# CLASSIFICATION OF AI SYSTEMS AS HIGH-RISK (CHAPTER III, SECTION 1) *

Alejandro Huergo Lora
Full Professor of Administrative Law
University of Oviedo

CONTENTS: 1. Classification of AI systems as high-risk. 2. Commentary to Article 6 (Classification rules for high-risk AI systems). 2.1. AI systems classified as high risk due to their relationship with products subject to European legislation listed in Annex I. 2.2. High-risk AI systems in application of Annex III [Article 6(2)]. 2.3. The different cases listed in Annex III. 2.3.1. Biometrics. 2.3.2. Critical infrastructure. 2.3.3. Education and vocational training. 2.3.4. Employment, workers' management and access to self-employment. 2.3.5. Access to essential goods and services. 2.3.5.1. Access to essential public assistance benefits and services. 2.3.5.2. AI systems used for the assessment of creditworthiness. 2.3.5.3. Systems used for risk assessment and pricing in life and health insurance. 2.3.5.4. High risk systems for classification and prioritisation of calls to emergency services. 2.3.6. Law enforcement. 2.3.7. Migration, asylum and border control. 2.3.8. Administration of justice and democratic processes. 2.3.8.1. Administration of justice. 2.3.8.2. Democratic processes. 2.4. The possible exemption of AI systems despite their inclusion in Annex III. 2.4.1. Material criteria determining exclusion. 2.4.2. The specific cases in which it is understood that the system does not influence the content of the decision. 2.4.3. The counter-exception: profiling. 2.5. The application of the exceptions in Article 6(3). 2.6. Approval of guidelines by the Commission. 2.7. The Commission's power to add or delete cases of AI systems exempted from high-risk status (paragraphs 6-8). 3. Commentary to Article 7

* This paper has been written within the framework of the Research Project 'Algorithmic tools for citizens and public administrations', funded by the Spanish Ministry of Science and Innovation (PID2021-126881OB-I00).

(Amendments to Annex III). 3.1. Relationship between Article 7 and Article 6(3). 3.2. Criteria to be taken into account by the Commission for the inclusion in Annex III of AI systems not initially included therein. 3.3. The Commission's power to remove AI systems from Annex III.

## 1. *Classification of AI systems as high-risk*

Articles 6 and 7 establish a complex system for identifying high-risk AI systems. This complexity is due to the need to adapt the classification to the changing reality of AI, where technological changes are constantly occurring.

In addition, the Regulation aims to avoid administrative overload and allows a system that should, in principle, be classified as high risk to be exempted if it is shown to produce limited risks.

The general principle is to limit the number of high-risk AI systems to those that pose a significant risk and to limit the burden arising from that classification: 'AI systems identified as high-risk should be limited to those that have a significant harmful impact on the health, safety and fundamental rights of persons in the Union and such limitation should minimise any potential restriction to international trade' (paragraph 46).

The classification follows the following scheme:

- AI systems that are high risk due to their relationship with Annex I products (products subject to European legislation due to the risk involved).
- Stand-alone AI systems that are high risk because they are listed in Annex III.
- Exceptions to Annex III: An AI system listed in Annex III may not be considered high risk if it is justified that the circumstances of Article 6.3 are met and the procedural requirements of Article 6.4 are fulfilled.

In addition, and in order to facilitate the adaptation of the Regulation to the changing circumstances of AI, it is foreseen:

- Delegated rules adopted by the Commission to amend Article 6(3) in any of the following ways: by increasing the cases in which an Annex III system may be allowed to be considered as not high risk, by modifying them or by deleting any of those cases.
- Delegated rules adopted by the Commission to amend Annex III, also in any way: by adding, modifying or deleting assumptions.

## 2. *Commentary to Article 6 (Classification rules for high-risk AI systems)*

### 2.1. *AI systems classified as high risk due to their relationship with products subject to European legislation listed in Annex I*

Article 6(1) classifies as high risk AI systems which are products falling within the scope of the European legislation listed in Annex I, or which are safety components of one such product (provided, as we shall see, that the particular product is not only subject to one such standard but that a third party conformity assessment is required).

Therefore, the AI system itself may be considered a product covered by these standards, however, AI systems that are intended to be used as a safety component of such a product are also classified as high risk.

This definition seems very strict since there are many ways in which a product can be 'enriched' or completed with an AI system. Apparently, the AI system can improve the performance of the product in many different ways, and not only as a security component. It can improve its efficiency, allow for customisation of results, etc. I believe that, in view of the intention and meaning of the Regulation, a broad interpretation of the concept should be chosen, since if the aim is to address the risks created by AI systems in conjunction with products covered by the rules listed in Annex I, all of them must be addressed, whatever their purpose, whether it is to become safety components or something else.

The Regulation closes the circle in Articles 102-110 by amending some of the rules listed in Annex I to mandate that, when adopting acts implementing those rules, the requirements listed in the Regulation are taken into account for products which are high-risk AI systems. AI systems that are covered by the rules listed in Annex I are high risk. It is assumed that they will be regulated as high-risk AI systems in the Regulation, but it is also required that the acts implementing these Annex I standards take into account the requirements set out in the Regulation. The aim is to avoid a mismatch between the two rules. This applies to AI systems that are themselves products subject to one of the Annex I rules, rather than to systems that are adjective, ancillary or complementary to products covered by one of the Annex I rules.

Annex I lists 20 European harmonisation standards which cover 20 products and establish a common minimum framework to reduce their risks and avoid the proliferation of national standards that would hamper the free movement of such products. These are:

- machinery,

- toys,
- pleasure craft and personal watercraft,
- lifts,
- equipment and protective systems intended for use in potentially explosive atmospheres,
- radio equipment,
- pressure equipment,
- cableway installations,
- personal protective equipment,
- appliances burning gaseous fuels,
- medical devices,
- *in vitro* diagnostic medical devices,
- civil aviation security,
- two- or three-wheeled vehicles and quadricycles,
- agricultural or forestry vehicles,
- marine equipment,
- railway system,
- motor vehicles and their trailers, and systems, components and separate technical units intended for such vehicles,
- manned and unmanned aircraft, their engines, propellers, components and equipment to control them remotely.

