibition of Article 5(1)(d) AI Act.

b) Assessment of personality traits and characteristics

(197) The prohibition also applies if the risk assessment to assess or predict the risk of the person committing a criminal offence is only based on assessing the person's personality traits and characteristics. Such an assessment or prediction is often included in the concept of profiling, but it could also be seen as an alternative, if profiling as defined in Article 4(4) GDPR cannot be established.

(198) As mentioned in section 4.2.1.c), personality traits and characteristics constitute a broad category of characteristics related to a particular natural person, for which there is no generally agreed taxonomy. Recital 42 AI Act provides examples of personality traits and characteristics which may be assessed for predicting the risk of a person committing an offence, such as 'nationality, place of birth, place of residence, number of children, level of debt or type of car'. This is only an illustrative and not an exhaustive list.

c) ***'Solely'***

(199) Article 5(1)(d) AI Act provides that the risk assessments covered by that provision are only prohibited where they are based 'solely' on the profiling of a person or the assessment of their personality traits and characteristics. It is clear from Recital 42 AI Act that 'solely' is intended to apply both to profiling or to the assessment of personality traits and characteristics.

(200) The condition that the risk assessment must be based 'solely' on profiling or assessing personality traits and characteristics may not be fulfilled in a number of situations.

(201) As is evident from the last phrase of Article 5(1)(d) AI Act, such a situation arises, in any event, where the AI system is 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. As Recital 42 makes clear, in this context one should think, in particular but not necessarily exclusively, of a situation in which a reasonable suspicion in respect of the natural person concerned already exists. After all, in such cases there will normally have been a human assessment, which will normally be based on relevant objective and verifiable facts.

(202) However, there can also be other situations, which will always need to be assessed on a case-by-case basis. On the one hand, the use of the term 'solely' leaves open the possibility of various other elements being taken into account in the risk assessment, which makes that it is no longer based on profiling or assessing personality traits or characteristics alone. On the other hand, in order to avoid circumvention of the prohibition and ensure its effectiveness, any such other elements will have to be real,

substantial and meaningful for them to be able to justify the conclusion that the prohibition does not apply. A reading of the prohibition of Article 5(1)(d) AI Act together with the exclusion contained in the last phrase thereof suggests that, in particular, the existence of certain pre-established objective and verifiable facts may justify that conclusion.

For example,

- A law enforcement authority uses an AI system to predict criminal behaviour for crimes such as terrorism solely based on individuals' age, nationality, address, type of a car, and marital status. With that system, individuals are deemed more likely to commit future offences that they have not yet committed solely based on their personal characteristics. Such a system may be assumed to be prohibited under Article 5(1)(d) AI Act.

- National tax authorities use an AI predictive tool to review all taxpayers' tax returns to predict potential criminal tax offences to identify cases requiring further investigation. This is done solely on the basis of the profile built by the AI system, which uses for its assessment personality traits, such as double nationality, place of birth, number of children, and opaque variables, especially inferred information that is predictive and therefore non-objective and hard to verify. Such a system will normally fall under the prohibition of Article 5(1)(d) AI Act, since there is no reasonable suspicion of the involvement of a particular person in a criminal activity or other objective and verifiable facts linking that to that criminal activity. This is also an example that falls within the scope of social scoring prohibited under Article 5(1)(c) AI Act involving unfavourable treatment with data from unrelated social contexts.

- A police department uses AI-based risk assessment tool to assess the risk of young children and adolescents being involved in 'future violent and property offending'. The system assesses children based on their relationships with other people and their supposed risk levels, meaning that children may be deemed at a higher risk of offending simply by being linked to another individual with a high-risk assessment, such as a sibling or a friend. The parents' risk levels may also impact a child's risk level. The risk assessments result in police 'registering' these children in their systems, monitoring them with additional inspections, and referring them to youth 'care' services. Such a system is also likely to fall under the prohibition of Article 5(1)(d) AI Act.

5.2.3. Exclusion of AI systems to support the human assessment based on objective and verifiable facts directly linked to a criminal activity

(203) Article 5(1)(d) AI Act provides, in its last phrase, that the prohibition does 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. Although, as noted, the situation described in this express exclusion is not necessarily the only one in which the prohibition does not apply, including that situation expressly in that provision offers legal certainty by delineating the scope of the prohibition and by making it clear that, where that situation is at issue, the prohibition does not in any event apply.

(204) Where the system falls within the scope of the exclusion and is therefore not prohibited, it will be classified as a high-risk AI system (as referred to in Annex III, point 6(d), AI Act) if intended to be used by law enforcement authorities or on their behalf and therefore subject to the requirements and safeguards, including human oversight (Article 14 and Article 26 AI Act). These requirements include that the human oversight must be assigned to persons with the necessary competence, training and authority who should be able to properly understand the capabilities and limitations of the AI system, correctly interpret its output and address the risk of automation bias. Those persons should have clear procedures, training and the necessary competence and authority to meaningfully assess the outputs of the AI system. In this specific case, their human assessment should ensure that any AI prediction or assessment of the risk of a person committing a crime is based on objective and verifiable facts linked to a criminal activity. Those persons should also intervene in order to avoid negative consequences or risks, or stop the use of the AI system if it does not perform as intended.

(205) Furthermore, the concept of 'human intervention' has been subject to CJEU case-law, in particular in the context of solely automated decision-making predicting the risk of air passengers being involved in serious crimes. That case-law may also be relevant for the application of the concept of 'human assessment' as used in Article 5(1)(d) AI Act.

In the *Ligue des droits humains* case[138], the CJEU examined the legality of the use of an advanced AI system for the systematic processing of passenger name record (PNR) data of air travellers to assess their likelihood of being involved in terrorism and other serious crimes.

The CJEU interpreted the rule in Directive (EU) 2016/681 ("PNR Directive") that prohibits adverse legal decisions based solely on automated processing and required **individual human assessment** and review for any positive matches by non-automated means to identify false positives and ensure non-discriminatory results.

According to the CJEU, that human assessment**,** subject to which any results of automated processing of PNR data must be based, must rely on objective criteria to evaluate whether a positive match concerns someone who might be involved in this specific case in terrorist offenses or serious crime, and to ensure the non-discriminatory nature of automated processing.

[138] Judgment of the Court of Justice of 21 June 2022, *Ligue des droits humains*, C-817/19, ECLI:EU:C:2022:491.

(206) As to the content of the exclusion, one of its central elements is that the AI system is used to support human assessment, rather than involving the AI system itself making the risk assessment as occurs in the situations covered by the prohibition. However, for the exclusion to apply, that human assessment must, in addition, already be based on objective and verifiable facts directly linked to a criminal activity.

5.2.4. Extent to which private actors' activities may fall within scope

(207) Besides law enforcement authorities that are in principle the main deployers of AI crime predictive systems, the activities of private entities may also be covered by the prohibition in Article 5(1)(d) AI Act in some cases. That follows from the fact that, based on its wording, the prohibition does not apply exclusively to law enforcement authorities. Moreover, otherwise the prohibition might be easily circumvented, which would call into question its effectiveness.

(208) That being so, the prohibition may be assumed to apply, in particular, when private actors are entrusted by law to exercise public authority and public powers for the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties[139]. Private actors may be also explicitly requested on a case-by-case basis to act on behalf of law enforcement authorities and carry out individual crime risk predictions. In those cases, the activities of those private actors could also fall within the scope of the prohibition, if the applicable conditions are fulfilled and the exclusion does not apply.

> For example, a private company providing advanced AI-based crime analytic software may be asked by a law enforcement authority to analyse a large amount of data from multiple sources and databases, such as national registers, banking transactions, communication data, geo-spatial data, etc., to predict or assess the risk of individuals as potential offenders of human trafficking offences. If all the criteria for Article 5(1)(d) are met, such a use case could be prohibited.

(209) Furthermore, the prohibition may apply to private entities assessing or predicting the risk of a person of committing a crime where this is objectively necessary for compliance with a legal obligation to which that private operator is subject to assess or predict the risk of persons committing specific criminal offences (e.g., in case of anti-money laundering, terrorism financing).

> For example, a banking institution has an obligation under Union anti-money laundering legislation[140] to screen and profile customers for money-laundering offences. If the bank uses an AI system to fulfil its obligations, that should be done based only on the data as specified in that law which are objective and verifiable to ensure that the persons singled out as suspect are reasonably likely to commit anti-

[139] See definition of law enforcement authorities in Article 3(45) AI Act.
[140] Anti-Money Laundering Regulation (EU) 2024/1624 of 31 May 2024.

money laundering offences. The predictions must also be subject to human assessment and verification in compliance with that legislation[141] in order to ensure the accuracy and appropriateness of such assessments. Compliance with that legislation will ensure that the use of individual crime prediction AI system for anti-money laundering purposes fall outside the scope of the prohibition in Article 5(1)(d) AI Act.

(210) However, having regard to the focus on risk assessments relating specifically and exclusively to the commission of criminal offences that is evident from the wording of the prohibition as well as to the purpose of the prohibition as explained in Recital 42, if a private entity profiles customers for its regular business operations and safety or to protect its financial interests (e.g. detecting financial irregularities) without the purpose of assessing or predicting the risk of the customer committing a specific criminal offence, the activities of the private entities should not be considered to fall under the scope of the prohibition of Article 5(1)(d) AI Act.

(211) In other words, in the absence of private parties having been entrusted by law certain specific law enforcement tasks, acting on behalf of law enforcement authorities or being subject to specific legal obligations as described above, the use of AI systems for making risk assessments in the context of private entities' ordinary course of business and with the aim of protecting their own private interests, whilst the fact that those risk assessments may relate to the risk of criminal offences being committed merely as a purely accidental and secondary circumstance, is not deemed to be covered by the prohibition.

5.3. Out of scope

5.3.1. Location-based or geospatial predictive or place-based crime predictions

(212) Location-based or geospatial or place-based crime predictions is based on the place or location of crime or the likelihood that in those areas a crime would be committed. In principle, such policing does not involve an assessment of a specific individual. They therefore fall outside the scope of the prohibition.

Examples of location-based or geospatial predictive or place-based crime predictions

- AI-based predictive policing system provides a score of the likelihood of criminality in different areas in a city based on previous criminality rates by area and other supporting information such as street maps, to highlight elevated risk of specific types of criminalities e.g. burglaries, knife crime etc. and help law enforcement authorities determine where to deploy less or more police patrols/presence to carry out community policing to disrupt and deter criminal activity.

[141] Article 20 of Regulation (EU) 2024/1624.

- A customs authority uses AI risk analytic tools to predict the likelihood of the location of narcotics or illicit goods, for example on the basis of known trafficking routes.

- A police department uses AI-driven systems to detect and locate gunshots in real time. The system employs acoustic sensors in urban areas to identify gunfire sounds and triangulate their location, providing officers with actionable data to aid detection and investigation of crimes.

(213) It may, however, not always be evident how to distinguish location-based crime predictive systems from individual predictive systems assessing the risk of a person committing a crime. To the extent that an AI system carries out location-based predictive policing and then considers the risk score of the location as an aspect in the profiling of a person, that system should be considered person-based and in principle covered by Article 5(1)(d) AI Act, although it may fall outside the scope of the prohibition on other grounds.

For example, if location-based or geospatial information or place-based information is linked to information relating to an individual (e.g., the place of residences of a given person) and the AI system would assess the risk that that person is likely to commit crime based solely on profiling of the individual concerned, including based on his or her residence where crime is high, that system should be considered person-based.

5.3.2. AI systems that support human assessments based on objective and verifiable facts linked to a criminal activity

(214) Article 5(1)(d) AI Act provides that the prohibition does not apply to AI systems used to support the human assessment of the involvement of a natural person in a criminal activity, which is already based on objective and verifiable facts directly linked to a criminal activity. In such a case, the individual crime risk assessments and predictions would also not be based *solely* on profiling or the assessment of personal characteristics and would therefore not be prohibited.

Examples of AI systems falling outside the scope of the prohibition for this reason include:

- The use of an AI system for profiling and categorisation of actual behaviour, such as reasonably suspicious dangerous behaviour in a crowd that someone is preparing and is likely to commit a crime, and there is a meaningful human assessment of the AI classification. In this case, the risk assessment made by the human with the support of AI is not solely based on the personal traits or the profiling, but on objective and verifiable facts linked to the threatening criminal behaviour of that person that has been reviewed by a human before action taken.

- The police is investigating the risk of a possible armed robbery and suspects two individuals. Several verifiable and objective facts are present on which that suspicion is based, such as verifiable participation and dialogues in dark web chat groups for

purchasing arms. An AI system combining geospatial predictive or place-based policing information and Automated Number Plate Registration (ANPR) information of vehicles belonging to the suspects supports the human assessment in the investigation based on verifiable and objective facts directly linked to a specific criminal activity.

- The use of an AI system that assesses the risk whether a prisoner should receive the benefit of an early release. The AI profile of the affected person or assessment of their personality traits and characteristics only support the human assessment of objective and verifiable facts related to past criminal offences and demonstrated behaviour relevant to rehabilitation.

- A judge conducts a pre-trial detention hearing for a person accused of a serious criminal offence to assess whether non-custodial measures can be applied. The decision is based on an assessment of the existence of valid grounds to imposed pre-trial detention, such as the likelihood that the suspect or accused person will commit another offence if not detained or that they will abscond or obstruct the proper conduct of the investigation. To assist in this process, the judge uses an AI risk assessment tool trained on data including past criminal history of individuals in similar cases, as well as factors such as age group, social behaviour, income, and employment status.

- AI system is used to support the assessment of a human officer to assess the risk of an individual serving a non-custodial sentence violating release conditions or absconding based on past criminal behaviours and objective facts that give grounds for suspicion such as adherence to conditions of release, psychological assessment outcomes and recommendations from other community services the individual may be using. Based on this information, the officer decides whether to maintain the status quo or revise the conditions of release.

- AI systems used by custom authorities to assess the risk of goods entering the EU not complying with the legislation applicable at the border (e.g. which may include bans on import of illicit drugs, export sanctions contravention or other illegal activity) to identify situations where a customs control should be carried out. The AI system assesses objective and verifiable information provided to the customs related to the goods and their supply chains (e.g. nature and value of the goods, container number, means of transport for concealment of other goods, prior knowledge relating to the compliance of goods of the particular description and origin with requirements for their importation to or exportation from the Union). In certain cases, it may also process information about the prior involvement of the importer or exporter in irregularities related to import of goods, their affiliation to criminal organizations or a criminal record for drug trafficking. Such systems are out of scope of the prohibition because any prediction of a likelihood of a natural person to be involved in an import or export of illicit goods is not solely based on profiling, but on objective and verifiable information related to the goods and the importer or exporter's prior

involvement in criminal activity and subject to a human review to determine whether or not the situation requires a customs control or risk mitigation action.

5.3.3. AI systems used for crime predictions and assessments in relation to legal entities

(215) The prohibition in Article 5(1)(d) AI Act applies only to individual predictions and risk assessments of natural persons, thus typically excluding crime predictive systems that profile legal entities such as companies or non-governmental organisations.

For example,

- A tax or customs authority is using an AI system to analyse large amounts of data on transactions and tax declarations and customs data of companies for assessing the risk of a company committing tax or customs fraud constituting a criminal offence.
- AI systems used to assist customs authorities to help identify situations where an instruction not to send illicit goods to the EU should be issued to legal entities.

(216) At the same time, there may be borderline cases where a natural person acts via a legal entity as a 'sole trader' or as an independent professional (e.g. a lawyer). In such circumstances, the prohibition in Article 5(1)(d) AI Act may apply provided that all conditions are fulfilled, since the AI system profiles a specific natural person and assesses or predicts their risk of committing a criminal offence, even if this is done for purposes in relation to the commercial activity undertaken by the natural person.

5.3.4. AI systems used for individual predictions of administrative offences

(217) The prohibition in Article 5(1)(d) AI Act applies only for the prediction of criminal offences, thus excluding administrative offences from its scope, the prosecution of which is, in principle, less intrusive for the fundamental rights and freedoms of people.

For example, a public authority using AI in the context of an administrative investigation to assess the risk of potential offenders of committing minor offences (such as petty traffic offences) or irregularities in tax, procurement or expenditure processes would not fall within the scope of the prohibition in Article 5(1)(d) AI Act, even in cases where information might be gathered for possible involvement of the natural persons in criminal offences as a result of the administrative investigations and checks.

(218) Whether an offence is administrative or criminal in nature may depend on Union or national law. For offences that are not directly regulated by Union law, the national qualification of the offence is subject to scrutiny by CJEU since 'criminal offence' is a concept that has autonomous meaning within the EU law and should be interpreted consistently across Member States. The CJEU has concluded, in a different context, that the classification of the offences by the Member States is not conclusive in that

respect[142]. Relevant criteria used to assess the nature of the offence (criminal or not) may be found in relevant case-law of the CJEU and the European Court of Human Rights (ECtHR).[143]

5.4. Interplay with other Union legal acts

(219) The interplay of the prohibition in Article 5(1)(d) AI Act with the LED and GDPR is relevant when assessing the lawfulness of personal data processing under Union data protection law, such as the GDPR and the LED. In particular, Article 5(1)(d) AI Act imposes a specific prohibition for law enforcement authorities, other public authorities and private entities falling within the scope of the prohibition to assess or predict the risk of a natural person committing a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics. With regard to LED, Article 5(1)(d) AI Act is without prejudice to Article 11(3) LED, which prohibits profiling resulting in (direct or indirect) discrimination.

(220) The interplay of the prohibition in Article 5(1)(d) AI Act with Directive (EU) 2016/343 on the presumption of innocence is also relevant, since both acts are concerned - directly in the case of the Directive and indirectly in the case of the AI Act (see its Recital 42), with the fundamental right to be presumed innocent until proven guilty according to law.[144] While the Directive applies from the moment that a person is suspected or accused of having committed a criminal offence[145], the AI Act has a broader scope of application and applies already at the stage of prediction and crime prevention before a formal criminal investigation is opened against a particular person and even in cases when such predictions and risk assessments are made by private actors falling within the scope of Article 5(1)d) AI Act and not by competent law enforcement authorities, including judicial authorities.