The fact that an AI system is itself a product subject to a standard listed in Annex I, or is used as a safety component of such a product is insufficient in itself to classify it as a high-risk system. It is also necessary that the product in question is subject to a conformity assessment procedure involving third party intervention for that Annex I standard.

The rules in Annex I, as well as the AI Regulation itself, set different levels of requirements depending on the level of risk of each product. At the lowest levels, conformity assessment can be done internally without needing third party intervention. Third party intervention is necessary when a higher level of risk is reached.

The different modalities of conformity assessment of products are listed in Decision 768/2008 of the Parliament and of the Council of 9 July 2008 on a common framework for the marketing of products. A third-party conformity assessment is used when a notified body or conformity assessment body is involved.

As Recital 51 recalls, the concept of high risk in the Regulation is not the same as in these standards. Thus, some of them, such as those on medical devices, require third party conformity assessment not only for high risk

devices, but also for medium risk devices. Any AI system that is used in relation to these products, or is itself one of them, will be considered as a high-risk AI system.

Therefore, the 'hinge' linking the AI Regulation to the standards listed in Annex I is the concept of third party conformity assessment. Whatever level of risk is attached to this requirement in the standard in question, it is understood in the AI Regulation as making the system high-risk.

It is important to note that while all products covered by one of the standards listed in Annex I, for which these standards provide for third-party conformity assessment, are considered high-risk AI systems, *not every high-risk AI system requires third-party conformity assessment*. However, the possibility that a high-risk AI system does not require third-party conformity assessment refers to stand-alone systems (i.e. those in Article 6.2 and Annex III) because when an AI system is high-risk under Article 6.1, the conformity assessment shall be that which corresponds to it under the relevant Annex I standard (that which corresponds to that AI system as a product subject to that Annex I standard, or that which corresponds to the product subject to an Annex I standard of which that AI system is a safety component or, more broadly, to which it is related), according to Article 43(3) of the Regulation.

Article 6(1) clarifies that it is irrelevant whether or not the AI system is placed on the market or put into service separately from the product, which is subject to one of the standards listed in Annex I. This is presumably intended to prevent the apparently independent placing on the market of an AI system, which is an accessory to a product subject to one of the standards listed in Annex I, from being used to circumvent the requirements for high-risk AI systems. This does not apply to AI systems, which are themselves products subject to one of the standards listed in Annex I because they cannot be placed independently on the market.

A problem or doubt may arise here with regard to 'dual-use' systems, i.e. systems which are placed on the market independently of a product covered by one of the standards listed in Annex I but which may or may not be used as an accessory to one of those products. Uncertain situations may arise where one of these systems may be classified as high risk when it is not, in fact, high risk because the system may be used with one of these products, but not necessarily so.

### 2.2. *High-risk AI systems in application of Annex III [Article 6(2)]*

AI systems listed in Annex III, referred to as stand-alone AI systems in Recital 52, are also considered high risk.

This recital indicates that the intended purpose of each system (in specific areas), as well as the severity of the possible harm (to the health and safety or the fundamental rights of persons) and the probability of occurrence, have been taken into account in drawing up Annex III.

Clearly, the risks of AI depend on the context in which it is used, what the Regulation identifies as intended use and the area. Fundamentally, what characterises these areas is that the decision taken with the help of AI may affect third parties other than the subject using the AI system and also, especially, that these are third parties placed in a situation of weakness, inferiority or vulnerability (consumers, insured persons, users of public services, students, etc.).

The combination of severity of harm and probability of occurrence is a must in any risk assessment. An assessment based solely on probabilities can lead to erroneous decisions because a given probability may be low in relation to assumable risks, but in the case of systemic or severe damage (so-called fat tail risks), any probability is too high to be assumed.

Regarding the legal assets that may suffer damage due to the use of AI systems and damage that the regulation of high-risk systems seeks to prevent, the Regulation constantly mentions health, safety and fundamental rights, with a brief explanation in recitals 47 and 48.

The case of safety and health is easier to see through examples such as autonomous cars or diagnostics based on AI systems. In these cases, the potential harm arises when the human operator leaves the decision to the AI. The problem is not so much in the existence of the system, which can enrich the human decision and even make it safer by providing contrast or a filter, but the abandonment of the human operator to the judgement of the AI system, whose possible failures become fatal.

The reference to fundamental rights becomes much more complex because fundamental rights are many and, moreover, go in different directions. Recital 48 mentions human dignity, respect for private and family life, protection of personal data, freedom of expression and information, freedom of assembly and association, the right to non-discrimination, the right to education, consumer protection, workers' rights, the rights of persons with disabilities, gender equality, intellectual property rights, the right to an effective remedy and a fair trial, the right of defence and the

presumption of innocence, and the right to good administration, as well as the rights of children and the right to a high level of environmental protection.

There are different ways in which harm can be caused to these rights through the use of AI systems. Sometimes, that harm can come from AI errors. This occurs when individuals are to be treated differently based on their conduct or merit, and when the decision is made on the basis of an AI system, the AI system makes the wrong decision and harms those affected. For example, decisions on access to employment or admission to educational institutions. In other cases, such as for example, when talking about dignity or privacy, the problem is not so much the adoption of 'wrong' decisions as the fact that it is considered necessary, in order to respect fundamental rights, to treat all subjects equally, instead of profiling them in order to give them individualised treatment. For example, treating consumers differently because of the information provided by an AI system may be considered contrary to their rights. The AI system would, in these cases, be the instrument that would enable conduct that is considered to be detrimental to fundamental rights. As in the previous case, the harm would not lie so much in the AI system but in the decision taken by human operators to engage in a certain (discriminatory) conduct with the help of the AI system, which would make possible something that would be very difficult (if not impossible) to do in the absence of such technology.

Therefore, these are two different approaches. In one case, the AI system can be harmful when it works poorly. On the other, the AI system is harmful even if it works well.

However, a theoretical effort is still needed to explain in which cases the use of AI systems can violate fundamental rights. Long lists of rights, such as the one contained in recital 48, are not very indicative because the possible infringements are very varied and different from one another.