(221) Even in cases where the prohibition in Article 5(1)(d) AI Act does not apply, it is important to emphasize that applicable Union and national law remains fully applicable, including in particular data protection, criminal procedural and police law and safeguards that may further restrict or impose additional conditions on the use of individual crime predictive AI systems.

---

[142] See, for example, Judgment of the Court (Grand Chamber) of 14 November 2013 Proceedings concerning the enforcement of a financial penalty issued against - Marián Baláž, Case C-60/12, ECLI identifier: ECLI:EU:C:2013:733.

[143] According to the CJEU's case law, it is for national courts to determine whether a non-criminal penalty may be regarded as 'criminal' in light of the so-called 'Engel criteria', See: ECtHR, judgment of 8 June 1976, Engel and Others v. the Netherlands, Application nos. 5100/71, 5101/71, 5102/71, 5354/72 and 5370/72, CE:ECHR:1976:0608JUD000510071, paragraph 82. Originally developed by the European Court of Human Rights (ECtHR) and subsequently endorsed by the CJEU, these criteria are alternative and not cumulative. When examining whether a penalty has a criminal nature, the competent national court should assess: (1) the classification of the relevant provisions under domestic law; (2) the very nature of the offence; and (3) the severity of the penalty. In evaluating the nature of the offence, aspects taken into account include inter alia whether the proceedings are instituted by a public body with statutory powers of enforcement; whether the legal rule has a punitive or deterrent purpose; whether the legal rule seeks to protect the general interests of society usually protected by criminal law; whether the imposition of any penalty is dependent upon a finding of guilt. Regarding the severity of the penalty, relevant reference is the maximum potential penalty provided in the national law. These criteria are alternative and not necessarily cumulative. See European Court of Human Rights, Guide on Article 6 of the European Convention on Human Rights, Right to a fair trial (criminal limb), updated 29 February 2024. See also CJEU, judgment of 5 June 2012, Bonda, Case C-489/10, EU:C:2012:319, paragraphs 37ff.; CJEU, judgment of 26 February 2013, Åkerberg Fransson, Case C-617/10, EU:C:2013:105, paragraph 35.

[144] The presumption of innocence is a fundamental right enshrined in Article 48 of the EU Charter of Fundamental Rights.

[145] As specified by the CJEU, it is not required that this person is made aware of their status as suspect/accused persons by the competent authorities for the Directive to apply.

6. ARTICLE 5(1)(E) AI ACT - UNTARGETED SCRAPING OF FACIAL IMAGES

(222) Article 5(1)(e) AI Act prohibits the placing on the market, putting into service for this specific purpose, or the use of AI systems that create or expand facial recognition databases through the untargeted scraping of facial images from the Internet or CCTV footage.

6.1. Rationale and objectives

(223) The untargeted scraping of facial images from the internet and from CCTV footage seriously interferes with individuals' rights to privacy and data protection and deny those individuals the right to remain anonymous. Recital 43 AI Act therefore justifies the prohibition established in Article 5(1)(e) AI Act based on the 'feeling of mass surveillance' and the risks of 'gross violations of fundamental rights, including the right to privacy'.

6.2. Main concepts and components of the prohibition

> ***Article 5(1)(e) AI Act provides***
>
> The following AI practices shall be prohibited:
>
> (e) the placing on the market, the putting into service for this specific purpose, or the use of AI systems that create or expand facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage;

(224) Several cumulative conditions must be fulfilled for the prohibition in Article 5(1)(e) AI Act to apply:

(i) The practice must constitute the 'placing on the market', 'the putting into service for this specific purpose' or the 'use' of an AI system;

(ii) for the purpose of creating or expanding facial recognition databases;

(iii) the means to populate the database are through AI tools for untargeted scraping; and

(iv) the sources of the images are either from the internet or CCTV footage.

(225) For the prohibition to apply all four conditions must be simultaneously fulfilled. The first element of placing on the market, putting into service or use of the AI system has been already analysed in section 2.3. The prohibition, therefore, applies to both providers and deployers of AI systems, each within their respective responsibilities, not to place on the market, put into service or use such AI systems. The specific criteria related to the prohibition of untargeted scraping are further described and analysed below. The prohibition applies to scraping tools that are placed on the market or being put into service 'for this specific purpose' of untargeted scraping of facial images from the internet or CCTV footage. This implies that the prohibition does not apply to any

scraping tool with which a database for face recognition may be constructed or expanded, but only to tools for untargeted scraping.

6.2.1. Facial recognition databases

(226) The prohibition in Article 5(1)(e) AI Act covers AI systems used to create or expand facial recognition databases. 'Database' in this context should be understood to refer to any collection of data, or information, that is specially organized for rapid search and retrieval by a computer. A facial recognition database is capable of matching a human face from a digital image or video frame against a database of faces, comparing it to images in the database and determining whether there is a likely match between the two. Such a facial recognition database may be temporary, centralised or decentralised. Article 5(1)(e) does not require that the sole purpose of the database is to be used for facial recognition; it is sufficient that the database can be used for facial recognition.

6.2.2. Through untargeted scraping of facial images

(227) 'Scraping' typically refers to using web crawlers, bots, or other means to extract data or content from different sources, including CCTV, websites or social media, automatically. These tools are software programmed to sift through databases and extract information and to make use of that information for another purpose.

(228) 'Untargeted' relates to a technique that operates like a 'vacuum cleaner', absorbing as much data and information as possible, without targeting specifically and individually intended subject(s) of the scarping. Scraping indiscriminately harvests data or content. Thus, the notion of 'untargeted' means without a specific focus on a given individual or group of individuals. The respect of opt-out of internet protocols such as robot.txt does not affect the untargeted nature of the scraping.

(229) If a scraping tool is instructed to collect images or video containing human faces only of specific individuals or a pre-defined group of persons, then the scraping becomes targeted, for example to find one specific criminal or to identify a group of victims. Such scraping is not covered by the prohibition in Article 5(1)(e) AI Act.

(230) For example, the targeted collection of images focusing on a class of victims, by using crawlers to pick up on images of victims that human traffickers post/advertise on social media channels, is not covered by the prohibition. Untargeted scraping should be interpreted in a manner that does not allow circumvention of the prohibition. The scraping of the Internet or CCTV footage for the creation of a database step-by-step, thereby selecting specific groups of individuals or other criteria each time, should fall within the prohibition of Article 5(1)(e) AI Act where the end-result is functionally the same as pursuing untargeted scraping from the outset.

(231) Where systems combine targeted searches for images or videos with untargeted searches, the untargeted scraping is prohibited.

6.2.3. From the Internet and CCTV footage

(232) For the prohibition in Article 5(1)(e) AI Act to apply, the source of the facial images may either be the Internet or CCTV footage. Regarding the internet, the fact that a person has published facial images of themselves on a social media platform does not mean that that person has given his or her consent for those images to be included in a facial recognition database. Examples of scraping facial images from CCTV footage include images acquired by surveillance cameras operated in places such as airports, streets, parks, etc.

> Example:
>
> A facial recognition software company collects pictures of faces. The photographs held by the company have been scraped from social media (e.g. Facebook, YouTube, Twitter, Venmo) with an 'automated image scraper' that searches the internet and detects images containing human faces. It collects those images with any associated information (such as the source of the image (URL), the geo-localisation, and sometimes the names of the individuals). The facial features are then extracted from the images and transformed into mathematical representations, which are hashed for indexation and future comparison. When a user uploads the image of an individual to the AI system, that system will determine whether that image matches a face in the database. The uploaded image will go through the same mathematical transformation as the scraped images.

(233) Where an AI system receives a picture of a person and searches the face on the internet for matches, i.e. 'reverse engineering image search engines', this will be considered to be targeted scraping. Moreover, it is questionable whether the matches would appear in a 'database'.

6.3. Out of scope

(234) The prohibition in Article 5(1)(e) AI Act does not apply to the untargeted scraping of biometric data other than facial images (such as voice samples). The prohibition does also not apply where no AI systems are involved in the scraping. Facial image databases that are not used for the recognition of persons are also out of scope, such as facial image databases used for AI model training or testing purposes, where the persons are not identified.

(235) The prohibition in Article 5(1)(e) AI Act does not apply to AI systems which harvest large amounts of facial images from the internet to build AI models that generate new images about fictitious persons because such systems would not result in the recognition of real persons. Such AI systems could fall under the transparency requirements of Article 50 AI Act.

(236) The prohibition in Article 5(1)(e) AI Act covers AI systems used to create or expand facial recognition databases. When it comes to existing facial databases built up prior

to the entry into application of the prohibition, which are not further expanded through AI-enabled untargeted scraping, those databases and their use must comply with the applicable Union data protection rules.

(237) The prohibition in Article 5(1)(e) AI Act is targeted at the creation or expansion of facial recognition databases. The concrete act of biometric identification is subject to specific rules in the AI Act and other relevant Union legislation.

6.4. Interplay with other Union legal acts

(238) In relation to Union data protection law, the untargeted scraping of the internet or CCTV material to build-up or expand face recognition databases, i.e. the processing of personal data (collection of data and use of databases) would be unlawful and no legal basis under the GDPR, EUDPR and the LED could be relied upon.

7. ARTICLE 5(1)(F) AI ACT EMOTION RECOGNITION

(239) Article 5(1)(f) AI Act prohibits AI systems to infer emotions of a natural person in the areas of workplace and education institutions, except where the use of the system is intended for medical or safety reasons. Emotion recognition systems that do not fall under the prohibition are considered high-risk pursuant to point (1)(c) of Annex III AI Act. Article 50(3) AI Act lays down certain transparency requirements for the use of emotion recognition systems.

7.1. Rationale and objectives

(240) Emotion recognition technology is quickly evolving and comprehends different technologies and processing operations to detect, collect, analyse, categorise, react, interact and learn emotions from persons. Such technology is also referred to as 'affect technology'. Emotion recognition can be used in multiple areas and domains for a wide range of applications[146] such as for analysing customer behaviour[147] and targeted advertising and neuromarketing[148]; in the entertainment industry, for example to provide personalised recommendations or to predict reactions to movies; in medicine and healthcare, for example to detect depression, for suicide prevention or to detect autism, in education, for example to monitor attention or engagement of learners (pupils and students at different ages); in employment, for example to accompany the recruitment process, to monitor emotions or boredom of employees, but also well-being applications for 'making workers happier'[149]; for law enforcement and public safety,

[146] The use of emotions for economic purposes is also referred to as 'emotionomics'.
[147] See e.g., G. Mangano, A. Ferrari, C. Rafale, E. Vezzetti, F. Marcolin, 'Willingness of sharing facial data for emotion recognition: a case study in the insurance market' in *AI & Society*, London, Springer, 2023.
[148] See N. Lee, A. J. Broderick, & L. Chamberlain, 'What is 'neuromarketing'? A discussion and agenda for future research', in *International Journal of Psychophysiology*, 63(2), 2007,199- 204 defining neuromarketing as a field of study as "*the application of neuroscientific methods to analyze and understand human behaviour in relation to markets and marketing exchanges*" (p. 200).
[149] See E. Ackerman, & E. Strickland, 'Are you Ready for Workplace Brain Scanning? Extracting and using brain data will make workers happier and more productive, backers say', *IEEE Spectrum*, 19 November 2022, https://spectrum.ieee.org/neurotech-workplace-innereye-emotiv The authors explain that 'sensors detect electrical activity across different areas of the brain, and the patterns in that activity can be broadly correlated with different feelings or physiological responses, such as stress, focus, or a reaction to external stimuli'.

for example with lie detectors or emotion screening at big events; and for many other purposes.

(241) Emotion recognition is frequently doubted as to its effectiveness or as to its accuracy.[150] Recital 44 AI Act explains that there are 'serious concerns about the scientific basis of AI systems aiming to identify or infer emotions, particularly as expression of emotions vary considerably across cultures and situations, and even within a single individual. Among the key shortcomings of such systems are the limited reliability, the lack of specificity and the limited generalisability.' It further explains that emotion recognition can lead to 'discriminatory outcomes and can be intrusive to the rights and freedoms of the concerned persons', in particular the rights to privacy, human dignity and freedom of thought. This plays an important role in asymmetric relationships especially in the context of the workplace and education and training institutions, where both workers and students are in particularly vulnerable positions. At the same time, emotion recognition in specific use contexts, such as for safety and medical care (e.g. health treatment and diagnosis) has benefits.[151]

7.2. Main concepts and components of the prohibition

> ***Article 5(1) (f) AI Act provides:***
>
> The following AI practices shall be prohibited:
>
> f) the placing on the market, the putting into service for this specific purpose, or the use of AI systems to infer emotions of a natural person in the areas of workplace and education institutions, except where the use of the AI system is intended to be put in place or into the market for medical or safety reasons.

(242) Several cumulative conditions must be fulfilled for the prohibition in Article 5(1)(f) AI Act to apply:

(i) The practice must constitute the 'placing on the market', 'the putting into service for this specific purpose' or the 'use' of an AI system;

(ii) AI system to infer emotions[152];

(iii) in the area of the workplace or education and training institutions; and

(iv) excluded from the prohibition are AI systems intended for medical or safety reasons.

(243) For the prohibition to apply all four conditions must be simultaneously fulfilled. The first element, i.e. the placing on the market, putting into service or use of the AI system, has already been analysed in section 2.3.. The prohibition, therefore, applies to both

[150] See e.g., J. Stanley, *Experts Say 'Emotion Recognition' lacks Scientific Foundation,* 18.7.2019, ACLU, referring to a study by L. Feldman Barrett e.a., 'Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements', *Psychological Science in the Public Interest*, 2019, pp.iii-90.
[151] See e.g., R. El Kaliouby and R. Picard and S. Baron-Cohen, 'Affective Computing and Autism', *Annals New York Academy of Sciences,* 2007, pp.228-248
[152] Or the technology is capable of inferring emotions (i.e. when placing it on the market).

providers and deployers of AI systems, each within their respective responsibilities, not to place on the market, put into service or use such AI systems. The other conditions related to the prohibition are further described and analysed below.

7.2.1. AI systems to infer emotions

a) AI systems to infer emotions versus emotion recognition systems

(244) Article 3(39) AI Act defines 'emotion recognition systems' as AI systems 'for the purpose of identifying and inferring emotions or intentions of natural persons on the basis of their biometric data. The prohibition in Article 5(1)(f) AI Act does not refer to 'emotion recognition systems', but only to 'AI systems to infer emotions of a natural person'. Recital 44 further clarifies that that prohibition covers AI systems 'to identify or infer emotions'.

(245) Inferring generally encompasses identifying as a prerequisite, so that the prohibition should be understood as including both AI systems identifying or inferring emotions or intentions.[153] For consistency reasons, it is also important to construe the prohibition in Article 5(1)(f) AI Act as having a similar scope as the rules applicable to other emotion recognition systems (Annex III, point 1(c), and Article 50 AI Act) and to limit it to inferences based on a person's biometric data. The definition in Article 3(39) AI Act of emotion recognition systems should therefore be considered relevant in relation to Article 5(1)(f) AI Act.

b) Identification and inference of emotions or intentions

(246) 'Identification' occurs where the processing of the biometric data (for example, of the voice or a facial expression) of a natural person allows to directly compare and identify an emotion with one that has been previously programmed in the emotion recognition system. 'Inferring' is done by deducing information generated by analytical and other processes by the system itself. In such a case, the information about the emotion is not solely based on data collected on the natural person, but it is inferred from other data, including machine learning approaches that learn from data how to detect emotions.[154]

c) Emotions

(247) For the purpose of Article 5(1)(f) AI Act, the concept of emotions or intentions should be understood in a wide sense and not interpreted restrictively. Recital 18 AI Act provides some detail, listing emotions 'such as happiness, sadness, anger, surprise, disgust, embarrassment, excitement, shame, contempt, satisfaction and amusement'. These examples are not exhaustive.

[153] See also Recital 18 AI Act.
[154] See Recital 12 AI Act. Inferred data is hence also often the result of probability-based analytical (big data) processes aimed at finding correlations and finding patterns in data sets.

(248) The prohibition should not be circumvented by referring to attitudes, and includes cases where the AI system finds on the basis of the biometric data that a person is showing for example an angry attitude.

(249) Recital 18 AI Act clarifies that emotions or intentions do not include 'physical states, such as pain or fatigue, including, for example, systems used in detecting the state of fatigue of professional pilots or drivers for the purpose of preventing accidents.' It further clarifies that emotion recognition systems do not include 'the mere detection of readily apparent expressions, gestures or movements, unless they are used for identifying or inferring emotions', which should be understood to also apply to Article 5(1)(f) AI Act. Those expressions can be basic facial expressions, such as a frown or a smile, or gestures such as the movement of hands, arms or head, or characteristics of a person's voice, such as a raised voice or whispering. However, when these readily apparent expressions or gestures are used for identifying or inferring emotions or intentions, they are covered by the prohibition.

For example,

- The observation that a person is smiling is not emotion recognition.
- Identifying whether a person is sick is not emotion recognition.
- A TV broadcaster using a device that allows to track how many times its news presenters smile to the camera is not emotion recognition.
- Concluding that a person is happy is emotion recognition. An AI system that infers that an employee is unhappy, sad or angry towards customers (e.g. from body gestures, a frown or the lack of a smile) is 'emotion recognition'.
- Systems inferring from voice or body gestures, that a student is furious and about to become violent, is 'emotion recognition'.
- Using AI recognition systems to infer a professional pilot's or driver's fatigue to alert them and suggest when to take brakes to avoid accidents is not 'emotion recognition', since emotion recognition does not include physical states such as pain or fatigue.

d) On the basis of their biometric data

(250) According to the definition in Article 3(39) AI Act, only AI systems identifying or inferring emotions or intentions based on biometric data constitute emotion recognition systems.[155]

[155] Article 3(34) AI Act: defines 'biometric data' as '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'' See also Recital 18 AI Act. About emotion inferences from voice and speech.

(251) Personal characteristics from which biometric data can be extracted are physical or behavioural attributes. Physiological biometrics employ physical, structural, and relatively static attributes of a person, such as their fingerprints, the pattern of their iris, contours of their face, or the geometry of veins in their hands. Some modalities are microscopic in nature, but still exhibit biological and chemical structures that can be acquired and identified e.g., DNA and odour[156]. Behavioural biometrics monitor the distinctive characteristics of movements, gestures, and motor-skills of individuals as they perform a task or series of tasks. This means that human movements, such as walking (gait analysis) or finger contact with a keyboard (keystrokes), are captured and analysed. Behavioural biometrics encompass a variety of modalities that exhibit both voluntary and involuntary repeated motions and associated rhythmic timings/pressures of body features ranging from signatures, gait, voice, and keystrokes through to eye tracking and heartbeats[157], electroencephalography (EEG)[158], or electrocardiograms (ECG)[159]. The biometric input can relate to one modality (e.g., facial images) or multiple modalities (e.g., facial information combined with electroencephalogram (EEG)). Recital 18 gives as examples facial expressions, gestures such as movement of hands or characteristics of a person's voice.

For example,

- An AI system inferring emotions from written text (content/sentiment analyses) to define the style or the tone of a certain article is not based on biometric data and therefore does not fall within the scope of the prohibition.

- An AI system inferring emotions from key stroke (way of typing), facial expressions, body postures or movements is based on biometric data and falls within the scope of the prohibition.

(252) The AI Act definition of biometric data is therefore broad and includes any biometric data used for emotion recognition, biometric categorisation or other purposes.[160]

7.2.2. Limitation of the prohibition to workplace and educational institutions

(253) The prohibition in Article 5(1)(f) AI Act is limited to emotion recognition systems in the 'areas of workplace and educational institutions'. As clarified in recital 44 AI Act,

[156] Physiological and Behavioural Biometrics - Biometrics Institute
[157] Physiological and Behavioural Biometrics - Biometrics Institute
[158] See EDPS, TechDispatch 1/2024 – Neurodata, 3.6.2024, in which the use of brain data and related technology is discussed, as well as the legal implication, including the proposition of new 'neurorights', including mental privacy and integrity. In S. O'Sullivan, H. Chneiweiss, A. Pierucci and K. Rommelfanger, *Neurotechnologies and Human Rights Framework: Do we need new Human Rights?,* Report, OECD and CoE, 9.11.2021, p.33 , a state of the art and legal aspects of neurotech is discussed.
[159] See Hasnul et al., 2021, Electrocardiogram-Based Emotion Recognition Systems and Their Applications in Healthcare.
[160] In the AI Act, the definition of biometric data does not include the wording 'which allow or confirm the unique identification' (the functional use of biometric data), contrary to the definition of biometric data in the GDPR that includes this requirement. The GDPR definition of biometric data will apply under data protection rules with regard to the processing of personal data (and when for example Article 9(1) and 9(2) GDPR would be applicable).

this limitation is meant to address the imbalance of power in the context of work or education.

a) 'Workplace'

(254) The notion of 'workplace' should be interpreted broadly. That notion relates to any specific physical or virtual space where natural persons engage in tasks and responsibilities assigned by their employer or by the organisation they are affiliated to, for example in case of self-employment. This includes any setting where the work is performed and can vary widely based on the nature of the job, spanning from indoor office spaces, factories and warehouses to publicly accessible spaces like shops, stadiums or museums, to open-air sites or cars, as well as temporary or mobile work sites. This is independent from the status as an employee, contractor, trainee, volunteer, etc.[161] The notion of 'workplace' in Article 5(1)(f) AI Act should also be understood to apply to candidates during the selection and hiring process, consistently with other provisions of the AI Act addressing the placing on the market, putting into service or use of AI systems in the area of employment, workers management and access to self-employment, since there is an imbalance of powers and the intrusive nature of emotion recognition may already apply at the recruitment stage.

For example:

- Using webcams and voice recognition systems by a call centre to track their employee's emotions, such as anger, is prohibited.[162] If only deployed for personal training purposes, emotion recognition systems are allowed if the results are not shared with HR responsible persons and cannot impact the assessment, promotion etc. of the person trained, provided that the prohibition is not circumvented and the use of the emotion recognition system does not have any impact on the work relationship.

- Using voice recognition systems by a call centre to track their customers emotions, such as anger or impatience, is not prohibited by Article 5(1)(f) AI Act (for example to help the employees cope with certain angry customers).

- AI systems monitoring the emotional tone in hybrid work teams by identifying and inferring emotions from voice and imagery of hybrid video calls, which would typically serve the purpose of fostering social awareness, emotional dynamics management, and conflict prevention, are prohibited.

- Using emotion recognition AI systems during the recruitment process is prohibited.

- Using emotion recognition AI systems during the probationary period is prohibited.

[161] See also the recitals in relation to the high-risk AI systems in the workplace, such as Recital 56, deploying a broad interpretation. See also the list of high-risk AI systems in Annex III, referring to self-employment at 4. Self-employment is also broadly covered by EU anti-discrimination law.

[162] Example from Boyd et al., 2023, Automated Emotion Recognition in the Workplace: How Proposed Technologies Reveal Potential Futures of Work..

- Using cameras by a supermarket to track its employees' emotions, such as happiness, is prohibited.

- Using cameras by a supermarket or a bank to detect suspicious customers, for example to conclude that somebody is about to commit a robbery, is not prohibited under Article 5(1)(f) AI Act, when it is ensured that no employees are being tracked and there are sufficient safeguards.

b) 'Education institutions'

(255) The reference to **education institutions** is broad and should be understood to include both public and private institutions. There is no limitation as regards the types or ages of pupils or students or of a specific environment (online, in person, in a blended mode[163]etc). For example, education and training institutions at all levels fall under the scope of the prohibition in Article 5(1)(f) AI Act, including vocational schools, i.e. schools where students learn skills involving the use of their hands[164] and continuous training[165]. Education institutions are normally accredited or sanctioned by the relevant national education authorities or equivalent authorities. A key feature is that education institutions may provide a certificate (respectively participation is a precondition for obtaining a certificate). The prohibition should be understood to also apply to candidates during the admissibility process.

For example:

- An AI-based application using emotion recognition for learning a language online outside an education institution is not prohibited under Article 5(1)(f) AI Act. By contrast, if students are required to use the application by an education institution, the use of such emotion recognition system is prohibited.

- An education institution using AI-based eye tracking software when examining students online to track the fixation point and movement of the eyes (gaze point, e.g., to detect if unauthorized material is used) is not prohibited, because the system does not identify or infer emotions. By contrast, if the system is also used to detect emotions, such as emotional arousal and anxiousness, this would fall within the scope of the prohibition.

- Using an emotion recognition AI system by an education institution to infer the interest and attention of students is prohibited. By contrast, if only deployed for learning purposes in the context of a role-play (for example, for training actors or

[163] Blended learning is to be understood as taking more than one approach in the education and training process, including blending digital (including online learning) and non-digital learning tools.
[164] See e.g., the impact assessment accompanying the proposal of the Commission, where specific AI uses by vocational training institutions were mentioned as posing intense interference with a broad range of fundamental rights, e.g., when assessing: EU Commission, *Commission Staff Working Document. Impact Assessment. Annexes, SWD(2021)84 final, Part2/2,* p. 43. See also I. Tuomi, T*he , The use of Artificial Intelligence (AI) in education*, European Parliament, 2020, pp. 9- 10.
[165] See Article 14 Charter.

teachers), emotion recognition systems are allowed if the results cannot impact the evaluation or certification of the person being trained.

- Using an emotion recognition AI system by an education institution during admissibility tests for new students is prohibited.

- Using an AI system that allows to capture students talking to each other via their phones or other channels during online lectures by an education institution is not prohibited, since it does not infer emotions. By contrast, if the system is also used to detect emotions, such as emotional arousal, anxiousness and interest, this would fall within the scope of the prohibition.

- An education institution employing an emotion recognition AI system on both teachers (workplace) and students (education) is prohibited.

7.2.3. Exceptions for medical and safety reasons

(256) The prohibition in Article 5(1)(f) AI Act contains an explicit exception for emotion recognition systems used in the area of the workplace and education institutions for medical or safety reasons, such as systems for therapeutical use.[166] In light of the AI Act's objective to ensure a high-level of fundamental rights protection, this exception should be narrowly interpreted.

(257) In particular, therapeutic uses should be understood to mean uses of CE-marked medical devices. Moreover, this exception does not comprise the use of emotion recognition systems to detect general aspects of wellbeing. The general monitoring of stress levels at the workplace is not permitted under health or safety aspects. For example, an AI system intended to detect burnout or depression at the workplace or in education institutions would not be covered by the exception and would remain prohibited.

(258) The notion of safety reasons within this exception should be understood to apply only in relation to the protection of life and health and not to protect other interests, for example property against theft or fraud.

(259) It follows from this narrow interpretation of the exception that any use for medical and safety reasons should always remain limited to what is strictly necessary and proportionate, including limits in time, personal application and scale, and should be accompanied by sufficient safeguards. Such safeguards could include, for example, prior written and motivated expert opinion relating to the specific use case. The necessity should be assessed on an objective basis in relation to the medical and safety purpose, and not refer to the employer's or educational institution's 'needs'. This assessment should inquire whether less intrusive alternative means exist which would achieve the same purpose.

[166] Recital 44 AI Act.

(260) Employers and educators should only deploy emotion recognition systems for medical and safety reasons in case of an explicit need[167]. Data collected and processed in this context may not be used for any other purpose. This is particularly important given that the use of AI management software at work has proven to potentially negatively impact workers' health and safety. Continuous monitoring via wearables, for instance, may increase work-stress while affecting productivity[168].

(261) Since Recital 18 AI Act excludes from the definition of emotion recognition systems physical states, such as pain or fatigue, a number of AI systems used for safety reasons would already not fall under that definition, including, for example, systems used in detecting the state of fatigue of professional pilots or drivers for the purpose of preventing accidents.

(262) Other laws, including data protection rules, remain applicable to emotion recognition systems that fulfil the conditions of the exception in Article 5(1)(f) AI Act[169].

(263) Emotion recognition systems that classify as high-risk systems pursuant to Article 6(2) and Annex III(1)(c) AI Act will need to comply with the high-risk requirements in Chapter III Section 2 AI Act and the transparency obligation of Article 50(3) AI Act.

For example:

Emotion recognition may be deployed for medical reasons to assist employees or students with autism and improve accessibility for those who are blind or deaf[170]. Such uses would fall within the exception for medical reasons in Article 5(1)(f) AI Act.

By contrast, emotion recognition for assessing students' or employees' well-being, motivation levels, and job or learning satisfaction do not qualify as 'use for medical reasons' and would be prohibited.

An employer would be prohibited from deploying AI-enabled devices or digital assistants at the workplace for measuring anxiety based on measured stress levels or for measuring boredom of employees, unless the elevated stress level/lack of concentration would pose a specific danger, for example when deploying dangerous machines or dealing with dangerous chemicals. In the latter case, the employer may not use the data for other purposes, such as assessing the employee's work performance.

7.3. More favourable Member State law

(264) Article 2(11) AI Act provides that the Union or Member States may keep or introduce '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'.

[167] In conformity with EU employment law, if such new technologies are introduced, employers shall also consult with workers or their representatives, conform national procedures. Without respecting these procedural requirements, such systems cannot be introduced by reference to the AI Act as such. They will require also consent from the point of view of data protection legislation, which remains applicable.
[168] The Interconnection between the AI Act and the EU's Occupational Safety and Health Legal Framework - Global Workplace Law & Policy (kluwerlawonline.com).
[169] From December 2026 Directive (EU) 2024/2831 of the European Parliament and of the Council of 23 October 2024 on improving working conditions in platform work will apply.
[170] Systems could be usefully employed for helping employees or students/pupils to understand the emotions of colleagues etc.

Collective bargaining agreements which are more favourable to workers may also be allowed or encouraged.

For example, Member States may adopt laws providing that the use of emotion recognition systems in the area of work may not be applied for medical purposes.

7.4. Out of scope

(265) As mentioned before, out of scope are:

- AI systems inferring emotions and sentiments not on the basis of biometric data,
- AI systems inferring physical states such as pain and fatigue.

(266) Emotion recognition systems used in all other domains other than in the areas of the workplace and education institutions do not fall under the prohibition in Article 5(1)(f) AI Act. Such systems are, however, considered high-risk AI systems.[171] At the same time, such systems may be prohibited in certain cases by virtue of Article 5(1)(a) and (b) AI Act (harmful manipulation and exploitation), or by virtue of other Union legislation. All other applicable legislation, such as Union data protection law, consumer protection etc. continue to apply to such systems.

For example: Emotion recognition systems used in a commercial context for addressing customers do not fall under the prohibition of Article 5(1)(f) AI Act, whether based on biometric data or not. Hence, examples such as AI systems that enable emotion recognition based on keystroke or based on voice messages of customers (e.g., chat messages, use of virtual voice assistants), used in online marketing for applications for displaying personalized messages and for advertisement purposes including in smart environments ('intelligent billboards') are not covered by the prohibition. Nevertheless, such practices may be covered by the prohibitions of harmful manipulation and exploitation in Article 5(1)(a) and (b) AI Act[172], if all conditions for the application of those prohibitions are met.

a) Other systems out of scope

(267) 'Crowd control' generally refers to the control and monitoring of the behaviour of groups to maintain (public) order and event safety. It is often associated with large crowd events (e.g., soccer or football games, concerts, etc) or specific places, such as airports or trains. Crowd control systems can operate without inferring emotions of individual persons, when for example analysing the general noise and mood level at a given place. In that case, the system would not fall within the scope of Article 5(1)(f) AI Act, because it does not infer emotions of (a concrete) natural person.

[171] Article 6(2) AI Act and Annex III, 1 letter c).
[172] These situations might also be prohibited under other rules, such as data or consumer protection.

(268) However, there may be instances where such crowd control systems infer emotions of individuals, for example whether there are many angry faces. Normally, such AI systems would not fall under the prohibition of Article 5(1)(f) AI Act, since they are typically not used in the workplace or in education institutions.

(269) Also out of scope are systems that are used in the medical field for example care robots, or medical practitioners using emotion recognition systems during an examination at their workplace, and voice monitors that analyse emergency calls.

(270) Such systems will often screen persons that are there in a work context, for example the security staff at a football stadium or at a central station (where such systems are used to recognize aggressive behaviour), or employees in the medical field. In such cases, deployers must employ safeguards to avoid the screening of employees. However, it cannot be completely avoided that such systems also infer the emotions of those employees. Since the primary objective of the system is not targeted at assessing employees' emotions, these systems should be considered to be outside the scope of the prohibition. Deployers of such systems remain responsible to ensure that employees are not adversely affected by their use.

8. ARTICLE 5(1)(G) AI ACT: BIOMETRIC CATEGORISATION FOR CERTAIN 'SENSITIVE' CHARACTERISTICS

(271) Article 5(1)(g) AI Act prohibits 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 the labelling, filtering, or categorisation of biometric data sets acquired in line with Union or national law, which may be used, for example, for law enforcement purposes.[173]

8.1. Rationale and objectives

(272) A wide variety of information, including 'sensitive' information, may be extracted, deduced or inferred from biometric information, even without the knowledge of the persons concerned, to categorise those persons. This may lead to unfair and discriminatory treatment, for example when a service is denied because somebody is considered to be of a certain race. AI-based biometric categorisation systems for the purpose of assigning natural persons to specific groups or categories, relating to aspects such as sexual or political orientation or race, violate human dignity and pose significant risks to other fundamental rights, such as privacy and non-discrimination. They are therefore prohibited by Article 5(1)(g) AI Act.

8.2. Main concepts and components of the prohibition

Article 5(1)(g) AI Act provides

[173] Recital 30 AI Act.

The following AI practices shall be prohibited:

g) the placing on the market, the putting into service for this specific purpose, or the 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 categorising of biometric data in the area of law enforcement;

(273) Several cumulative conditions must be fulfilled for the prohibition in Article 5(1)(g) AI Act to apply:

(i) The practice must constitute the 'placing on the market', 'the putting into service for this specific purpose' or 'the use' of an AI system;

(ii) The system must be a biometric categorisation system;

(iii) individual persons must be categorised;

(iv) based on their biometric data;

(v) to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life, or sexual orientation.

(274) For the prohibition to apply, all five conditions must be simultaneously fulfilled. The first condition, i.e. the placing on the market, the putting into service or the use of the AI system, is analysed in section 2.3. The prohibition, therefore, applies to both providers and deployers of AI systems, each within their respective responsibilities, not to place on the market, put into service or use such AI systems. The other conditions for the application of the prohibition[174] are further described and analysed below.

(275) The prohibition does not cover the labelling or filtering of lawfully acquired biometric datasets, including for law enforcement purposes.

8.2.1. Biometric categorisation system

(276) 'The categorisation of an individual by a biometric system is typically the process of establishing whether the biometric data of an individual belongs to a group with some pre-defined characteristic. It is not about identifying an individual or verifying their identity, but about assigning an individual to a certain category. For instance, an advertising display may show different adverts depending on the individual that is looking at it based on their age or gender.'[175] Persons may also simply be categorised for statistical reasons, without being identified and without the objective to identify them.

[174] For the criterion of 'AI system', 'the 'placing on the market', 'the putting into service for this specific purpose' or the use, see *above*.
[175] See the Article 29 Working Party, *Opinion 3/2012 on developments in biometric technologies*, WP193, 27.4.2012, p. 6.

(277) Article 3(40) AI Act defines a biometric categorisation system as an AI system for the purpose of assigning natural persons to specific categories on the basis of their biometric data, unless it is ancillary to another commercial service and strictly necessary for objective technical reasons. As explained in section 7.2.1.d), 'biometric data' is defined in Article 3(34) AI Act. In particular, biometric data comprises behavioural characteristics that are based on biometric features. The scope of biometric categorisation excludes categorisation according to clothes or accessories, such as scarfs or crosses, as well as social media activity.