### 2.3. *The different cases listed in Annex III*

#### 2.3.1. *Biometrics*

Some 'biometrics' AI systems are considered high-risk, i.e. systems that, based on images of people, can provide information about them, be it their identity (by comparing the image with a database), personal characteristics (gender, age, ethnicity, etc.) or states of mind (emotion recognition). We talk about biometrics because these systems break down images (mainly of faces,

but other body images can also be analysed, or videos of people in movement) into multiple parameters that can be expressed mathematically and be subject to calculation (eye colour, proportion between different dimensions of the face, gait, etc.).

Biometrics is one of the success stories of AI, as can be, for example, translation or natural language processing (dubbing or transcription of audio or video recordings). It is not so 'creative' a task that it involves the as yet non-existent general AI, but neither is it so automatic that it could be done with purely deterministic programming as has existed and been used for decades. Biometrics requires handling extensive data sets to extract correlations that make it possible, from images of people, to produce those results in the form of identification or other types of information that have been mentioned.

I refer in particular to the use of 'real-time' biometric identification systems in publicly accessible spaces for the purposes of law enforcement, biometric categorisation systems aimed at ascertaining the race, political opinions, sexual orientation, religion, beliefs or opinions of natural persons (irrespective of the intended purpose of such information), or systems that seek to infer emotions of a natural person in the areas of workplace and education institutions (with exceptions). These practices are not totally prohibited because states are allowed to authorise 'real-time' biometric identification in publicly accessible spaces for the purposes of law enforcement if some strict requirements (relating to the crime to be prosecuted and judicial or similar authorisation of the use of such a practice) are met.

Should biometric AI systems that are prohibited by Article 5 be considered to be covered by Annex III(1), or does Annex III(1) only cover systems that are not prohibited? In my view, to the extent that the prohibition in Article 5 is not strict, but is in a sense a qualified authorisation, these AI systems should be considered high-risk. It makes no sense that 'real-time' biometric identification systems in publicly accessible spaces for the purposes of law enforcement, which can be used because they meet the requirements of Article 5, should be subject to lesser requirements than other biometric systems that fall within the scope of Annex III(1).

In this respect, it should be recalled that Article 5 (prohibition) is a different form of regulation from that which applies to high-risk AI systems. The prohibition of certain uses of AI (always subject to exceptions) only implies that certain purposes or use cases are ruled out but does not impose any regulation on AI systems that can be used. In other words, this form of AI regulation does not go into the 'guts' or 'kitchens' of AI systems, it only

prohibits AI from doing certain things. In contrast, the regulation of high-risk AI systems is different and, in some ways, more intrusive (or so it has been perceived by many operators) because it imposes performance obligations that affect (and make it more expensive) AI systems.

Precisely, this difference explains why, in my opinion, all biometric AI systems (strictly speaking, biometric verification is another matter) that are not expressly prohibited (i.e. also those that benefit from the exceptions set out in Article 5) should be understood to be included in paragraph 1. Thus, for example, 'real-time' biometric identification systems in publicly accessible spaces that do meet the requirements or emotion detection systems can be used

Recital 54 clarifies why these systems are considered high-risk: the possibility of 'technical inaccuracies', i.e. errors. Such errors may not be random or individualised but may be more general in scope and may have an impact on previously existing biases or discriminatory treatments (to the detriment of minorities, women, etc.), which is what is meant by 'biased results' and 'discriminatory effects'.

Three biometric AI systems are classified as high risk in this section 1 of Annex III.

Firstly, remote biometric identification systems. These are systems aimed at finding out who is the person in an image or searching for a person in a set of images (e.g. from cameras located in multiple public or private spaces). All systems with such a purpose are high-risk, regardless of whether they are intended for use in publicly accessible spaces or elsewhere (in this sense, the concept is broader than the delimitation of the analogous prohibited practice, which only includes systems used in publicly accessible spaces).

On the other hand, it is clarified that the concept of a remote biometric identification system 'shall not include AI systems intended to be used for biometric verification, the sole purpose of which is to confirm that a specific natural person is the person he or she claims to be'.

This is not an exception, but the delimitation of two different concepts. In identification, a system attributes to a person, without his or her request or consent, a certain identity, right or wrong. In verification, on the other hand, the system checks, at a person's request, whether he is who he claims to be, for the purposes of, for example, being able to cross a border post, to enter a sports club, or to be considered as having gone to work. [1] It seems

[1] Article 3 sets out the two concepts. 'Biometric identification' means 'the

clear that this second activity is less invasive, although it is true that the consent given by the person concerned may be purely fictitious if alternative ways of identifying oneself are not established in order to gain access or achieve the intended purpose. If no other valid form of proof of identity is provided, the worker is left with no alternative but to submit to biometric identification or leave the job. In fact, data protection regulators have imposed several sanctions, for example, in Spain, on biometric verification systems in the workplace or also used in hotel establishments to verify the identity of the person ordering a meal, on the grounds that data protection requirements, starting with sufficient consent and the principle of proportionality, were not met. [2]

In both cases, the greatest risk comes from misidentification errors, in the sense that a person is assigned a different identity and prosecuted as an alleged offender, for example. From individual error, one can move on to collective bias, where the system errs more frequently and more harmfully concerning certain groups. In the case of verification systems, in principle, the error does not have the same consequences since it only leads to a person being prevented from exercising their rights (e.g. a member of a club who cannot gain access because the system does not recognise them) or being prevented from proving that they have fulfilled their obligations (time and attendance system that does not recognise a worker). However, an excessive and disproportionate accumulation of data may lead to a violation of the right to privacy insofar as, completely unnecessarily (the same purpose could be achieved by showing a document or signing a document), biometric information is deposited, which can lead, for example, to the detection of a disease or even the prior commission of a crime (think of the

automated recognition of physical, 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', while 'biometric verification' means 'the automated, one-to-one verification, including authentication, of the identity of natural persons by comparing their biometric data to previously provided biometric data'.