(278) Biometric categorisation may rely on categories of physical characteristics (e.g. facial features and form, skin colour) based on which persons are assigned to specific categories. Some of these categories may be of a special 'sensitive' nature' or characteristics protected under Union non-discrimination law, such as race. However, biometric categorisation may also be based on DNA or on behavioural aspects, such as keystroke analysis or a person's gait[176].

(279) To fall outside the scope of the definition of biometric categorisation under the AI Act, two conditions – being 'ancillary to another commercial service and strictly necessary for objective technical reasons' – must be cumulatively fulfilled.

(280) According to recital 16 AI Act, a purely ancillary feature is a feature that is intrinsically linked to another commercial service, meaning that the feature cannot, for objective technical reasons, be used without the principal service, and the integration of that feature or functionality is not a means to circumvent the applicability of the rules of the AI Act.

For example, the following uses of AI are permitted under Article 5(1)(g) AI Act:

- Filters categorising facial or bodily features used on online marketplaces to allow a consumer to preview a product on him or herself could constitute such an ancillary feature, since they can be only used in relation to the principal service which consists in selling a product.

- Filters integrated into online social network services which categorise facial or bodily features to allow users to add or modify pictures or videos could also be considered to be ancillary feature, since such a filter cannot be used without the principal service of the social network services consisting in the sharing of content online.

In contrast, examples of uses that would be prohibited include:

- An AI system that categorises persons active on a social media platform according to their assumed political orientation, by analysing the biometric data from the photos they have uploaded on the platform, to send them targeted political messages. While such a system may only be ancillary to the political advertising, it would not be

[176] See e.g., the Article 29 Working Party, *Opinion 3/2012 on developments in biometric technologies*, WP193, 27.4.2012, pp.16-17. The Group refers here to 'soft recognition' (p. 17), i.e. 'detection of behaviour or specific needs of people'.

'strictly necessary for objective technical reasons', hence the conditions for excluding it from the definition of biometric categorisation are not fulfilled.

- An AI system that categorises persons active on a social media platform according to their assumed sexual orientation by analysing the biometric data from photos shared on that platform and on that basis serves those persons advertisements would qualify as biometric categorisation within the meaning of the AI Act. Also in this case there is no strict necessity for this 'ancillary service', hence the exclusion from the prohibition does not apply.

8.2.2. Persons are individually categorised based on their biometric data

(281) The use of biometric data for the categorisation of natural persons is an essential element for the prohibition to apply (see above section 8.2.1. and 7.2.1.d)).

(282) Furthermore, for the prohibition to apply, natural persons must be 'individually' categorised. If this is not the purpose or outcome of the biometric categorisation, the prohibition does not apply, for example if a whole group is categorised without looking at the individual.

Examples of individual categorisation include:

- AI systems that conduct 'Attribute Estimation' (count demographics), including for example 'age, gender, ethnicity', on the basis of for example bodily features, such as face, height, or skin, eye and hair colour (or a combination thereof).

- AI systems capable of categorising individuals and singling them out based on a specific feature (e.g. a scar under the right eye), or because they have a tattoo on their right hand.

These use-cases are examples for individual biometric categorisation. For these examples to fall within the prohibition of Article 5(1)(g) AI Act all conditions of that provision must be fulfilled.

8.2.3. To deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation

(283) Article 5(1)(g) AI Act prohibits only biometric categorisation systems which have as their objective to deduce or infer a limited number of sensitive characteristics: race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation.

For example, systems prohibited under Article 5(1)(g) AI Act include:

- a biometric categorisation system that claims to be capable of deducing an individual's race from their voice (This is different from a system that categorises persons

according to skin or eye colour, or a system that analyses the DNA of victims of crimes in view of their origin. Those systems would not be prohibited).

- a biometric categorisation system that claims to be capable of deducing an individual's religious orientation from their tattoos or faces.

8.3. Out of scope

(284) The prohibition in Article 5(1)(g) AI Act does not cover AI systems engaged in the labelling or filtering of lawfully acquired biometric datasets, such as images, based on biometric data, including in the area of law enforcement. This is further explained in recital 30 AI Act[177].

(285) The labelling or filtering of biometric datasets may be done by biometric categorisation systems precisely to guarantee that the data equally represent all demographic groups, and not, for example, over-represent one specific group. If the data used for training an algorithm are biased against a specific group (i.e. systematic differences in the data exist between groups due to the way the data are collected, or data is historically biased), the algorithm may replicate this bias, possibly resulting in unlawful discrimination against persons or groups of persons.[178] For this reason, labelling on the basis of some protected sensitive information may be necessary for high-quality data, precisely to prevent discrimination. The AI Act may even require labelling operations to conform to the AI Act's requirements for high-risk AI systems.[179] Such labelling or filtering of biometric data is therefore explicitly exempted from the prohibition in Article 5(1)(g) AI Act. The prohibition only applies where biometric data is categorised to infer race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation.

Examples of permissible labelling or filtering include:

- the labelling of biometric data to avoid cases where a member of an ethnic group has a lower chance of being invited to a job interview because the algorithm was 'trained' based on data where that particular group performs worse, i.e. has worse outcomes than other groups.[180]

- the categorisation of patients using images according to their skin or eye colour may be important for medical diagnosis, for example cancer diagnoses.

[177] Recital 30 AI Act: 'That prohibition should not cover the lawful labelling, filtering or categorisation of biometric data sets acquired in line with Union or national law according to biometric data, such as the sorting of images according to hair colour or eye colour, which can for example be used in the area of law enforcement'.
[178] Ibid.
[179] See e.g., Article 10 and 17 AI Act.
[180] FRA, #BigData. Discrimination in data supported decision making, Luxemburg, 2018, 14, p. 5.

(286) Article 5(1)(f) AI Act also provides that the prohibition in that provision does not apply to the labelling or filtering of lawfully acquired datasets in the area of law enforcement[181].

> For example, this covers the use by a law enforcement authority of an AI system that allows labelling and filtering of a dataset suspected of containing child sexual abuse material. In a first step, law enforcement would use the support of AI systems to detect and redact sensitive data from images. Furthermore, filtering and labelling according to gender, age, biometric data such as eye and hair colour, scars and marking could help with identifying the victims or creating links with other cases. Similarly filtering and labelling abusers' hands based on specific characteristics such as length of fingers or any distinguishing markings or tattoos to help with identifying possible suspects is permitted.

8.4. Interplay with other Union law

(287) AI systems intended to be used for biometric categorisation according to sensitive attributes or characteristics protected under Article 9(1) GDPR on the basis of biometric data, in so far as these are not prohibited under this Regulation, are classified as high-risk[182] under the AI Act[183].

(288) Article 5(1)(g) AI Act further restricts the possibilities for a lawful personal data processing under Union data protection law, such as the GDPR, LED, EUDPR. In particular, Article 5(1)(g) AI Act excludes the possibilities for biometric categorisation of natural persons, based on their biometric data, as defined in the AI Act, to infer race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation, subject to the exception for labelling or filtering of lawfully acquired biometric data sets, including in the area of law enforcement, as described above. Moreover, the prohibition in Article 5(1)(g) AI Act is consistent with Article 11(3) LED, which explicitly prohibits any 'profiling' that results in discrimination on the basis of special categories of personal data, such as race, ethnic origin, sexual orientation, political opinion, or religious beliefs.

9. ARTICLE 5(1)(H) AI ACT - REAL-TIME REMOTE BIOMETRIC IDENTIFICATION (RBI) SYSTEMS FOR LAW ENFORCEMENT PURPOSES

(289) Article 5(1)(h) AI Act prohibits the use of real-time RBI systems in publicly accessible spaces for law enforcement purposes, subject to limited exceptions exhaustively set out in the AI Act. Specifically, Article 5(1)(h)(i)-(iii) AI Act envisages three situations in which the use of such systems may be permitted where authorised by national

[181] The AI Act concerning the use of biometric categorisation systems for law enforcement is based on Article 16 TFEU. See also Recital 3 AI Act.
[182] Recital 54 and Annex III, point 1 letter b) AI Act.
[183] Recital 54 and Annex III, point 1 letter b). AI Act.

legislation and where the conditions and safeguards of Article 5(2) to (7) AI Act are met.

(290) In accordance with Article 5(5) AI Act, Member States are free to decide whether and in which of the three situations the use of real-time RBI systems in publicly accessible spaces for law enforcement purposes is permitted in their territory. In the absence of national legislation allowing and regulating such use, law enforcement authorities and entities acting on their behalf may not deploy such systems for law enforcement purposes. The existence of national legislation that complies with the relevant requirements of the AI Act is therefore a pre-requisite of such use.

(291) Article 5(1)(h) AI Act only prohibits the use of real-time RBI systems in publicly accessible spaces for law enforcement purposes, so that only deployers of such systems are concerned by that provision. The placing on the market and the putting into service of such systems, as well as the use of other RBI systems, is not prohibited, but subject to the rules for high-risk AI systems in accordance with Article 6(2) and point a) of Annex III AI Act[184]. Where a Member States authorises the use of real-time RBI systems in publicly accessible spaces for law enforcement purposes for any of the three objectives listed in Article 5(1)(h) AI Act, the rules for high-risk AI systems also apply to that use.

(292) Finally, specific rules apply to the retrospective use of RBI systems for law enforcement purposes. Such non-real-time use is not prohibited, but subject to additional safeguards for the deployment of high-risk AI systems (Article 26(10) AI Act).

9.1. Rationale and objectives

(293) Recital 32 AI Act acknowledges the intrusive nature of real-time RBI systems in publicly accessible spaces for law enforcement purposes to the rights and freedoms of persons concerned, to the extent that it may affect the private life of a large part of the population, evoke a feeling of constant surveillance, and indirectly dissuade the exercise of the freedom of assembly and other fundamental rights. Technical inaccuracies of AI systems intended for the remote biometric identification of natural persons can lead to biased results and entail discriminatory effects. Such possibly biased results and discriminatory effects are particularly relevant with regard to age, ethnicity, race, sex or disabilities. In addition, the immediacy of the impact and the limited opportunities for further checks or corrections in relation to the use of such systems operating in real-time carry heightened risks for the rights and freedoms of the persons concerned in the context of, or impacted by, law enforcement activities.'

(294) However, where the use of such systems is strictly necessary to achieve a substantial public interest and where the situations in which such use may occur are exhaustively listed and narrowly defined, that use outweighs the risks to fundamental rights (Recital

[184] In addition, specific rules applicable to the retrospective use of RBI systems for a law enforcement purposes (Article 26(10) AI Act).

33 AI Act). To ensure that such systems are used in a 'responsible and proportionate manner', their use is subject to the safeguards and the specific obligations and requirements in Article 5(2)-(7) AI Act.

9.2. Main concepts and components of the prohibition

Article 5(1)(h) AI Act

The following AI practices shall be prohibited:

h) the use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement, unless and in so 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 or sexual exploitation of human beings, as well as the 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 purpose of conducting a criminal investigation or prosecution or executing a criminal penalty for offences referred to in Annex II and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least four years.

Point (h) of the first subparagraph is without prejudice to Article 9 of Regulation (EU) 2016/679 for the processing of biometric data for purposes other than law enforcement.

(295) Several cumulative conditions must be fulfilled for the prohibition in Article 5(1)(h) AI Act to apply:

(i) The AI system must be a RBI system;

(ii) The activity consists of the 'use' of that system;

(iii) in 'real-time',

(iv) in publicly accessible spaces, and

(v) for law enforcement purposes.

(296) The second condition, i.e. the 'use' of the AI system, has been already analysed in section 2.3. of these Guidelines. The other conditions listed above are further described and analysed below.

9.2.1. The Notion of Remote Biometric Identification

(297) Biometric recognition technologies detect, capture, and transform measurable physical characteristics (such as eye distance and size, nose length, etc.) or behavioural

characteristics (such as gait or voice) into machine-readable biometric data (see section 7.2.1.d) above). These data are available in different forms: images or templates, which are a mathematical representation of the salient features of an individual, used for recognition purposes. Biometric recognition technologies are used for verification and identification purposes. [185]

(298) According to Article 3(41) AI Act, a RBI system is

> *[a]n 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.*

(299) This definition covers only the identification functionality of biometric recognition systems, which implies the absence of active involvement of the persons concerned (i.e. no active participation) and results in the capture of the characteristics of those persons typically at a distance. For identification performance, the captured biometric data are compared with biometric data already stored in a reference database (such as a repository, e.g. a criminal database containing facial images or templates of suspects).

a) Identification purposes only

(300) The notion of 'biometric identification' is defined in Article 3(35) of the AI Act as

> *the automated recognition of physical, physiological and behavioural or psychological human features for the purpose of establishing the identity of a natural person by comparing biometric data of that individual to biometric data of individuals stored in a database.*

(301) Recital 15 AI Act further clarifies that such human features may comprise

> *the face, eye movement, body shape, voice, prosody, gait, posture, heart rate, blood pressure, odour, keystroke characteristics,*

(302) AI systems used for following natural persons can also be included in the definition of biometric identification, for example to see in which direction a suspect escapes. This can be concluded from Article 5(1)(h)(iii) AI Act, that allows for the localisation of suspects of crimes. The localisation is possible when a person is being followed.

(303) AI systems that are intended to be used for biometric verification fall outside the scope of the prohibition in Article 5(1)(h) AI Act.[186] Biometric verification (or authentication) consists of comparing data presented at a sensor with another set of previously recorded data stored on a device, such as a smartphone, a passport, or an ID card. The purpose of biometric verification is to verify that a specific person is who they claim to be.

[185] As defined by the biometrics community in ISO/IEC Standard 2382-37:2022 Information Technology - Vocabulary, Biometric recognition, Term 37.01.03.
[186] Recital 17 AI Act.

An example of biometric verification is the comparison of a traveller's face scanned at an e-gate with the facial image contained in their passport.

b) Remoteness

(304) According to Article 3(41) AI Act, remoteness implies the ability of biometric systems to identify individuals 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.

(305) The use of biometric systems to confirm the identity of a natural person for the sole purpose of having access to a service, unlocking a device, or having security access to premises is excluded from the concept of 'remote' (Recital 15 AI Act). This modality is used, for example, in access control.[187]

For example, a face identification system is deployed to enter a restricted area (e.g. power plant premises) through face scanning technology; the system compares the face of the individual presented at the entrance camera with a reference image contained in a reference database of persons allowed to enter the building.

(306) Recital 17 AI Act clarifies that this exclusion from the scope of the prohibition is justified by the fact that such systems are likely to have a minor impact on fundamental rights of natural persons as compared to RBI systems which may be used for the processing of the biometric data of a large number of persons without their active involvement. That recital further clarifies that RBI systems are typically used to perceive multiple persons or their behaviour simultaneously in order to facilitate significantly the identification of natural persons without their active involvement. For active involvement, it is not sufficient that persons are informed about the presence of cameras, but they need to step actively and consciously in front of a camera that is installed in a way fostering active participation.

For example,

- RBI systems that are used in cameras installed at walls or ceiling of metro stations for surveillance purposes. Such a system fulfils the condition of remoteness.

- Systems that are used to give access to the metro station, such as biometric metro tickets, where persons are actively involved and consciously approach the biometric sensor to obtain access, do not fulfil that condition.

[187] E.g. Ross A, Jain AK (2015) 'Biometrics, Overview' in Li S.Z. and Jain A.K. (eds) Encyclopedia of Biometrics, (1st ed. Springer Science, New York), pp. 289-294.

(307) Biometric recognition systems that process (contactless) fingerprints, gait, voice, DNA, keystrokes and other (biometric) behavioural signals may also constitute RBI systems.[188]

> For example:
>
> - A voice biometric technology system may be deployed to identify a person speaking. The microphone then collects the biometric sample.
> - A gait recognition system may be used via CCTV and the videos are automatically checked for matches with previously captured templates.
> - Keystroke biometric technology may be used to identify the person typing a fraudulent message.
>
> The fact that these systems are given as examples of RBI systems does not imply that they are prohibited under Article 5 AI Act.

(308) In the case of body-cams capable of RBI used by individual law enforcement agents, the untargeted filming during, for example, a demonstration with hundreds of participants will be considered to fulfil the condition of remoteness.

c) Reference database

(309) Identification is not possible without a reference database containing biometric data for comparison purposes. Thus, the existence of a reference database is **indispensable** to perform the comparison for identification purposes.[189]

> For example, in the case of missing persons, the Schengen Information System[190] database could be used as the reference database for facial recognition purposes (once operational).

9.2.2. Real-time

(310) Real-time means that the system captures and further processes biometric data 'instantaneously, near-instantaneously or in any event without any significant delay.'[191] All the processing steps, i.e. the capture, comparison, and identification of biometric data, occur simultaneously or almost simultaneously, which may include a 'limited short delay' to avoid the prohibition being circumvented through the retrospective use of RBI systems.[192] The notion of 'without a significant delay' is not defined in the AI Act; it will have to be assessed on a case-by-case basis. As the devices used for real-time or post-remote identification are increasingly one and the same with different

[188] EDPB-EDPS, Joint Opinion 5/2021, p. 11; Council of the European Union, 'Opinion of the Legal Service', 12302/22, 12 September 2022, paragraph 33, and Recital 15 AI Act.
[189] Recital 34 AI Act.
[190] Alerts on missing persons (Article 32 of the SIS II decision); Council Decision 2007/533/JHA of 12 June 2007 on the establishment, operation and use of the second generation Schengen Information System (SIS II)
[191] Recital 17 AI Act.
[192] Article 3(42) AI Act.

functionalities, the distinction is temporal. Generally speaking, a delay is significant at least when the person is likely to have left the place where the biometric data was taken.

(311) Real-time systems in general are used at a given place to facilitate a quick reaction and not to retrospectively identify persons. They provide the user of the system with a means to track the movements of persons under surveillance and to monitor them.

a) An AI system screens all incoming visitors to a concert venue: Real-time RBI
b) A system films all incoming visitors to a concert. An incident happens at the concert. After the concert, the identification system is operated on the video material in order to identify the offender: Post-RBI.

(312) When a law enforcement authority covertly takes a picture of a person via a mobile device and submits it to a database for immediate search, depending on the circumstances, this may fall under the prohibition of Article 5(1)(h) AI Act.

9.2.3. In publicly accessible spaces .

(313) Article 3(44) AI Act defines **publicly accessible spaces** as

> *any publicly or privately owned physical space accessible to an undetermined number of natural persons, regardless of whether certain conditions for access may apply, and regardless of the potential capacity restrictions.*

(314) Recital 19 AI Act lists several elements that characterise such spaces:

- Accessibility to an undetermined number of persons, independently of the potential capacity or security restrictions, such as purchasing a ticket or title of transport, prior registration or having a certain age. The possibility of getting access to a space through an unlocked door does not mean that the space is publicly accessible if indications or circumstances suggest the contrary (such as a sign restricting access). Moreover, the access to a space can be limited to certain persons, as defined by law, linked to public safety or security, or to the decision of the person having the relevant authority over the space.

For example, publicly accessible spaces are in principle:

- a concert venue for which participants pay an entrance fee.
- an event location where a trade fair is organised targeting participants over the age of 50.

A space closed by a gate, even if the gate is unlocked, such as the gated entrance of a fenced residential area of several houses, will normally not be considered a publicly accessible space. By contrast, a park in a gated residence with public opening hours without any access restrictions during those hours will generally constitute a publicly accessible space during those hours and a closed space outside those hours.

- Irrelevance of ownership, i.e. a space does not need to be in public ownership for it to be considered as a publicly accessible space.

For example, the space may be owned by a private entity, a public entity, or a public entity and managed by a private party, without impacting the nature of the space.

- No specific activity for which the space is used; a publicly accessible area is not necessarily a space linked to a public service. Moreover, a space linked to a public service may include non-publicly accessible spaces, i.e. the offices of the civil servants working at a municipality.

For example, publicly accessible spaces may be used for commerce, such as shops, restaurants, cafés, etc.; for services, such as banks, professional activities (a doctor's office as well as an accountant's office), hospitality (e.g. a hotel), etc.; for sport, such as swimming pools, gyms, stadiums, etc.; for transport, such as bus, metro and railway stations, airports, means of transport, etc.; for entertainment, such as cinemas, theatres, museums, concert and conference halls, etc.; or for leisure or otherwise, public roads and squares, parks, forests, playgrounds.[193]

(315) The following spaces do not constitute publicly accessible spaces within the meaning of Article 5(1)(h) AI Act:

- online spaces, since they do not constitute a physical space within the meaning of Article 3(44) AI Act.

For example, chat rooms, social media, online platforms, etc., are therefore excluded from the scope of the prohibition.

- Certain spaces meant to be accessed by a limited number of persons, such as factories, companies and workplaces with entry control or limitations to relevant employees or service providers, since they are intended to be accessed only by relevant employees and service providers[194].

For example, a workplace accessible with a badge is in principle not considered a publicly accessible space, whereas an office without entry controls falls may be.

- **prisons and border control** are not publicly accessible spaces[195]

(316) For example, a border crossing point is not a publicly accessible space, while the street leading to the border crossing point or a forest in the vicinity normally is.

[193] Recital 19 AI Act.
[194] Recital 19 AI Act.
[195] Recital 19 AI Act. In a different context, border control has been defined as the activity carried out at a border, in accordance with and for the purposes of Regulation (EU) 2016/399 (Schengen Borders Code), in response exclusively to an intention to cross or the act of crossing that border. This does not comprise the so-called border area, which may extend to a maximum of 50 kilometres on either side of the border.

(317) Some spaces can have a dual function. For example, an airport is generally considered a publicly accessible space as regards its common areas, but the area dedicated to border control (where the customs officials stand and passports or ID checks occur) is excluded from the scope of the prohibition.

(318) As clarified in Recital 19 AI Act, assessing whether a space is accessible to the public should be done based on a case-by-case analysis.

9.2.4. For law enforcement purposes

(319) The prohibition in Article 5(1)(h) AI Act applies to the use of RBI systems for law enforcement purposes, irrespective of the entity, authority, or body carrying out the law enforcement activities.

(320) Law enforcement is defined in Article 3(46) AI Act as the '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 safeguarding against and preventing threats to public security.' These purposes are the same as those listed in Article 1 LED.[196] Thus, any interpretation of those purposes in relation to LED may also be relevant for the purpose of interpreting the notion of 'law enforcement' used in the AI Act.

(321) Law enforcement purposes comprise the investigation, detection, and prosecution of criminal offences. They also comprise activities in relation to the prevention of criminal offences, including safeguarding against and the prevention of threats to public security, before any crime has actually been committed. For instance, the police may take 'coercive measures at demonstrations, major sporting events or riots' in the context of crime prevention.[197] Finally, those activities comprise the execution of penalties, such as the execution of sentences.

(322) According to Article 3(46) AI Act, law enforcement activities may be performed by law enforcement authorities or on their behalf. Law enforcement authorities are further defined in Article 3(45) AI Act in the same manner as national competent authorities are defined in the LED.[198] That definition covers law enforcement authorities and entrusted bodies or entities (which may be private parties):

> *(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*

[196] Some activities of law enforcement authorities are excluded from the scope of the LED, such as when they perform administrative tasks (such as human resources), these activities are carried outside the law enforcement framework. They fall under the GDPR. See Recital 19 GDPR.
[197] Recital 12 LED.
[198] Article 3(7) LED.

For example, such public authorities include police authorities and criminal justice authorities (such as prosecutors) when they carry out a law enforcement task.

(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;

(323) Under the AI Act other entities, bodies, or persons, may exercise law enforcement activities after being entrusted by Member States law entrusting them public authority and public powers for the purposes listed above.

(324) 'On behalf of' means that a law enforcement authority has delegated the performance of a law enforcement activity (or part of it) to another entity or person, including private parties, or has requested in specific cases another entity or person to act to support law enforcement activities. In both cases, the law enforcement authorities must instruct on all major aspects and supervise the other entity, as this requirement is inherent to the notion of acting "on behalf" of a person.

Delegation of tasks to other bodies may include, for example,

- Public transport companies requested by law enforcement authorities to ensure security on the public transport networks under their instructions and supervision.
- Sports federations requested by law enforcement authorities to act under their instructions and supervision to provide security at sporting events.
- Banks that are requested by law enforcement authorities to conduct certain actions to 'counter certain crimes in specific cases' under the instructions and supervision of law enforcement authorities.

These activities fall within the definition of "for the purpose of law enforcement" since those entities act 'on behalf' of law enforcement authorities. If those entities act on their 'own behalf' when detecting and countering crimes (such as fraud, money laundering), they will not be considered to fall under the prohibition of Article 5(1)(h) AI Act.

(325) Only when those other bodies or entities have been entrusted with a specific law enforcement task will their activities fall under the definition of 'law enforcement'.

9.3. Exceptions to the prohibition

(326) The AI Act provides three exceptions to the general prohibition on the use of real-time RBI in publicly accessible spaces for law enforcement purposes. Article 5(1)(h)(i) to (iii) AI Act exhaustively lists three objectives for which real-time RBI may be

authorised, while Article 5(2) to 5(7) AI Act lays down the conditions and safeguards for such authorisation. Article 5(1)(h)(i)-(iii) AI Act does not in itself constitute a legal basis for the real-time use of RBI systems in publicly accessible spaces. Rather, only a domestic Member State law that fulfils, in particular, the requirements in Article 5(2)-(7) AI Act can allow the use of real-time RBI, as provided by Article 5(2) AI Act. Consequently, in the absence of Member State legislation authorising the use of real-time RBI for one or more of those objectives, such use is prohibited as from 2 February 2025.

9.3.1. Rationale and objectives

(327) The objectives set out in Article 5(1)(h)(i)-(iii) AI Act aim to allow the use of certain AI and investigative tools for law enforcement purposes. These objectives are:

(i) the targeted search of victims of three specific serious crimes and missing persons [protection];

(ii) the prevention of imminent threats to life or physical safety or a genuine threat of terrorist attacks [prevention]; and

(iii) the localisation and identification of suspects and offenders of certain serious crimes as listed in Annex II [prosecution/investigation].

(328) In those scenarios, the Union legislature has balanced the security needs of society against the risk that real-time RBI systems pose to the fundamental rights of individuals subject to those systems. According to Recital 33 AI Act, the objectives for which the use of real-time RBI systems for law enforcement purposes in publicly accessible spaces is allowed must be strictly, exhaustively, and narrowly defined, and appear when there is a 'strict necessity' to achieve 'a substantial public interest' which 'outweighs the risks' posed to fundamental rights. Any other use of real-time RBI systems in publicly accessible spaces for law enforcement purposes which is not listed in Article 5(1)(i)-(iii) AI Act is prohibited.

For instance, the use of real-time RBI systems by the police to identify a shoplifter and compare their facial images against criminal databases is prohibited, as it does not fall under any of the objectives listed in Article 5(1)(h)(i)-(iii) AI Act.

9.3.2. Targeted search for the victims of three serious crimes and missing persons

(329) According to Article 5(1)(h)(i) AI Act, the use of real-time RBI in publicly accessible spaces for law enforcement purposes is allowed, subject to strict necessity and the conditions in Article 5(2)-(7) AI Act, for the targeted search of victims of abduction, trafficking in human beings, or sexual exploitation of human beings, as well as the search for missing persons.

a) Targeted search for victims of three types of crimes

(330) The scenario described in Article 5(1)(h)(i) AI Act seeks to assist law enforcement authorities to search for victims of three serious crimes.

(331) A targeted search would involve the localisation and identification of victims.

Three types of crimes

(332) The targeted search for specific victims of three serious crimes are covered by the scenario listed in Article 5(1)(h)(i) AI Act: the abduction of, trafficking in, and sexual exploitation of human beings[199].

> If, for example, a child is kidnapped and there are concrete indications that the kidnapper intends to bring the child from one place to another by car, the police may use a real-time RBI system for the targeted search of that child, but it must define a perimeter of deployment and duration of use to identify the child.

b) Searching for missing persons

(333) The first scenario also covers the search for a missing person.[200]

(334) A distinction may be made between missing children and missing adults, since the voluntary disappearance of a missing adult will not always trigger a search. The applicable rules regarding missing children vary considerably from one Member State to another.[201] In any event, Article 5(1)(h)(i) AI Act only allows the use of a real-time RBI system to search for missing persons for law enforcement purposes.

(335) The disappearance of an adult does not always lead to a search of that person by police, as adults have the right to disappear. A search could be linked to the legal status of the person ('under curatorship'), their health condition (a mental illness), the existence of a suicidal note, but also the departure without personal belongings. If the circumstances of the disappearance are a cause for concern, the disappearance may be filed with the police so that a search can start.

(336) In some Member States, the search for a missing person may occur under an administrative procedure and not for law enforcement purposes. For example, where a vulnerable person is missing, but there is no suspicion of a crime or any other law enforcement purpose, the use of real-time RBI systems to search for that person would

[199] Kidnapping, trafficking in human beings, and sexual exploitation are three crimes that can trigger a European Arrest Warrant (EAW) to arrest and transfer a criminal suspect or a sentenced person to the country that issued the EAW. The three crimes relate mostly but not exclusively to women and children. According to the European Commission's DG Migration and Home Affairs, almost 40 percent of the victims are EU citizens, and most of them are women and children trafficked for sexual exploitation. The number of men victims has nearly doubled in ten years. They are trafficked for forced labour and forced begging, while most of the women and children are trafficked for sexual exploitation. https://home-affairs.ec.europa.eu/policies/internal-security/organised-crime-and-human-trafficking/together-against-trafficking-human-beings_en

[200] A 'missing person' is not defined at EU level. But in Council Conclusions of December 2021 on 'Stepping Up Cross-Border Police Cooperation in the area of Missing Persons', the Council takes as reference both the definition of a missing person in the Council of Europe's Recommendation CM/Rec (2009) 12 and in national regulations. Council Conclusions (2021) 14808/21, para 11, page 4.

[201] European Commission, European Migration Network, 'How do EU Member States treat cases of missing unaccompanied minors?' EMN Inform, 2020.

not be deemed to be for law enforcement purposes and would therefore fall under the rules for such use under the GDPR.

9.3.3. Prevention of imminent threats to life or terrorist attacks

(337) Article 5(1)(h)(ii) AI Act lists the second scenario in which the use of real-time RBI in publicly accessible spaces for law enforcement purposes is allowed, subject to strict necessity and the conditions established in Article 5(2)-(7) AI Act

> *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.*

(338)

a) Specific, substantial and imminent threat to life or physical safety of natural persons

(339) In application of Article 2 of the Charter, which guarantees the right to life, the Union and its Member States must safeguard and, thus, protect the lives of individuals. The criteria in Article 5(1)(h)(ii) AI Act concerning the threat to life to allow for the use of real-time RBI systems in publicly accessible spaces require the existence of (1) a specific, (2) substantial and (3) imminent threat to the life or physical safety of (4) natural persons. The threat does not need to be limited to identified individuals or a group, as it relates to natural persons in general.

(340) Recital 33 AI Act clarifies that an imminent threat to life or the physical safety of natural persons may also include an imminent threat to critical infrastructure[202] 'where the disruption or destruction of such critical infrastructure would result in an imminent threat to the life, physical integrity or safety of a person, including through serious harm to the provision of basic supplies to the population or to the exercise of the core functions of the State.'

For example,[203] A serious disruption and destruction of critical infrastructure (e.g., a power plant, water supply, or a hospital) may result in an imminent threat to the life or the physical safety of a person when there is serious harm of cessation of basic supplies to the population (deprivation of electricity or drinkable water for a long period, in a particularly warm or cold weather, etc).

(341) What constitutes an imminent threat to life or the physical safety of natural persons is ultimately defined and assessed at the level of the Member State based on its national laws, in accordance with EU law, in particular taking into account the key elements and rationale of Article 5 AI Act. This will have to be laid down/referred to in the laws

[202] As defined in Article 2(4) of Directive 2022/2557.
[203] Recital 33 AI Act.

Member States must adopt to make use of the exceptions to the prohibition on the use of real-time RBI for law enforcement purposes in publicly accessible spaces.

(342) An **imminent** threat to life or physical safety is a threat that can occur at any moment and requires '**immediate action to be taken.**'[204] **A substantial** threat to physical safety relates to serious bodily injuries.

(343) A specific threat means that the threat is clearly defined, individualised and concrete, in that it should not be hypothetical or relate to certain dangers in general.

> For example, the police are informed that a former student plans a deadly attack at his former university as he seeks revenge on several former classmates. The police receives information about the imminence of the attack, the targeted school, and the weapons he plans to use to execute his plans.

(344) A specific threat needs not be intentional. Non-intentional actions could also result in a threat to life or physical safety.

b) A genuine and present or genuine and foreseeable threat of a terrorist attack

(345) This part of the second scenario described in Article 5(1)(h)(ii) AI Act is comprised of several elements: the existence of a threat of a terrorist attack and the characteristics of the threat, which must be genuine and present or genuine and foreseeable.

Threat of a terrorist attack

(346) The assessment concerning the existence and seriousness of the threat is made at national level when assessing the actual circumstances of a measure to be taken to safeguard national security, and more specifically, in case of a terrorist attack. The **terrorist threat level is defined at national level** and varies from one Member State to another. For example, the Netherlands has established five levels of threats,[205] Belgium four,[206] France three, [207] and Sweden five.[208] However, the concept of 'a genuine and present or genuine and foreseeable threat', as used in Article 5(1)(h)(ii), is an autonomous notion of Union law and should therefore be assessed, in principle, independently of national definitions. The threat relates not to terrorism in general, but specifically to a threat of a terrorist attack.

Characteristics of the threat: genuine and present or genuine and foreseeable

(347) The threshold of seriousness that a threat needs to reach to allow for the use of real-time RBI systems in publicly accessible spaces for law enforcement purposes was

[204] Recital 37 of Regulation 2023/1543.
[205] https://www.government.nl/topics/counterterrorism-and-national-security/risk-of-an-attack-threat-level
[206] https://cuta.belgium.be
https://crisiscenter.be/en/risks-belgium/security-risks/terrorism-and-extremism
[207] https://www.sgdsn.gouv.fr/vigipirate#
https://www.sgdsn.gouv.fr/files/files/Vigipirate/20160130-np-sgdsn-pse-tackling-terrorism-together.pdf
[208] https://www.krisinformation.se/en/hazards-and-risks/terrorism

inspired by the CJEU's case law on data retention and passenger name record measures aimed at safeguarding national security, in particular, against terrorist attacks. According to the CJEU, in those contexts, 'a threat to national security must be genuine and present, or at the very least, foreseeable, which presupposes that sufficiently concrete circumstances have arisen.'[209]

Prevention

(348) Contrary to Article 5(1)(h)(i) and Article 5(1)(h)(iii) AI Act, the scenario described in Article 5(1)(h)(ii) does not specify that the use of real-time RBI is permitted to locate or identify a concrete person. Its purpose is the prevention of a particular threat. Accordingly, the scenario may also cover the use of real-time RBI to detect and follow 'terrorists on the move', i.e. several persons linked to the same threat, if there are concrete indications that those persons plan to commit a terrorist attack, but it is not clear where.

Real-time RBI to prevent a terrorist attack in a park The police are informed that a person is running around a park looking for people to attack with a knife while he is screaming violent extremist slogans usually linked to terrorist attacks and terrorist groups. If the Member State has authorised the use of real-time RBI in the scenario described by Article 5(1)(h)(ii) AI Act, law enforcement authorities may use real-time RBI to identify and locate the person in and around the park to prevent the attack, provided the other conditions of Article 5(2)-(7) AI Act have been met.

9.3.4. Localisation and identification of suspects of certain crimes

(349) Article 5(1)(h)(iii) AI Act allows the real-time use of RBI in publicly accessible spaces for 'the localisation and identification of a person suspected of having committed a criminal offence, for the purpose of conducting a criminal investigation or prosecution or executing a criminal penalty for offences referred to in Annex II and punishable in the Member States concerned by a custodial sentence or a detention order for a maximum period of at least four years.'