[2] On the sanction imposed by the Spanish Data Protection Agency for the use of biometric identity verification systems in hotel establishments (file 78/2021, fine of March 2022), more information can be found here: https://www.aepd.es/prensa-y-comunicacion/notas-de-prensa/comunicado-registro-de-datos-de-ciudadanos-por-alojamientos-turisticos. Also, the decision of November 2024 in case 202304834, for fingerprint identification of employees of a notary's office.

cases of detection of the perpetrator of a past crime, thanks to the information deposited in genetic banks, supposedly to find out where a person's ancestors come from).

In addition to the risk of error, there is also the risk of loss of privacy, similar to that produced by any form of video surveillance, only in this case, its identification potential is increased because there is no need for a laborious and imperfect manual checking of images, but the screening is automatic and more effortless.

'AI systems intended to be used for biometric categorisation, according to sensitive or protected attributes or characteristics based on the inference of those attributes or characteristics' are also considered high risk. The concept of biometric categorisation is defined in Article 3 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'. These are, therefore, systems that are not aimed at establishing who is in the image is a specific person (i.e. to prove that this person is the one in the image and, therefore, was in that place at that time) but are aimed at determining that the person in the image (whether identified or not) has certain inherent characteristics. What are these characteristics? It seems necessary, although there is no clear thread in this direction in the regulation, to link this concept of 'sensitive or protected attributes or characteristics' with the 'special categories of personal data' referred to in Article 9.1 of the GDPR: 'racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, health or a natural person's sex life or sexual orientation'.

One of the utilities of AI can be precisely to infer these attributes (or many other not so protected attributes, such as preferences or predispositions for consumption) without the need for a statement from the subject, operating from the known data of that person (in this case, biometric data revealed in an image) and on the basis of correlations that allow profiles to be established. Therefore, AI makes it possible to carry out this categorisation from images (with the corresponding margin of error, of course).

The risk here derives from the possibility of errors. There is a prior, fundamental question, and that is whether it is legitimate to try to classify people based on these personal circumstances, because this can only serve to treat people differently (something that may be incompatible with the constitutional principle of equality) and because constitutional texts and declarations of rights often prohibit citizens from having to indicate their

choice (for example, Article 16.2 of the Spanish Constitution states that '[n]o-one may be obliged to declare their ideology, religion or beliefs'). However, this issue is not decided by the Regulation since, as we will see in a moment, the fact that these biometric systems are high-risk does not mean that their use is allowed (as long as they meet the requirements of high-risk systems), but that this decision is left to other European or national rules.

Finally, 'AI systems intended to be used for emotion recognition' are also high risk. Emotion recognition is defined as 'an AI system for the purpose of identifying or inferring emotions or intentions of natural persons on the basis of their biometric data' (Article 3(39)). Certain emotion recognition systems are prohibited by Article 5. Any emotion recognition system is high-risk and, even if permitted, must comply with the requirements applicable to such systems.

Any emotion recognition system (AI is a tool for this, there could be others based not on it but on other technologies, although it is this one that has allowed progress to be made in achieving this objective) involves an intrusion into people's privacy, by allowing third parties to know something that in principle is internal and reserved and that the person decides whether to tell, how to tell and to what extent. In some cases, emotion recognition may border on violating the right not to testify against oneself. Rating the system as high risk does not eliminate this intrusive effect, it only tries to prevent the system from making mistakes.

One of the keys to this paragraph 1 is the formula (which does not appear in all the paragraphs of Annex III) 'in so far as their use is permitted under relevant Union or national law'. In other words, the Regulation imposes obligations on the use of these AI systems but does not permit their use, which depends on the provisions of other rules, European or national. It could be the case (in other words) that the use of such biometric systems would be prohibited, in which case they could not be used even if they met the requirements for high-risk AI systems. This decision of the Regulation makes sense because even if the risk of errors is addressed (which is the main purpose of the regulation of high-risk systems), the use of these systems raises privacy issues (to which I have referred) that the Regulation does not address.

As to what such European or national regulation authorising the use should look like, strictly speaking, a rule expressly authorising such use is not required. The formula used, 'permitted under relevant Union or national law', is different from that used, for example, in Article 5.5 for the possible authorisation of the use of 'real-time' remote biometric identification

systems in publicly accessible spaces for the purposes of law enforcement, since in this case 'in their national law the necessary detailed rules for the request, issuance and exercise of, as well as supervision and reporting' are required.

However, the reference to 'relevant Union or national law' must, in my opinion, be interpreted as more than a generic 'provided that it is not prohibited'. There will need to be some kind of regulation that, with a broader or stricter content, admits the use of biometrics for this purpose. In the case of biometrics for verification purposes, protection legislation does not prohibit it, and no express provision or permission is deemed necessary. It is simply subject to the same filter as the processing of particularly sensitive data. On the other hand, I believe that biometric categorisation and emotion recognition are particularly sensitive and require explicit regulatory authorisation and acknowledgement.

As for the choice between whether such a permit is to be derived from European or national legislation, the principle of primacy must be understood to apply. A European prohibition would preclude possible national authorisation. Moreover, a European permit could be binding on national law or be subject to the reservation that there is no national prohibition (which is precisely the model followed by Annex III).

### 2.3.2. *Critical infrastructure*

AI systems intended to be used as safety components in managing and operating critical digital infrastructure, road traffic, or in the supply of water, gas, heating or electricity' are high risk.

As provided for in Article 3(62), 'critical infrastructure' is as defined in Article 2(4) of Directive 2022/2557.

A safety component is defined as 'a component of a product or of an AI system which fulfils a safety function for that product or AI system, or the failure or malfunctioning of which endangers the health and safety of persons or property'. As we have already seen, this definition is ambiguous, misleading or does not fully coincide with the plain meaning of the words since it includes both components that have the function of providing safety to the critical infrastructure and those that, whatever their purpose, may cause insecurity in case of malfunctioning.

Recital 55 makes the same misunderstanding by not considering as safety components those that serve to make the critical infrastructure work (or

may affect its operation in case of failure or malfunctioning): 'As regards the management and operation of critical infrastructure, it is appropriate to classify as high-risk the AI systems intended to be used as safety components in the management and operation of critical digital infrastructure as listed in point (8) of the Annex to Directive (EU) 2022/2557, road traffic and the supply of water, gas, heating and electricity, since their failure or malfunctioning may put at risk the life and health of persons at large scale and lead to appreciable disruptions in the ordinary conduct of social and economic activities. Safety components of critical infrastructure, including critical digital infrastructure, are systems used to directly protect the physical integrity of critical infrastructure or the health and safety of persons and property, but which are not necessary in order for the system to function. The failure or malfunctioning of such components might directly lead to risks to the physical integrity of critical infrastructure and thus to risks to the health and safety of persons and property. Components intended to be used solely for cybersecurity purposes should not qualify as safety components. Examples of safety components of such critical infrastructure may include systems for monitoring water pressure or fire alarm controlling systems in cloud computing centres'.

One of the cases where it is clearest that an AI system is high risk is when it is used to operate or manage a critical infrastructure. The danger here is, above all, the possibility of it falling prey to some kind of cyber-attack that could allow sabotage with potentially severe results. The implementation, no longer of AI systems, but of IT systems that allow these critical infrastructures to be remotely operated and their functioning controlled while facilitating their management and making it possible to obtain instrumental data, makes them more susceptible to attack or sabotage increases the risk (the malfunctioning of these systems affects or may affect the functioning of the critical infrastructure) and threatens to turn them into weapons in any kind of confrontation (hybrid warfare). In contrast to other high-risk systems, where the greatest danger is that individuals are harmed when the AI system issues wrong predictions or information about them that leads to unfounded treatment (danger of discrimination), here there is a clear risk to public safety.

Recital 55 again falls back on the ambiguity in using the term safety component because it again implies that only those components whose function is to ensure the safety of the critical infrastructure are important. However, many times, the AI system operating in the infrastructure is not aimed precisely at improving safety (which is not the primary function of AI

systems), but at improving its efficiency in another way, and the safety problem comes from the fact that a malfunctioning AI system does put the safety of the infrastructure at risk. This is adequately covered by Article 3, paragraph 14, but not by recital 55.

This provision of the Regulation acts in coordination with Directive 2022/2557, which deals with the security of critical infrastructure and provides an additional layer of security, in this case, focusing on the AI systems that can be added to it.

### 2.3.3. *Education and vocational training*

There are several scenarios in which an AI system used in the educational environment is considered high-risk and subject to consequent requirements.

Firstly, systems that can be used to allow admission to educational institutions: 'AI systems intended to be used to determine access or admission or assign natural persons to educational and vocational training institutions at all levels'.

AI systems that can be used to evaluate learning outcomes, whatever the consequence of that evaluation, are also considered high risk, also when it contributes to defining the treatment of those concerned in the education system: 'AI systems intended to be used to evaluate learning outcomes, including when those outcomes are used to steer the learning process of natural persons in educational and vocational training institutions at all levels'.

Also closely related to admission, systems indicating the appropriate level of education for a subject are considered high risk: 'AI systems intended to be used for the purpose of assessing the appropriate level of education that an individual will receive or will be able to access, in the context of or within educational and vocational training institutions at all levels'.

Finally, and related, in this case, to outcome assessment, systems used to detect fraudulent behaviour are also considered high risk: 'AI systems intended to be used for monitoring and detecting prohibited behaviour of students during tests in the context of or within educational and vocational training institutions at all levels'. These systems are particularly, but not only, used in e-learning. In any case, whatever the educational modality in which they are used, they would be high risk systems.

The common element in all these areas, which probably justifies the

classification of these systems as high risk, is the relationship with the principle of equality and equal opportunities. Insofar as access to certain educational levels (or qualifications, or centres) can have a powerful influence, not only on a person's intellectual development but also on access to better paid jobs (and this is probably the dominant perspective at the moment), we want this access not to be conditioned or predetermined by discriminations embedded in AI systems.

This is confirmed by Recital 56 of the Regulation. On the one hand, it begins (as usual) with a hymn to the advantages that AI can offer in this sector: 'The deployment of AI systems in education is important to promote high-quality digital education and training and to allow all learners and teachers to acquire and share the necessary digital skills and competences, including media literacy, and critical thinking, to take an active part in the economy, society, and in democratic processes'. I understand that AI offers the same possibilities and advantages in the educational field as in others: efficiency (predicting which courses a person might be interested in, for example, or predicting what will be more difficult for them to insist on) and savings in simple tasks (generative AI that can help prepare teaching materials, for example).

However, then it is recalled that such systems can determine the future of individuals (the opportunities they receive) and are therefore classified as high risk: 'However, AI systems used in education or vocational training, in particular for determining access or admission, for assigning persons to educational and vocational training institutions or programmes at all levels, for evaluating learning outcomes of persons, for assessing the appropriate level of education for an individual and materially influencing the level of education and training that individuals will receive or will be able to access or for monitoring and detecting prohibited behaviour of students during tests should be classified as high-risk AI systems, *since they may determine the educational and professional course of a person's life and therefore may affect that person's ability to secure a livelihood*'.

Moreover, it makes explicit how people can be harmed, identifying this risk, fundamentally, with discrimination, both at the individual level (individual errors that affect specific people) and at the collective level (bias against groups, perpetuation of traditional patterns, discrimination): 'When improperly designed and used, such systems may be particularly intrusive and may violate the right to education and training as well as the right not to be discriminated against and perpetuate historical patterns of discrimination, for example against women, certain age groups, persons with

disabilities, or persons of certain racial or ethnic origins or sexual orientation'.

It seems necessary to specify and explain what is implicit in these provisions and to help interpret their scope of application.

Access (or non-access) to education (or specific segments of it) has always had an impact on people's future, on their economic possibilities and personal development. Although this influence is never total (there are always people who are very successful with very little education, and the income range that corresponds to a certain level of education is, in reality, extensive), the fact is that leaving school very early or, on the other hand, obtaining certain qualifications can have a significant influence, for better or worse, on the income that a person obtains throughout his or her life. The concept of the 'social lift' is a further illustration of this idea, showing that the income obtained by a person (taken as the factor that most influences his or her quality of life) will depend more on the education received (which depends on his or her abilities) than on the income of the household into which he or she was born. In reality, the fact that education determines people's income does not in itself indicate that such a 'social lift' exists since access to educational levels can be (and often is) determined by economic reasons or reasons linked to a person's social origin. 'Social lift' requires equality of access.