Annex II AI Act provides an exhaustive list of serious crimes for which the use of real-time RBI may be authorised for the aforementioned objective. Those criminal offences are: - terrorism, - trafficking in human beings,

[209] Judgment of the Court of Justice of 20 September 2022, *SpaceNet*, C-793/19 (Joined Cases C-793/19, C-794/19), ECLI:EU:C:2022:702, paragraph 93.

- sexual exploitation of children, and child pornography,
- illicit trafficking in narcotic drugs or psychotropic substances,
- illicit trafficking in weapons, munitions or explosives,
- murder, grievous bodily injury,
- illicit trade in human organs or tissue,
- illicit trafficking in nuclear or radioactive materials,
- kidnapping, illegal restraint or hostage-taking,
- crimes within the jurisdiction of the International Criminal Court,
- unlawful seizure of aircraft or ships,
- rape,
- environmental crime,
- organised or armed robbery,
- sabotage,
- participation in a criminal organisation involved in one or more of the offences listed above.

a) Localisation and identification

(350) A Member State may authorise the use of real-time RBI in publicly accessible spaces for law enforcement purposes to locate and identify a suspect of a criminal offence to conduct a criminal investigation, prosecute that person for the crime committed, or execute an existing sentence.

b) Suspects and Perpetrators

(351) Article 5(1)(h)(iii) AI Act covers two categories of individuals: suspects and perpetrators. A suspect is a person with regard to whom there are serious grounds for believing that they have committed a criminal offence, and sufficient evidence of that person's involvement in the offence has already been gathered. A perpetrator is a person who is accused or convicted of having committed a criminal offence. The same conditions (crime listed in Annex II and maximum punishment of at least 4 years) apply to locate or identify the accomplice of the crimes listed in Annex II AI Act.

c) List of serious crimes

(352) Only serious crimes justify the use of real-time RBI systems in publicly accessible spaces for law enforcement purposes.

(353) The first five offences listed in Annex II AI Act are the same as the 'euro crimes; listed in Article 83 TFEU, while the other offences constitute priorities for law enforcement

cooperation.[210] Some of them (e.g. kidnapping, illicit trafficking in nuclear or radioactive materials) may be linked to terrorism.[211]

(354) Although all the criminal offences listed in Annex II may trigger the issuance of a European Arrest Warrant ('EAW') against a suspect or perpetrator, the use of real-time RBI to locate and identify a suspect for one of these serious criminal offences does not require that an EAW has been issued.

(355) Moreover, to use real-time RBI for this purpose, the respective criminal offence must be punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least four years.

During a busy festival in a city, police authorities deploy live facial recognition technologies to monitor the area around the festival and identify wanted individuals with outstanding arrest warrants for illegal drug trafficking and sexual offences. At different entrances to the festival, the police use live video footage of people passing in front of a mobile camera to compare their faces with a watchlist of faces of wanted individuals.

First, concerning the offence types, RBI can be used in case of illegal drug trafficking. However sexual offences are not on the list of offences, unless they relate to the sexual exploitation of children, child sexual abuse material, or rape. The police are not allowed to deploy real-time facial recognition technologies in a broad, untargeted manner, i.e. in the hope of finding wanted criminals and taking them off the streets.

The case is different if the police have received a physical description with a photograph of a wanted individual that is subject to a European Arrest Warrant for drug trafficking and they have reasons to believe that he will be present at the festival. In those circumstances, deploying real-time facial recognition technologies to identify a targeted individual may be covered by Article 5(1)(h)(iii) AI Act.

After a serious terror attack at a Christmas market with 12 deaths, the police uses real-time facial recognition technologies to identify the offender and to see to where he is escaping. In that context they also use the real-time facial recognition technologies of the nearby train station and at destination stations of the trains leaving from there shortly after the attack. In the case of a terror attack, such use can be permitted under Article 5(1)(h)(iii) AI Act.

(356) A link between Article 5(1)(h)(i) and Article 5(1)(h)(iii) AI Act may be made for the crimes covered by the scenario described in Article 5(1)(h)(i) AI Act. While real-time RBI systems may be deployed to find a victim or a missing person, those systems may

[210] Europol priorities.
[211] See Recital 33 and see definition of terrorist offences in Article 3 of Directive 2017/541.

also be used to locate and identify the perpetrator or suspect of trafficking in human beings, sexual exploitation as far as it concerns children (as listed in Annex II), and kidnapping (as far as the abduction mentioned in Article 5(1)(h)(i) AI Act qualifies as kidnapping as listed in Annex II AI Act). A link may also be made between Article 5(1)(h)(ii) and (iii) AI Act: real-time RBI systems may be used to prevent a threat falling within the scope of Article 5(1)(h)(ii) and, if that threat materialises, those systems may be used to identify/locate the perpetrator 'on the move'.

10. SAFEGUARDS AND CONDITIONS FOR THE EXCEPTIONS (ARTICLE 5(2)-(7) AI ACT)

10.1. Targeted individual and safeguards (Article 5(2) AI Act)

Article 5(2) AI Act provides:

'The use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement for any of the objectives referred to in paragraph 1, first subparagraph, point (h), shall be deployed for the purposes set out in that point only to confirm the identity of the specifically targeted individual, 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 that would be caused if the system were not used;

(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 purposes of law enforcement for any of the objectives referred to in paragraph 1, first subparagraph, point (h), of this Article shall comply with necessary and proportionate safeguards and conditions in relation to the use in accordance with the national law authorising 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 be authorised only if the law enforcement authority has completed a fundamental rights impact assessment as provided for in Article 27 and has registered the system in the EU database according to Article 49. However, in duly justified cases of urgency, the use of such systems may be commenced without the registration in the EU database, provided that such registration is completed without undue delay.'

(357) The use of real-time RBI systems for one of the objectives listed in Article 5(1)(h)(i) to (iii) AI Act is subject to certain safeguards and conditions, which are detailed in Article 5(2) to Article 5(7) AI Act.

(358) First, the use of real-time RBI systems in publicly accessible spaces for law enforcement purposes is only allowed to 'confirm the identity of the specifically

targeted individual'. This first condition aims to balance the seriousness of the situation and the harm resulting from not using the system with the impact of the technology on individuals' rights and freedoms. It aims to avoid mass surveillance by targeting an individual for the deployment of real-time RBI. As a consequence, the deployment of a real-time RBI system in publicly accessible spaces for law enforcement purposes should only be authorized for targeted individuals.

(359) The use of the expression 'confirming the identity', as opposed to 'identification', is meant as an additional safeguard for the fundamental rights limiting the risk of indiscriminate surveillance and implies that the identification of an individual within the meaning of Article 5(1)(h) AI Act must be targeted. That expression should be understood as meaning that the use of real-time RBI may only be initiated to search for specific individuals for which the law enforcement authorities have reasons to believe or are informed that they are victims of the crimes listed in Article 5(1)(h)(i) AI Act or are involved in one of the scenarios described in Article 5(1)(h)(ii) or Article 5(1)(h)(iii) AI Act. This means, in practice, a comparison of the data collected real-time with the data contained in the reference database. As regards the use of real-time RBI system in the scenarios described in Article 5(1)(h)(ii) AI Act and for conducting a criminal investigation within the meaning of 5(1)(h)(iii) AI Act, law enforcement authorities do not necessarily need to know the identity of the individuals they are searching for before using the system. If they have factual indications and information about a planned terrorist attack by a terrorist group (without knowing who will execute the plan) at a specific time and place, the RBI system may be used to identify the offender from the terrorist group, provided the law enforcement authorities have constituted a reference database containing the biometric data of the individuals forming part of the terrorist group. In all three scenarios described in Article 5(1)(h)(i) to (iii) AI Act, 'confirming the identity' may also include the localisation of the person in question.

(360) Second, before using the system, the nature of the situation giving rise to the possible use, in particular, the seriousness, probability and scale of the harm for natural persons, society and law enforcement purposes that would be caused if the system were not used, should be assessed against the consequences of the use of the system on the rights and freedoms of the persons concerned, in particular, the seriousness, probability and scale of those consequences. This should include evaluating whether less intrusive alternative solutions are available to the law enforcement authorities or entities acting on their behalf.

For example, law enforcement authorities are prohibited from using real-time facial recognition systems in the street based on general security, crime prevention and over-crowding concerns, since that would involve the constant monitoring and surveillance of all persons, it is not limited in time, and it would therefore not meet the criteria for the exception from the prohibition laid down in Article 5(1)(h) AI Act.

(361) The '**seriousness**' criterion, applied here in connection to the possible harm and consequences, implies a variation in degrees of interference with the fundamental rights at stake, which is linked to the principle of proportionality.[212] Concerning the interferences with fundamental rights, some interferences are viewed as more serious than others.

(362) The '**scale**' criterion refers, in particular, to the number and categories of persons affected by the interference (including children and vulnerable or marginalised persons).

(363) Finally, the '**probability**' is the likelihood that an event will occur.

(364) The assessment of the seriousness, scale and probability of the harm and consequences should all be part of the Fundamental Rights Impact Assessment that the law enforcement authority is obliged to complete (see below)**.** That assessment will be concluded on a case-by-case basis.

(365) Third, the real-time use of RBI should be clearly limited in terms of geographic scope, duration, and the targeted person. This is to ensure that the RBI system is only used when strictly necessary.

(366) Concerning the **geographic restriction**, it may cover one or several geographical areas based on 'objective and non-discriminatory factors'. In the case of biometric identification this implies that the geographic restriction applies to a clearly delineated boundary for which there are indications that the event will take place. Such a delineation should -under normal circumstances- not comprise an entire city or country, but should be more targeted.

(367) Another safeguard relates to the **personal scope** of the measure, i.e. defining the **categories of persons concerned.** This would exclude the untargeted, indiscriminate identification of persons, without further indications of an incident.

(368) Finally, the **time limit** is a period limited to what is strictly necessary, but which may be extended in case of need in accordance with the applicable rules. The use of real-time RBI systems therefore cannot be for an indefinite or vague period of time. The period needs to be determined in light of the concrete indications that lead to the use of RBI systems.

(369) Fourth, before deployment, the law enforcement authority deploying the real-time RBI system must have conducted a Fundamental Right Impact Assessment (FRIA) and registered the system in the EU database (except in a duly justified case).

10.1.1. Fundamental Rights Impact Assessment

[212] Judgment of the Court of Justice of 2 October 2018, *Ministerio Fiscal*, C-207/16, ECLI:EU:C:2018:788, paragraph 55, where the Court states that 'access must be proportionate to the seriousness of the interference with the fundamental rights in question'.

(370) FRIAs carried out in application of Article 5(2) AI Act must comply with the conditions laid down in Article 27 AI Act. That provision sets out the requirements concerning FRIAs applicable to high-risk AI systems.

(371) In the period between when the prohibitions in Article 5 AI Act become applicable (after 2 February 2025) but the provisions on high-risk AI systems are not yet applicable (before 2 August 2026), the requirements for FRIA set out in Article 27 AI Act should be implemented by the deployers of real-time RBI systems meeting the conditions to benefit from one or more of the exceptions in Article 5(1)(h) AI Act. The following provisional guidance relates only to the use of real-time RBI in publicly accessible spaces for law enforcement purposes for the period before the obligations for high-risk AI systems become applicable and the Commission adopts the template for FRIA and provides further guidance on the obligation under Article 27 AI Act.

(372) A FRIA is a new type of impact assessment that aims to identify the impact that certain high-risk AI systems, including RBI systems, may produce on fundamental rights. A FRIA is an accountability tool. The FRIA does not replace the existing Data Protection Impact Assessment (DPIA) that data controllers (i.e. those responsible for the processing of personal data) must perform under Article 27 LED, Article 35 GDPR or Article 39 EUDPR.

> For example, a DPIA must be conducted when biometric data are processed through new technologies likely to result in a high risk to the rights and freedoms of natural persons (such as CCTV, AI facial recognition, and body-worn cameras) in publicly accessible spaces.

(373) Whereas a DPIA focuses on the risks to the rights and freedoms of individuals resulting from the processing of their personal data, a FRIA covers the possible impact of AI systems on individuals' fundamental rights more generally. The scope of a FRIA is therefore broader in terms of activities covered and fundamental rights assessed. Where personal data are processed by the AI system (which is the case of RBI systems), the FRIA should complement the DPIA performed by the deployer as data controller,[213] without covering aspects already addressed in the DPIA and avoiding overlaps. The analysis of the FRIA in these Guidelines is limited to the authorized use of RBI in real-time and is aimed to serve as a preliminary guidance for deployers in this interim period before the AI Office provides a template.[214]

(374) The obligation to carry out the FRIA under Article 5(2) AI Act is imposed on the deployers of the RBI system, and not entities or bodies or anyone else acting on their behalf. If other actors are acting on behalf of the deployer/the law enforcement authority, they will have to contribute to the preparation of the FRIA with all relevant information to ensure it is properly carried out.

[213] Article 27(4) AI Act.
[214] Thus, the analysis does not cover the case of high-risk AI systems in general.

(375) A FRIA must be carried out before the deployment of the authorized real-time RBI system.

(376) According to Article 27 AI Act, a FRIA should include the following information:

- A description of the RBI use and the deployer's processes for the use, together with the intended purpose of use:

The description should include: - the name of the deployer; - the law enforcement purpose(s) for which the real-time RBI system will be used; - the description of the reference database against which the biometric identification will be compared, including the sources of the biometric data (facial images, voice samples, etc.) that will be used; - the description of the technology underlying the system to explain its functioning (by referencing the available documentation provided by the provider and its name)[215]; - the legal basis on which the real-time RBI will be deployed.

- The period of use and frequency of use

Each individual use of a real-time RBI system for one of the permitted exceptions must be authorised prior to its deployment by a judicial authority or other independent authority under Article 5(3) AI Act. By contrast, for the FRIA, deployers must provide a general indication of the intended period of use and the expected frequency.

- The categories of persons and groups affected by the system

For the purpose of the exception in Article 5(1)(h) AI Act, the FRIA should distinguish between: - The targeted individual, who may be a victim of a crime, a perpetrator, or a suspect, - The individuals whose biometric data are included in the reference database, and - Categories of persons who are present in the surrounding areas where the RBI system will be deployed. The use of real-time RBI systems will not only affect the fundamental rights of the targeted individual. The rights of other individuals whose biometric data are used for comparison purposes, passersby, and people incidentally presented in the search area will also be affected. The description of the geographic scope of the search area(s) covered by the real-time RBI system will impact the number of persons affected by the system.

[215] Once the rules on high-risk AI system enter into force, this can be done by reference the registration number of the system in the EU database and the available information for the system contained therein.

- The specific risks of harm to the affected persons:

(377) The fundamental rights that may be affected by the use of real-time RBI in publicly accessible spaces for law enforcement purposes include, in particular:

- the right to private and family life, including people's reasonable expectation of anonymity in public spaces;
- the right to data protection, as RBI systems rely on the processing of biometric data and other personal data (e.g. names, ID numbers, as well as sensitive data such as ethnicity) to identify specific individuals;
- freedom of thought, conscience and religion, freedom of expression and freedom of assembly and association in the public spaces being searched, on which the use of RBI systems could have a chilling effect, preventing individuals to fully exercise their rights and freedoms, since if individuals know that they are monitored, they might change their behaviour, or even prevent themselves to behave in a certain manner;
- the right to an effective remedy and a fair trial;
- the right to non-discrimination if the system embeds biases (such as gender, ethnic or racial biases) and leads to the misidentification of a suspect or perpetrator;
- the right to human dignity by the feeling of being reduced to an object of the system;
- the presumption of innocence and right to defence; since no decision adversely affecting an individual may be solely taken on the output of the real-time RBI system.
- the rights of the child in case the victim, missing person or suspect is a minor;
- the rights of the elderly in the case of a missing person.

To assess the specific risks of harm likely to impact the identified affected person(s) or group(s), the FRIA must identify the fundamental rights of those persons and assess the impact on their fundamental rights, including the severity of the impact and its scale, taking into account the potentially affected persons.

This part of the FRIA should also include the assessment of whether the use of a real-time RBI system is necessary and proportionate considering the objectives and the circumstances in which its use is intended, including the existence or absence of less intrusive alternatives. The FRIA should describe the performance and accuracy level of the system, based on the technical documentation and, if available, the training data on which the technology was tested and developed to prevent biases and discrimination.

The FRIA should also identify the impact of use of a real-time RBI system on the fundamental rights of all individuals potentially affected, in particular the suspect or perpetrator, the victim searched and other individuals present in the publicly accessible spaces subject to the search. To the extent that the system processes the biometric data of these individuals, their rights to private and family life and data protection will be impacted, which will be assessed as part of the DPIA as far as the data processing

activities are concerned. For other activities related to the use of the real-time RBI systems and the impact on other fundamental rights, the FRIA will complement the DPIA. Depending on the context of deployment, other fundamental rights of these individuals, such as their rights to human dignity, freedom of thought, conscience and religion, assembly or freedom of expression, rights to an effective remedy and a fair trial, presumption of innocence and right of defence, rights of the child, may be impacted.

The assessment under the FRIA should be performed at an abstract level, prior to the first putting into service of the AI system. Specific context-dependent considerations that determine the impact of use in each individual case where a real-time RBI system is used should be further elaborated in the individual request for requesting the authorisation by a judicial authority or other independent administrative authority of each use of the RBI system (see section 10.23.8.3. below).

- Human oversight measures

According to Article 5(3) AI Act, no decision that would adversely affect an individual may be taken solely on the basis of the output of the real-time RBI system. As a consequence, the FRIA should describe the procedures that will be followed during the operation of the system and how the output will be interpreted in the context of decision-making process. The procedures should provide instructions on the deployment of the RBI system, clarify the role of a human agent in verifying and interpreting the output and provide training to operate the system. The person in charge of human oversight should have sufficient 'AI literacy, training and authority'[216] to understand how the system functions and when it underperforms or malfunctions.

Other considerations for human oversight and monitoring under Articles 14 and 26 AI Act are also relevant and should be described.

- Risk mitigation measures

Beyond implementing human oversight measures (including to avoid discriminatory measures), the deployer should explain redress measures in case a risk materialises, including the governance procedures and the complaint mechanisms (such as in the case of a misidentification).

10.1.2. Registration of the authorized RBI systems

(378) Article 5(2) AI Act also obliges the deployer of a real-time RBI system used in publicly accessible spaces for law enforcement purposes to register the system in the EU database provided for in Article 49 AI Act. However, in cases of a duly justified emergency (such as an imminent threat), deployment may start even prior to registration, provided the law enforcement authority registers the system without undue delay. Undue delay should be understood as meaning 'as soon as possible' considering

[216] Recital 91 AI Act.

the circumstances of the emergency that prevented the registration of the system prior to its use. Whether registration meets that criterion requires an appreciation on a case-by-case basis. It cannot be defined a priori with a precise time limit. The delay should not be caused by a deliberate action. According to Article 49(4) AI Act, RBI systems used for the purpose of law enforcement will be registered in a secure non-public section of the database, with limited information and limited access to that information.

For example, requesting law enforcement authorities to register the RBI system within 24 hours of the use might be considered a reasonable delay where the system was deployed in a situation of an imminent threat to life, such as in the scenario of a live shooter.

10.2. Need for prior authorisation

(379) Article 5(3) AI Act requires **prior authorisation** of each individual use of a real-time RBI system and prohibits automated decision-making based **solely on the output of such a system** which produces an adverse legal effect.

Article 5(3) AI Act provides:
For the purposes of paragraph 1, first subparagraph, point (h) and paragraph 2, each use for the purposes 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 5. However, in a duly justified situation of urgency, the use of such system may be commenced without an authorisation provided that such authorisation is requested without undue delay, at the latest within 24 hours. If such authorisation is rejected, the use shall be stopped with immediate effect and all the data, as well as the results and outputs of that use shall be immediately discarded and deleted.
The competent judicial authority or an independent administrative authority whose decision is binding shall grant the authorisation only where it is satisfied, on the basis of objective evidence or clear indications presented to it, that the use of the 'real-time' remote biometric identification system concerned is necessary for, and proportionate to, achieving one of the objectives specified in paragraph 1, first subparagraph, point (h), as identified in the request and, in particular, remains limited to what is strictly necessary concerning the period of time as well as the geographic and personal scope. In deciding on the request, that authority shall take into account the elements referred to in paragraph 2. No decision that produces an adverse legal effect on a person may be taken based solely on the output of the 'real-time' remote biometric identification system.

10.2.1. Objective

(380) The objective for requiring prior authorisation ('authorisation *ex ante')* for any use of a 'real-time' RBI system in publicly accessible spaces for law enforcement purposes is the need for an assessment and a decision as to whether any envisaged use of such a system for such purposes is:

- necessary and proportionate to achieve any one of the objectives listed in Article 5(1)(h)(i) to (iii), i.e., for the targeted search of specific victims, the prevention of specific threats, or the localisation or identification of offenders;

and

- limited to what is strictly necessary concerning the time period and the geographic and personal scope.

(381) The consequence of these requirements is that a double necessity and proportionality assessment should occur prior to the deployment of any real-time RBI system in publicly accessible spaces for law enforcement purposes. First, an assessment should be made by the user when performing a FRIA, as required by Article 5(2) AI Act. Second, in accordance with Article 5(3) AI Act, a judicial or independent administrative authority must also assess the necessity and proportionality of using such a system within the limits of the national law providing the legal basis for any such use, taking the Charter and other Union law into consideration. As a consequence, any such system may only be used 1) after a FRIA and 2) when the competent national authority has authorised such use.

(382) Article 5(3) AI Act must be read and understood in conjunction with Article 5(5) AI Act: for the use of a real-time RBI system to be authorized, a national law adopted in the Member States concerned must exist authorising such use.[217] Certain Member States already have a system of prior authorisation in place for the use of biometric systems under other Union or national law, such as data protection law.

10.2.2. The main principle: Prior authorisation by a judicial authority or an independent administrative authority

(383) The use of real-time RBI systems which pursue one of the objectives listed in Article 5(1)(h)(i) to (iii) AI Act and which has been provided for in the national law of the Member States concerned must be authorized by a judicial authority or an independent administrative authority **prior to its use**. This is the main principle.

(384) However, there is an exception in the case of urgency. This shall be **duly justified[218].** Urgency **is described as** 'situations where the need to use the systems concerned is such as to make it **effectively and objectively impossible to obtain an authorisation before commencing** the use of the AI system'[219] In such case of urgency, 'the use of the AI system should be restricted to the **absolute minimum necessary** and should be

[217] See also Article 5(2) AI Act: '(…) in accordance with the national law authorising the use thereof. (…)'.
[218] This means that 'the law enforcement authority should in such situations request such authorisation while providing the reasons for not having been able to request it earlier, without undue delay and at the latest within 24 hours'. (Recital 35 AI Act).
[219] Recital 35 AI Act.

subject to appropriate safeguards and conditions, as determined in national law and specified in the context of each individual urgent use case by the law enforcement authority itself'.

10.2.2.1. Prior and reasoned request in accordance with national procedural rules

a) Request by whom?

(385) Whilst not specified, it may be assumed that the request will normally be initiated by the deployer, i.e. **by the competent (law enforcement) authority**. According to the definition of law enforcement authority under Article 3(45) b) AI Act, 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' is considered a law enforcement authority and could also be the responsible as the 'competent authority' for submitting the request for prior authorisation.

(386) The use of a real-time RBI system for activities falling outside the scope of the AI Act does not need to be authorised under Article 5(3) AI Act. If subsequently such a system is being used for law enforcement purposes, the use would fall within the scope of the AI Act and an authorisation would be required where the requirements of Article 5(1)(h) AI Act are met.

b) Request for which use?

(387) Prior authorisation is needed for the use of 'real-time' RBI systems in publicly accessible spaces **for law enforcement purposes,** even if the systems are operated by other parties on behalf of law enforcement authorities, for example sport clubs or shopping malls**.**

For example:

- An organisation entrusted with resources for searching for missing children decides to use a real-time RBI system. It has no mandate to exercise public authority and public powers or for preventing criminal offences or tasks for the prevention of threats to public security. Such use does not fall under the prohibition laid down in Article 5(1)(h) AI Act, since it is not for law enforcement purposes. That system will, however, be categorised as 'high-risk' (point 1 (a) of Annex III) and a requirement of prior consultation of the supervisory data protection authority may be necessary pursuant to Article 36 GDPR. Depending on the applicable national law and whether one of the exceptions to Article 9(1) GDPR applies, a prior authorisation may also be required for such processing. By contrast, if the same organisation were requested by law enforcement authorities to act on their behalf for the search of missing children in a law enforcement context and under the supervision and instructions of the competent law enforcement authorities, prior authorisation would be needed pursuant to Article 5(3) AI Act.

- A private organisation entrusted with providing resources to aid persons who risk becoming victims of a natural disaster[220] decides to use a real-time RBI system for that purpose. Such use does not fall under the prohibition laid down in Article 5(1)(h) AI Act, since it is not for law enforcement purposes. That system will, however, be categorised as 'high-risk' (point 1 (a) of Annex III) and a requirement of prior consultation of the supervisory data protection authority may be necessary pursuant to Article 36 GDPR. Depending on the applicable national law and whether one of the exceptions to Article 9(1) GDPR applies, a prior authorisation may also be required for such processing.

c) When? 'Each use'

(388) In accordance with Article 5(3) AI Act, prior authorisation is needed for 'each use'. This implies that the decisive moment for obtaining such authorisation is not the moment prior to installing real-time RBI systems, but each concrete use thereof.

For example:

- the police install biometric ready CCTV cameras at the main train station of a city (no authorisation under the AI Act is needed, but the biometric system must comply with the requirements on high-risk systems, a FRIA must be prepared prior to the first use and an individual authorisation by a judicial or independent administrative authority is needed before each individual use of the system).

The police has concrete indications that a terrorist will arrive by train in the town (prior authorisation is needed for real-time identification).

d) Motivated Request

(389) Article 5(3) AI Act requires each individual request for the use of real-time RBI to be 'reasoned' and hence substantiated and motivated.

(390) Certain Member States allow such requests to be submitted online.[221] In accordance with Article 5(5) AI Act, national legislation should lay down requirements regarding the exact content of the request, while fully taking into account the requirements outlined above, including sufficient evidence to determine the strict necessity and proportionality for the use of real-time RBI and other relevant aspects to reflect the exceptional nature of authorising such use.

10.2.2.2. Authorisation by a judicial authority or an independent administrative authority

(391) The authorisation may only be granted by a judicial or an independent administrative authority whose decision is binding.

[220] Natural disasters include a river flood or fire of nature.
[221] See for example, the requests for authorisation to the French Data Protection Authority, the CNIL.

a) Independent authority

(392) The CJEU has interpreted the concept of 'independence' in different contexts. In *HK v Prokuratuur*, for example, the CJEU explained that independence means that the authority maintains a *'neutral stance'*[222]. The CJEU specified that an authority involved in previous investigations, in that case the public prosecutor, does not have such independence. Similar considerations may apply as regards the independence required by Article 5(3) AI Act, implying that the authorising authority needs to be independent from the authority using the RBI system. This would apply not only for the police, but also cases of investigative judges or prosecutors that supervise the work of police and the use of RBI for which authorisation is sought.

(393) In *Commission v Poland*, a case dealing with the question when a body can be considered to be independent in the context of railway safety, the CJEU found that, 'as regards public bodies, independence usually refers to a status that ensures that the body in question is able to act completely freely in relation to those bodies in respect of which its independence is to be ensured, shielded from any instructions or pressure'.[223] Similar indications may apply in the context of Article 5(3) AI Act.

(394) The judicial authorities in a democratic society are in general also independent authorities. The judiciary plays an important role when it is independent from the executive government(s) and from the legislator, covering and reviewing the application of the legislation and fundamental rights and freedoms in an autonomous and independent way. Judicial independence is one of the crucial facets of the rule of law and is guaranteed by Article 47 (Charter) and Article 6(1) ECHR.[224]

b) Authority of the place where the use will take place

(395) The authorisation must be addressed to the authority that is competent according to national law.[225]

c) Authorisation only if 'necessary and proportionate' to achieving one of the objectives set forth in the exceptions

(396) Any authorisation to use real-time RBI in publicly accessible spaces for law enforcement purposes must assess whether the requirements of Article 5(3) AI Act have been met.

Highly intrusive

(397) In the data protection context, the use of biometric data, in particular facial recognition technology, has been considered by the European Data Protection Board (EDPB) in its

[222] Judgment of the Court of Justice of 2 March 2021, *Prokuratuur*, C-746/18, ECLI:EU:C:2021:152, paragraph 54.
[223] Judgment of the Court of Justice of 13 June 2018, *Commission* v *Poland*, C-530/16, ECLI:EU:C:2018:430, paragraph 67.
[224] See R. Manko, *Judicial independence in the case law of the European Court of Human Rights*, Briefing, European Parliamentary Research Service (EPRS), 2022, 12 p. ; X, *ECJ case law on judicial independence. A Chronological overview*, Briefing, European Parliamentary Research Service (EPRS), 2023, p.12.
[225] See judgment of the Court of Justice of 6 October 2015, *Schrems*, C-362/14, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:62014CJ0362, ECLI:EU:C:2015:650, paragraph 44.

Guidelines 5/2022 and the European Data Protection Supervisor (EDPS) as affecting several fundamental rights and freedoms. That view is shared by The EU Agency for Fundamental Rights and by the Council of Europe.[226] Both the CJEU[227] and the ECtHR[228] have confirmed the sensitive nature of processing biometric data.

(398) Any interference **in fundamental rights and freedoms must always respect the essence of the rights and freedoms**. This follows from Article 52(1) Charter.

(399) The notion of 'essence' of fundamental rights and freedoms has been developed in the CJEU's case-law and is an independent value in the Union's legal order. If the essence of a fundamental right or freedom is not respected, it means that a right or freedom is unduly touched by a measure so that no interference shall be allowed upfront.

Only 'if necessary and proportionate'

(400) Any interference with fundamental rights and freedoms requires 'a law' that should in principle respect the necessity and proportionality pursuant to Article 52 Charter. (See below under Article 5(5) AI Act.) Article 5(3) AI Act requires that the national law permitting the use of real-time RBI in publicly accessible spaces for law enforcement purposes must provide that authorisation for such use is permitted 'only where it [the authority] is satisfied that the use is necessary for, and proportionate to, achieving one of the objectives specified' in Article 5(1)(h) AI Act. National authorities need to verify whether the biometric identification is strictly necessary.[229] This assessment should be based on the FRIA that should have already include an assessment of the necessity and the proportionality in a general manner prior to requesting the authorisation for each use with the specific circumstances.

10.2.2.3. The exception to the requirement of prior authorisation: request within 24 hours and consequences if rejected

(401) In cases of urgency, a user may submit a request for authorisation within 24 hours as from the moment that the real-time RBI system is used. In practice, that will generally be the moment that the biometric ready or capably cameras are 'switched on' and deployed and the first biometric comparison is made with the system. The logging of the processing activities should be made available to the authority to substantiate the timeliness of the request.[230]

(402) In such a case, the request should motivate why no prior request was submitted prior to starting using the system.

---

[226] Consultative Committee of the Convention for the protection of Individuals with regard to Automatic Processing of Personal Data (ETS 108), *Guidelines on Facial Recognition,* 2021.

227 Judgment of the Court of Justice of 26 January 2023, *Ministerstvo na vatreshnite raboti*, C-205/21, ECLI:EU:C:2023:49, paragraphs 60 to76 and 116 to 134.

[228] Judgment of the European Court of Human Rights of 4 July 2023, *Glukhin* v *Russia,* Application no. 11519/20, ECLI:CE:ECHR:2023:0704JUD001151920, paragraphs 88 and 90 (hereinafter referred to as the '*Glukhin* v *Russia* judgment').

[229] See also for data collection: Judgment of the Court of Justice of 28 November 2024, *Ministerstvo na vatreshnite raboti*, C-80/23, ECLI:EU:C:2024:991.

[230] Automatically generated data logs must be kept for at least 6 months for high-risk AI systems and shall include for the high-risk systems mentioned in point 1(a) of Annex III for each use the start and end data and time. See Article 12.3 (a) and Article 19 AI Act.

10.2.2.4. Immediate cessation in case the request for authorisation is rejected and deletion of the data

(403) Article 5(3) AI Act further provides that if an authorisation request in the case of urgency is rejected, the use of the real-time RBI system should be ceased with immediate effect. In such cases, all the data, including the results and outputs of that use, must be immediately discarded and deleted.[231] Article 5(3) AI Act is explicit in this regard, without exception. The deployer will have:

a) a reference database, containing the biometric information (e.g., facial images, voice snippets, ….) and related identifying information, if applicable, against which
b) captured biometric information from individuals present in the publicly accessible space is compared to identify and single out those individuals.
c) This comparison will lead to the comparison result.

(404) The requirement to discard and delete the data collected and processed also means that the reference database(s) used for the unauthorized biometric identification must be removed and deleted if it was built specifically for the contested search. *Only* where the law enforcement authorities had built and intended to maintain the database used for identification in *a lawful manner* for legitimate aims *other* than for the unauthorized use of real-time RBI may the database be maintained.

(405) Besides the deletion of any (unlawful) database with biometric information, all the collected images and other personal data, including the meta data, technical processing data, including the templates and other personal data, and other comparison - and output data obtained during the unlawful use of the real-time RBI system must also be deleted.

(406) Where the law enforcement authority challenges the rejection, the data may be kept by a trustee until a final decision has been taken on the request. During that period, those data should normally not be placed at the disposal of the law enforcement authority [310].

10.2.2.5. No decision-making solely on the output of the real-time RBI system

(407) In accordance with Article 5(3) AI Act, even where the deployer of a real-time RBI system obtains an authorisation, no decision that produces an adverse legal effect on a person may be taken based solely on the output of the 'real-time' RBI system.

> For example:
>
> - A person is arrested and imprisoned for a serious crime solely based on identification by a facial recognition system, without any further checks. This comes on top of the requirement under Article 14 AI Act for human oversight. Checks could relate for example to the question whether a given person has been at a different place or also whether there are other reasons for that the person cannot be the person searched.

Requirements of Article 14 AI Act for Human Oversight

[231] The supervisory authorities should also have the powers to do this post fact check and control. See Article 5(5) AI Act.

(408) The use of real-time RBI that is permitted because it pursues one of the objectives listed in Article 5(1)(h) and complies with Articles 5(2)-(6) AI Act still falls under the rules for high-risk systems. In accordance with Article 14 AI Act, 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 they are in use.' Pursuant to Article 14(5) AI Act, no action or decision may be taken by the deployer on the basis of the identification resulting from the system, 'unless that identification has been separately verified and confirmed by at least two natural persons with the necessary competence, training and authority' or unless 'Union or national law considers the application of this requirement to be disproportionate'. Article 4 AI Act prescribes AI literacy measures for providers and users of AI systems to ensure 'a sufficient level of AI literacy of their staff and other persons dealing with the operation and use of AI systems' and considering the persons on whom the systems are to be used.

(409) As stated by the EDPB in the data protection context, for human oversight to be effective it is crucial 'to enable the person to understand the (in that case facial recognition) system and its limits as well as to interpret its results properly. It is also necessary to establish a workplace and organisation that counteracts the effects of automation bias, and avoids fostering the uncritical acceptance of the results e.g. by time pressure, burdensome procedures, potential detrimental career effects etc.'[232] Similar considerations may apply in the context of the AI Act.

10.3. Notification to the authorities of each use of 'real-time' remote biometric identification systems in publicly accessible spaces for law enforcement

Article 5(4) AI Act provides:

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 5. The notification shall, as a minimum, contain the information specified under paragraph 6 and shall not include sensitive operational data.