The social state tends to guarantee not only equality of opportunity in access, but also the provision of a wide educational network at different levels of education, distributed territorially, so that access to specific centres is not so important. Possibly, in recent years, we have seen a regression. It is assumed that the network of public schools (university or non-university) does not guarantee this adequate and uniform level and that quality depends on each school. With the lure of competition between institutions, the idea that only some institutions are good is gaining ground, making admission processes more important. The attention paid to these processes, for example, by classifying the AI systems that may be used in them as high-risk, is a sad consolation in the face of this reality.

This paragraph of Annex III, unlike others (such as paragraph 1 on biometrics, paragraph 6 on law enforcement, paragraph 7 on migration and borders), does not include the caveat 'to the extent that the use of such systems is permitted by European or national law'. The introduction of such a caveat in paragraphs 1, 6 and 7 is possibly for different reasons. In the case of biometrics, because the regulation considers it to be particularly dangerous and, moreover, it is, in any case, regulated by data protection

law. In the case of law enforcement and border control, because they are obvious manifestations of public power that must be subject to the law, and it is essential that the instruments used in their service (rule of law) are regulated by law.

However, introducing such a precaution in one or other paragraphs of Annex III is not perfectly justified but rather raises many doubts. For example, it would also be necessary to introduce it in paragraph 8 concerning the administration of justice and electoral processes, also related to public processes (judicial and electoral, respectively), which are subject to exhaustive regulation as a guarantee of their purity.

Therefore, I understand that the fact that some sections of Annex III (such as, for example, section 3, which is now being discussed) do not expressly state that these high-risk systems may only be used if permitted by European or national regulations is of no significant significance. In other words, high-risk AI systems must meet the requirements of the Regulation, but that does not mean that they can be used for the purpose for which they are classified as high-risk systems (in this case, the admission of students). This does not depend on the AI regulations but on the regulations governing admission, assessment, etc. An explicit enabling regulation is probably unnecessary, but the Regulation alone does not settle the debate on the admissibility of such systems, which will depend on the relevant legal framework.

What role can AI play in student admissions? We are not talking about systems that simply apply a scale previously established by humans and entered into a computer programme. That would not have the 'autonomy' necessary to be an AI system. One could hypothesise the use of AI systems that 'predict' which candidates are most suitable based on criteria previously supplied by human operators and which cannot be clearly evidenced by the candidate's CV. If the aim is to identify, for example, more 'creative' candidates, an AI system could, by studying data from many professionals, help correlate creativity with other characteristics, creating a 'creative candidate' profile that could be applied to applicants. One can also imagine an AI system helping to filter CVs in the event of a massive number of applications. The system could be used to read the documentation and filter it according to regulated criteria, detecting who meets the requirements in terms of age, qualifications, etc. It seems to me that, in the current state of the art, the use of AI systems in admission procedures should be merely approximate, i.e. it would only serve to screen but not to make a final decision. Moreover, it would be essential that such 'predictions' be subject to

a human review procedure. They can serve, at most, as proposals because it can be assumed that, in many cases, they will be correct in the sense that they will only leave out applications that do not meet the requirements or that are blatantly unfounded, but cannot be definitive decisions in any case, neither to grant admission solely on the basis of the application of the system, nor to leave anyone out definitively for the same reason.

Something similar can be said of the systems mentioned in *c)*, i.e. those that determine the level of education that a subject may receive or have access to. At the outset, this seems to be a largely academic or overlapping assumption because the determination of the level of education is usually made on the occasion of admission to a particular educational establishment or grade. It is not very common for people to be tested for 'potential aptitude'. It could refer to 'level' tests such as those given in some schools (language schools, for example, or sports courses) where there are several levels and pupils are placed in one or another level according to their prior knowledge. But this does not seem to be the point of the paragraph. In any case, the use of AI seems ancillary here too because people should be judged or evaluated based on a direct assessment of their abilities or skills, not on the basis of a prediction based on the examination of third parties, i.e. those whose data have been used to build the model or profile.

The other two sections refer to the final stage of the educational process, the assessment. In fact, there is a clear relationship between all the sections because admission also involves an assessment of applicants, only in this case, the outcome of the assessment is to obtain (or not) a place, whereas in this second case, the assessment serves to obtain a mark. They also differ in that admission is always competitive (the number of places is limited; if there were enough places, the admission process would be a simple eligibility check), whereas in the case of assessment, the outcome is to obtain a mark.

Using an AI system to assess students and, simultaneously, determine admission to educational institutions was tried in the UK in 2020 when the UK did away with face-to-face exams in order to avoid overcrowding. An AI system was developed that 'predicted' the mark the student would have achieved in the exam that was not to be held, replacing that information (the mark, which did not exist because the exam was not held) with data that is known (such as the student's previous marks, their position in relation to their classmates, or the relationship, in previous years, between the marks their school gave and the marks their students achieved in external exams). In the end, there was criticism that students were to be judged partly on the basis of facts for which they were not responsible (e.g. the good or bad

performance of previous students at the same school), which was interpreted as also producing biases against more vulnerable groups, so that in the end the mark predicted by the AI system was used, not as the sole criterion for determining admission, but as an additional criterion, which could only be used when it favoured the student.

Barring anomalous circumstances such as the one indicated, it does not seem easy to justify assessing someone on the basis of a prediction rather than on the basis of their actual performance. It may be more efficient or cheaper to dispense with tests and rely on indications, but it is not acceptable, at least as a general rule, for the same reason that one cannot judge someone by a prediction regardless of the actual facts that have occurred. The application of AI systems may have a place as support and always in an approximate manner (to filter out extreme cases, which can be understood to be discarded due to their similarity to other cases that in previous years obtained a negative evaluation), and always as a proposal that may be subject to review at the initiative of the interested party.

Finally, systems for detecting prohibited conduct in the conduct of examinations are also classified as high-risk systems. Particularly during the COVID pandemic, when a large number of tests had to be carried out telematically on the spur of the moment, concerns were raised about possible fraud and detection procedures were also proposed, which in many cases included the obligation to undergo remote surveillance by using cameras. There were also rulings that rejected the application of these methods on the grounds that they were compulsory (no examination was allowed without undergoing them) or disproportionate (the introduction of cameras into the family home had to be admitted) or that formal requirements concerning the request and obtaining of consent were not met. AI systems may be an advantageous alternative, insofar as, by examining previous cases of fraud and taking into account that, in telematics examinations, other less invasive personal data than the image itself is available (e.g. the way a person types or the speed at which they type), they can detect patterns that lead to the presumption that the person taking the exam is not the student, or that they are receiving external assistance.

This possible use of AI systems is related to the detection of infringements in general, which is one of the most common areas of use. There is a difference, and that is that in predictive policing, the AI system identifies suspicious cases in order for the administrative inspection to target them preferentially, in an attempt to make the inspection more efficient. However, possible infringements will be established by direct verification of the facts

(usually at the inspection) and not by what the AI system says. In contrast, in the case of AI systems used to detect prohibited conduct, it is normal for the pupil's irregularity to be established by the evidence produced or evidenced by the system itself. This raises problems such as the application of the requirements of prima facie evidence (which must be fulfilled for the presumption of innocence to be destroyed and a sanction to be imposed) and the functioning of the reversal of the burden of proof: the AI system provides evidence that the person concerned can refute. Is this possibility of refutation sufficient for the imposition of the sanction to be in accordance with the principle of the presumption of innocence if the alleged offender does not provide evidence to the contrary?

The question arises about the type of risks that the classification of these systems as high risk is intended to avoid. The first is the risk of errors, both individual and collective. But there are other risks that go beyond this. For example, the use of AI systems indicates the use of certain admission and evaluation criteria (the criteria that the AI system is intended to predict or score), out of all the possible ones. Sometimes, the criteria are used as an alternative to the direct assessment of candidates. The classification of these systems as high-risk attempts to avoid errors (individual, or collective, or systematic) and undesirable outcomes but does not address the question of which criteria should be considered in the assessment or admission. Instead, this depends on the assessment or admission of the governing rules , which will also determine whether these systems can be used.

This section of Annex III refers to all levels of education and also expressly covers vocational training. It is doubtful whether it also includes non-formal education (language classes, training courses for employment, etc.). I think it is arguable that its scope does not go that far, especially as it is not normally the basic training network that will determine people's future.

### 2.3.4. *Employment, workers' management and access to self-employment*

Labour law relations are characterised by the weakness of one of the parties, which has led to the adoption of mandatory and protective regulations. The use of AI in this context is, therefore, particularly sensitive, which, in the context of the Regulation, results in the classification of the relevant systems as high risk.

All systems that can affect the working life of workers are qualified as high risk, starting with those that are prior to the birth of the relationship,

i.e. those that can be used in selection processes: 'AI systems intended to be used for the recruitment or selection of natural persons, in particular, to place targeted job advertisements, to analyse and filter job applications, and to evaluate candidates'.

Nevertheless, they are also high-risk systems that can be used by the employer to make decisions affecting the relationship, whether in the evaluation of performance, the allocation of tasks, the determination of other elements (e.g. remuneration) or dismissal: 'AI systems intended to be used to make decisions affecting terms of work-related relationships, the promotion or termination of work-related contractual relationships, to allocate tasks based on individual behaviour or personal traits or characteristics or to monitor and evaluate the performance and behaviour of persons in such relationships'.

The use of AI systems in recruitment is well known. One of the reasons for this is the massive participation in selection processes, which leads companies to use AI systems for filtering. On the one hand, general AI models allow applications to be 'read' and filtered (e.g. by checking compliance with the requirements of the job or by applying other filters). It is also known that recruiters use systems that use job seekers' data other than that contained in their applications (data that they themselves have provided through their digital activity and that the recruitment agency obtains from the market), and that lead to 'profiling' them and placing them better or worse in the selection process.

A broad concept of 'work-related relationships' is used to not be limited to relationships regulated by Labour Law. Explicitly, recital 57 states that 'Relevant work-related contractual relationships should, in a meaningful manner, involve employees and persons providing services through platforms as referred to in the Commission Work Programme 2021'. In many countries, the position of workers on platforms such as Glovo and similar platforms has given rise to many doubts and conflicts, oscillating between their qualifications as work-related and other contractual relationships. The Regulation aims to prevent AI systems used by platforms from ceasing to qualify as high-risk just because the relationship with riders is qualified as contractual in a Member State, hence the use of this term.

Recital 57 explains why these AI systems are classified as high risk, citing three different reasons. On the one hand, the relevance of these systems in the sense of how they affect workers' lives: 'those systems may have an appreciable impact on future career prospects, livelihoods of those persons

and workers' rights'. On the other hand, the danger of bias against historically disadvantaged groups: 'Throughout the recruitment process and in the evaluation, promotion, or retention of persons in work-related contractual relationships, such systems may perpetuate historical patterns of discrimination, for example against women, certain age groups, persons with disabilities, or persons of certain racial or ethnic origins or sexual orientation'. Finally, the danger of affecting their privacy and their right to data protection: 'AI systems used to monitor the performance and behaviour of such persons may also undermine their fundamental rights to data protection and privacy'.

We see three different reasons for this. Firstly, there is the possibility of individual mistakes, resulting, for example, in the exclusion of a job applicant, being left out of a promotion process or being dismissed. Secondly, collective mistakes, i.e. when such mistakes are systematically biased in a certain direction, to the disadvantage of those who are already at a disadvantage. Finally, the risk is that irrespective of whether the results of such systems are right or wrong, their use invades or affects the right to privacy or data protection.

The qualification of the AI system as high-risk will try to prevent individual and systematic errors (biases) from occurring in its application. However, in order to avoid the risks mentioned in recital 57, other keys must be used, namely those of labour law. It is primarily labour law (together with data protection) that regulates the use of employee data, e.g. cameras at workplaces to monitor performance or attendance, access to company e-mails to monitor their use, or the introduction of attendance checks based on personal data (fingerprints, biometric verification). On the other hand, labour law must also intervene to establish how the prohibition of discrimination in access to employment and in employment applies and what degree of motivation is required for business decisions in these areas because this is the key factor in deciding whether they can rely on AI systems or what kind of access to such systems needs to be given to workers affected by those decisions, as an explanation of those decisions.