(410) Each use of an RBI system pursuing one of the objectives listed in Article 5(1)(h)(i)-(iii) AI Act must be notified to the relevant market surveillance authority and the national data protection authority. Notification must take place after each use in order to be able to report about the number of authorisations and their result. The notification does not need to include sensitive operational data. According to Article 3 (38) AI Act, 'sensitive operational data' means operational data related to law enforcement activities (prevention, detection, investigation or prosecution of criminal offences), the disclosure of which could jeopardise the integrity of criminal proceedings.

[232] EDPB, *Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement* Version 2.0, 26 April 2023, p. 22.

(411) For the details on the reporting requirement, see section 10.6 below.

10.4. Need for national laws within the limits of the AI Act exceptions

10.4.1. Principle: national law required to provide the legal basis for the authorisation for all or some of the exceptions

(412) National laws are required for operationalising the use of 'real-time' RBI systems in publicly accessible spaces for the purposes of law enforcement. At the same time, Article 5(5) AI Act provides that Member States remain free to decide whether to adopt such national laws. If a national law authorising the use of real-time RBI is adopted, the AI Act specifies the substantive elements which the national laws must contain to comply with the requirements laid down in the AI Act.

Article 5(5) AI Act

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 purposes of law enforcement within the limits and under the conditions listed in paragraph 1, first subparagraph, point (h), and paragraphs 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, first subparagraph, point (h), including which of the criminal offences referred to in point (h)(iii) thereof, the competent authorities may be authorised to use those systems for the purposes 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.

10.4.2. National law shall respect the limits and conditions of Article 5(1)(h) AI Act

(413) Since the use of 'real-time' RBI systems in publicly accessible spaces for law enforcement purposes is considered an interference with fundamental rights, Article 5(5) AI Act provides that such use shall be established by national law in the Member States. Those national laws provide the legal basis for the use of such systems.

(414) National laws shall not exceed the limits set by Article 5(1)(h) AI Act and shall respect all further related conditions set forth in the AI Act. That implies that the Member States may not expand the objectives for which real-time RBI may be used in publicly accessible spaces for law enforcement purposes beyond those listed in Article 5(1)(h)(i)-(iii) AI Act[233].

233 See judgment of the Court of Justice of 5 April 2022, *Commissioner of An Garda Síochána*, C-140/20, ECLI:EU:C:2022:258, paragraph 54: 'In order to satisfy the requirement of proportionality, the national legislation must lay down clear and precise rules governing the scope and application of the measure in question and imposing minimum safeguards.'

(415) Member States shall notify their national laws to the Commission at the latest 30 days following the adoption thereof. Such notification does not provide a presumption of conformity of the Member States' law with the AI Act. The AI Office, after receiving the notification, will send a confirmation of receipt. Prior to its adoption, Member States are also encouraged to send a preliminary version of the national (or regional) law proposed to the AI Office. In any case, non-notification to the AI Office within the legal deadline of 30 days following the adoption set out in Article 5(5) may imply that the national law is unenforceable in legal proceedings, as has been held in different contexts.[234] The Commission will publish the Member States' laws on a public website.

(416) Member States may introduce, in accordance with Union law, more restrictive laws, i.e. laws with stricter requirements than those laid down in Article 5(1)(h) and (2) to (7) AI Act.

10.4.3. Detailed national law on the authorisation request, the issuance and the exercise

(417) As regards the detailed rules that apply for the request, the issuance and the exercise of the authorisation, these are to be determined by national law. Each Member State that wishes to allow the use of the systems at issue must specify in its national laws such rules, which are aimed at providing relevant and complete information as to the use of real-time RBI systems to the authorising authority to enable it to decide as to the strict necessity and proportionality of such use.

The national law permitting the use of real-time RBI systems may regulate for example,

- who are the competent authorities subject to Article 5(1)(h) AI Act and the independent authorities in the Member State competent for issuing (or refusing) an authorisation;

- the detailed scope of the objectives for which real-time RBI in publicly accessible spaces may be used for law enforcement purposes (without going beyond the objectives listed in Article 5(1)(h) (i) to (iii), but possibly narrowing them further down;

- providing that requests shall be in writing and requiring a detailed explanation of the specific use and the intended purpose of use for a specific criminal offence/situation which justifies its use;

- the requirement of motivation and the submission of supporting evidence (and need of translation if relevant) for justifying the use of the system pursuing the objectives listed in Article 5(1)(h)(i) to (iii) AI Act, in particular relating to place, period, and personal scope and justifying the strict necessity and proportionality, including the relevance, sufficiency and the efficiency of the use of the system and the absence of less intrusive means;

[234] See, by way of analogy, judgment of the Court of Justice of 19 December 2019, *Airbnb Ireland*, C-390/18,EU:C:2019:1112, paragraphs 96 to 97.

- the description of the technology that will be used and the location points of the data collection;

- the minimum reliability, threshold used, and accuracy rates of the systems used;

- the possibility of auditing the submitted information, including the technical details and accuracy criteria any time *ex ante* and *ex post* by the authorising authority;

- the specification of the reference databases used;

- the retention duration of the data captured and all other related personal data used;

- the security measures, including against unlawful access to the data;

- other safeguards (when relevant);

- the description of any cooperation with private or public authorities, including in other countries, and data transfers and exchanges;

- the traceability of the process;

- the name of the responsible personals of the deployers;

For the issuance *as to other formal elements*

- the possibility of a written procedure complemented with a hearing;

- the grounds of refusal;

- the rights of persons for which a search is conducted, the rights of persons whose data is captured, and the possible rights of third parties[235];

- the delays within which the authorities will have to take their decision;

- any need for formal notifications upon granting/refusal of the authorisation;

- sanctions for not complying with (formal and substantial) requirements;

- the right to appeal an authorisation that has been denied;

For the exercise

- the registration of the use of real-time RBI systems in a central register with a summary of the substantive elements;

- possible further reporting obligations;

- procedure for extending or modifying the authorisation.

10.4.4. Detailed national law on the supervision and the reporting relating to the authorisation

(418) Article 70 AI Act obliges the Member States to 'establish at least one notifying and one market surveillance authority.' Article 74(8) AI Act provides that 'Member States shall

[235] See e.g;, EDPB, *Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement* Version 2.0, 26 April 2023, p. 24 et seq.

designate as market surveillance authorities for the purposes of this Regulation either the competent data protection supervisory authorities under Regulation (EU) 2016/679 or Directive (EU) 2016/680, or any other authority designated pursuant to the same conditions laid down in Articles 41 to 44 of Directive (EU) 2016/680.'

(419) This comes on top of the designation of the authorising authority, which the Member State will have to establish before it can authorise the use of real-time RBI systems for any of the objectives listed in Article 5(1)(h)(i) to (iii) AI Act.

10.5. Annual reports by the national market surveillance authorities and the national data protection authorities of Member States

Article 5(6) AI Act provides

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 4 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.

(420) The national market surveillance authorities and the national data protection authorities of Member States that have been informed by deployers of the use of real-time RBI systems in publicly accessible spaces for law enforcement purposes (see Article 5(4)) must submit annual reports on such use to the Commission. These reports shall be made on the basis of a template provided by the Commission. This template will be established in due time.

(421) Where the deployer is an EU Institution, body, or agency, the EDPS is obliged to inform the Commission accordingly on an annual basis of the real-time RBI systems used in publicly accessible spaces for law enforcement purposes.

(422) Only the report of the national data protection authority will cover the period between 2 February 2025 and 2 August 2025, since the AI Act does not require Member States to appoint a national market surveillance authority before the latter date.

(423) National market surveillance authorities and national data protection authorities are free to decide whether they wish to submit individual reports or a joint report per Member State.

10.6. Annual reports by the Commission

Article 5(7) AI Act provides

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 on the basis of the annual reports referred to in paragraph 6. Those annual reports shall not include sensitive operational data of the related law enforcement activities.

(424) The AI Act requires the Commission to publish annual reports of the uses of real-time RBI systems in publicly accessible spaces for law enforcement purposes in the Member states and by Union institutions, agencies and bodies, based on aggregated data. These reports will be based on the information notified by the national authorities pursuant to Article 5(6) AI Act.

(425) The Commission's annual report shall not contain sensitive operational data. Sensitive operational data means 'operational data related to activities of prevention, detection, investigation or prosecution of criminal offences, the disclosure of which could jeopardise the integrity of criminal proceedings'.[236] This could mean that specific details that reveal ongoing or past investigations, such as e.g., locations, camera's used, shall not be published.

10.7. Out-of-Scope

(426) All other uses of RBI systems that are not covered by the prohibition of Article 5(1)(h) AI Act fall within the category of high-risk AI systems as defined by Article 6 and listed in point 1(a) of Annex III AI Act provided they fall within the scope of the AI Act.

(427) RBI systems that fall outside the scope of the prohibition in Article 5(1)(h) AI Act include biometric verification/authentication systems and the retrospective use of (post-) RBI systems in publicly accessible spaces for law enforcement purposes. For instance, police authorities might be authorised by national law to perform retrospective facial recognition to compare images of criminal suspects with recorded facial images in a criminal database.[237] Another use that falls outside the scope of the prohibition is the use of real-time RBI systems for law enforcement purposes in either a private (such as at somebody's place) or an online space (such as the use of a chat room or online game to identify a suspect of disseminating child sexual abuse material). Finally, the use of RBI systems by private actors, both in real-time and retrospective (such as the use of live facial recognition technology by a supermarket to identify known shoplifters, the use of live facial recognition technology by a sports arena to identify individuals banned from entering the arena, or the use of live facial recognition technology in schools for security purposes and school attendance) fall outside the scope of the prohibition.

(428) In addition to the rules that apply to high-risk AI systems generally, the **retrospective use of RBI systems** for law enforcement purposes is subject to additional conditions

[236] Article 3(38) AI Act.
[237] For instance, the *Traitement des Antécédents Judiciaires* database in France, created by *Décret no. 2012-652 du 4 mai 2012 relatif au Traitement des Antécédents Judiciaires* (Decree 2012-652).

and safeguards in accordance with Article 26(10) AI Act (in application as from 2 August 2026).[238]

(429) Uses for **purposes other than for law enforcement** must in any event comply with **data protection rules**. The cases below illustrate the interpretation of Article 9(2) GDPR in cases of such use and the exceptions to process biometric data.

For example,

- A French administrative Court found that the trial of live facial recognition technology in two public schools for access control and security purposes was neither necessary nor proportionate (under data protection rules). Alternative solutions that were less intrusive for the students were available, e.g. the use of badges. In addition, the conditions for explicit consent were not met. Therefore, consent could not be used as a valid legal basis to trial facial recognition technology in high schools. [239]

- A supermarket was not allowed to use live facial recognition technology to prevent shoplifting in the Netherlands. Without explicit consent from the customer or any legal basis allowing the processing for a substantial public interest (such as security purposes), the supermarket could not process biometric data and thus deploy facial recognition technology.[240]

- The use of live facial recognition technology at the entrance of a football club to identify supporters was prohibited in France [241] and to ensure the safety of spectators was prohibited in Spain. [242]

10.8. Examples of uses

The police instals mobile CCTV cameras equipped with AI-based facial recognition technologies on a police van around the main entrance of a football stadium during a European Championship match to secure the area and identify individuals whose faces are recorded in an ad hoc watchlist database of wanted individuals. This watchlist includes persons suspected of having committed a crime (ranging from serious crimes to frauds and burglaries), persons of possible interests for intelligence purposes, and vulnerable persons with mental issues. The police's use of live facial recognition technologies is not linked to information concerning the presence of a specific person at the event. Although there are likely to be people on the watchlist for the search of whom the use of real-time RBI would be allowed, this list is too

[238] Article 26(10) and Recital 94 AI Act.
[239] TA Marseille (Administrative Court in Marseille) 27 February 2020, no. 1901249.
[240] https://www.autoriteitpersoonsgegevens.nl/en/current/dutch-dpa-issues-formal-warning-to-supermarket-for-use-of-facial-recognition-technology.
[241] https://www.cnil.fr/fr/reconnaissance-faciale-et-interdiction-commerciale-de-stade-la-cnil-adresse-un-avertissement-un-club

[242] https://www.biometricupdate.com/202401/spanish-data-authority-opposes-facial-recognition-for-football-stadium-access

unspecific and is not linked to the event of the football match. Such use would therefore be prohibited.

A biometric identification system (not remote) verifies whether people have access to a nuclear energy plant. When people present themselves in front of the (obvious) camera and access is refused by the system, the system subsequently tries to identify whether the person is on a watchlist of terrorists. The system is not remote. Persons were actively participating in the verification exercise to gain admission to the plant. The use-case does not fall under the prohibition of Article 5 AI Act.

The police authorities of a busy city deploy AI-powered CCTV cameras, which can perform live facial recognition technologies. Possibly, different functionalities are being added, such as object detection and crowd movement, on top of facial recognition.

They place these cameras at multiple locations, including places of worship, a number of places frequented by the LGBT+ community, doctors' offices, pharmacies, and various restaurants and bars.

The installation of biometric-ready cameras as such is not prohibited under the AI Act.

Certain uses however, including the unspecified and indiscriminate identification of natural persons, is prohibited.

Several burglaries occurred in a residential neighbourhood during the summer break. The police obtain a description of the suspect from eyewitnesses, who saw the suspect at different moments in the neighbourhood ahead of the burglaries. To identify and arrest the suspect, the police use live facial recognition technology at different locations in the neighbourhood during a weekend. Based on the indications of eyewitnesses, the police created a facial composite of the suspect and extracted several pictures of individuals resembling the facial composite from a custody database.

Even if the police use live facial recognition technology against a targeted suspect and have defined a perimeter and time of use, the use is not allowed to be deployed in case of an offence, which is not listed in Annex II of the AI Act.

The police screens the emotions of fans in a football stadium with a biometric recognition system. The system spots some potential aggression and immediately deploys in that part of the stadium real-time RBI to identify hooligans that were violent in the past.

The screening of emotions in the stadium is not prohibited under the AI Act (It still falls under the high-risk category of the AI Act). The application of real-time RBI however would be prohibited under the AI Act, in particular where it is the biometric system that decides upon the necessity to identify the persons for the purposes of law enforcement.

The police relies on a CCTV network installed in the city and metro to identify a political protestor that organised a collective protest in the streets. In the Member State concerned, organisers of collective protests held on public roads and public areas, such as streets, must notify the municipal authorities three days in advance of a planned protest to prevent public disorder and violence. The absence of notification is a criminal offence punishable by up to six months imprisonment and a maximum fine of EUR 8 000. To identify the protestor, the police extracts the video feeds from the CCTV cameras installed in the streets and performs retrospective facial recognition by comparing the extracted images with photographs posted on social media.

The **retrospective use of facial recognition technology** is not prohibited by the AI Act. That use is considered high-risk and should comply with the requirements in the AI Act for such systems[243].

Further examples of NOT prohibited practices:

- Hotels using real-time RBI to recognise VIP guests. This is not law enforcement.
- Shopping malls using real-time RBI to find shoplifters. This is not law enforcement.

Prohibited:

Entrusted by the police, a shopping mall is using real-time RBI to find shoplifters. The system is deployed for law enforcement purposes, in a publicly accessible space. The use is prohibited because the search for shoplifters does not fall under any of the exceptions of Article 5(1)(h) AI Act.

11. ENTRY INTO APPLICATION

[243] The processing of biometric data for a law enforcement purpose remains subject to Article 10 of the LED, which needs to be implemented at national level. Their processing to perform the retrospective use of FRT should only be allowed if it is strictly necessary and should be subject to appropriate safeguards. Whether the retrospective use of FRT is strictly necessary to identify the demonstrator is questionable. In the *Glukhin* v *Russia* judgment that serves as a basis for this scenario, the ECtHR ruled that while crime detection can be a legitimate aim, the use of FRT, both retrospective and live, was disproportionate as there were no risks to public order or transport safety. The Court emphasized the 'highly intrusive' nature of FRTs. In that case, the Court concluded that using FRTs did not answer a pressing social need, nor was it necessary in a democratic society.

(430) According to Article 113 AI Act, Article 5 AI Act applies as from 2 February 2025. The prohibitions in that provision will apply in principle to all AI system regardless of whether they were placed on the market or put into service before or after that date[244].

(431) At the same time, the chapters on governance, enforcement and penalties will become applicable on 2 August 2025. Consequently, the provisions on penalties for non-compliance with the prohibitions in Article 5 AI Act will not apply before 2 August 2025. In this interim period, there will also be no market surveillance authorities to monitor whether the prohibitions are being properly complied with.

(432) Nevertheless, even in this interim period, the prohibitions are fully applicable and mandatory for providers and deployers of AI systems. Those operators should therefore take necessary measures to ensure that they do not place on the market, put into service or use AI systems that could constitute prohibited practices under Article 5 AI Act. Even if the provisions on monitoring and fines do not apply until a later date, the prohibitions themselves have direct effect and thus enable affected parties to enforce them in national courts and request interim injunctions against the prohibited practices.

12. REVIEW AND UPDATE OF THE COMMISSION GUIDELINES

(433) These Guidelines constitute a first interpretation with practical examples of the prohibitions in Article 5 AI Act. The Commission will provide additional support to operators and authorities how to understand the prohibitions and collect further practical use cases on an ongoing basis with input from providers and deployers of AI systems, the AI Board and other relevant stakeholders.

(434) The Commission will review these Guidelines as soon as this is necessary in view of practical experience gained in the implementation of the prohibitions and the pace of technological, societal, and regulatory developments in this area. This also includes any relevant experience from market surveillance enforcement actions and interpretations given by the CJEU on the prohibitions and other provisions of the AI Act examined in these Guidelines. During such a review, the Commission may decide to withdraw or amend these Guidelines. The Commission encourages providers and deployers of AI systems, national market surveillance authorities through the AI Board, the AI Advisory forum, the research community, and civil society organisations to contribute to this process by responding to future calls for public consultation.

[244] See Article 111(1) and (2) AI Act which specifies that the grandfathering clause is without prejudice to the application of Article 5 AI Act as referred to in Article 113(3)(a) AI Act.