That is why I consider that, although paragraph 4 does not incorporate, like some others in Annex III, the precaution that such systems must be authorised by European or national legislation, it is essential to reach this conclusion. Labour law is not national legislation that the harmonising nature of the Regulation should displace, and this is not only because of the reference to it in Article 2.11, but because they are regulations with a different purpose. The Regulation regulates AI systems, whereas labour law

regulates decisions affecting workers, regardless of whether or not AI systems are used as an instrument to make those decisions.

### 2.3.5. *Access to essential goods and services*

Paragraph 5 includes several scenarios that have in common that they are AI systems that are used to determine who has access to and who does not, or under what conditions, to benefits or services that are essential to life today ('Access to and enjoyment of essential private services and essential public services and benefits'). It is the fact that such services are of great importance to people's lives (as opposed to others, which would be less relevant) that leads the Regulation to qualify such AI systems as high risk: 'Another area in which the use of AI systems deserves special consideration is the access to and enjoyment of certain essential private and public services and benefits necessary for people to fully participate in society or to improve one's standard of living' (recital 58).

There are major differences between the four headings. In some, we are talking about 'benefits' within a service to which the citizen is entitled in the framework of a social security or other system (medical care, social benefits), while in others, we are talking about contracts between private parties (credit, insurance). In some cases, there is a situation of competition for access to scarce goods (prioritisation of some emergency calls over others), while in others this is not the case. And in some cases, public entities intervene, but not in others. Nor does it include the precaution that these services can only be used if the relevant European or national regulations so state, but, as I have already indicated, I understand that this is implicit and that the regulations governing these services or benefits, by stating how and with what guarantees and explanations decisions on the matter should be taken, indirectly affect this issue.

#### 2.3.5.1. *Access to essential public assistance benefits and services*

'AI systems intended to be used by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for essential public assistance benefits and services, including healthcare services, as well as to grant, reduce, revoke, or reclaim such benefits and services' are qualified as high risk.

Recital 58 further specifies the examples: 'In particular, natural persons

applying for or receiving essential public assistance benefits and services from public authorities, namely healthcare services, social security benefits, social services providing protection in cases such as maternity, illness, industrial accidents, dependency or old age and loss of employment and social and housing assistance, are typically dependent on those benefits and services and in a vulnerable position in relation to the responsible authorities'.

In this case, the delimitation of the factual situation does include the specification that the user of the system is a public authority or another subject acting on behalf of public authorities (such as a concessionaire, contractor, consultant in one way or another, etc.). This indicates, for example, that medical and health care in private facilities would be covered if they are facilities acting within national health systems (such as agents providing the health service on behalf of the public social security authority), but not when they act in a private health service (for patients who have private insurance or who are paying customers). It could also be said that AI systems that are used in the framework of private benefit provision (insurance contracts that include health care benefits to be authorised or recognised by the insurer) are not high risk (at least not under this section).

The Regulation operates outside the categories of public and private law. The borderline between high-risk and non-risk AI systems runs through the legal and economic world without at all conforming to the boundary between public law and private law, but rather crossing it on several occasions. This is even though these concepts (public and private law) are at least partly related to the same reasons. In other words, some entities that are subject to public law (which affects their entire legal regime and that of their staff, with enormous consequences for their operation) precisely because of the sensitive nature of the services they provide (social security, public health, public educational establishments). Here, the Regulation tells us that some of these services can be managed indifferently by public and private entities and that in order to guarantee or safeguard citizens' rights, it is sufficient to comply with certain specific obligations without affecting their entire legal regime.

In any case, precisely in this paragraph a), Annex III does distinguish between public and private operators, since for the AI system to be classified as high risk, it is necessary, in addition to the benefits having certain characteristics, for them to be managed by public bodies or private bodies but on behalf of public authorities.

Recital 58 specifies the reasons for this high-risk classification in the risk of

errors, which would be particularly serious in these cases, in view of the importance of these benefits and services for people's lives: 'If AI systems are used for determining whether such benefits and services should be granted, denied, reduced, revoked or reclaimed by authorities, including whether beneficiaries are legitimately entitled to such benefits or services, those systems may have a significant impact on persons' livelihood and may infringe their fundamental rights, such as the right to social protection, non-discrimination, human dignity or an effective remedy and should therefore be classified as high-risk'.

In those cases where access to these services has been constitutionalised (without prejudice to their indispensable regulation or development in ordinary legislation), erroneous decisions coming from AI systems may affect these rights, as Recital 58 recalls: 'may infringe their fundamental rights, such as the right to social protection, non-discrimination, human dignity or an effective remedy and should therefore be classified as high-risk'.

Some of these AI systems may be 'dual-use' in the sense that they can be used either to assist in public benefit or private benefit decision-making, the former being classified as high risk, but not the latter. This implies that the system provider will have to meet the requirements of high-risk AI systems if the system is to be suitable for such uses, even if, in some cases, it is to be used for non-high risk uses.

It is important to note that systems that serve not only to grant or deny but also to revoke or claim benefits are classified as high risk. These are systems that can be used to detect cases of fraud and are related to other sections of Annex III. Some of them have had a significant prominence as pioneering use cases that have led to some well-known rulings, such as the Syri case in the Netherlands.

In an attempt to reconcile two apparently contradictory objectives, Recital 58 states that this classification as high risk system should not prevent or hinder innovation in the administrative field: 'this Regulation should not hamper the development and use of innovative approaches in the public administration, which would stand to benefit from a wider use of compliant and safe AI systems, provided that those systems do not entail a high risk to legal and natural persons'. Indeed, the classification of a certain type of system as high risk implies costs that will make their use more difficult. However, the Regulation facilitates innovation by stipulating that its requirements will only apply from the moment the system is put into operation, not while it is still in the experimental phase. In addition, the

facilitating clauses in Articles 6 and 7, which relax the requirements for systems listed in Annex III but with low actual risk, should be able to help.

2.3.5.2. *AI systems used for the assessment of creditworthiness*

This point of Annex III(5) qualifies as high risk 'AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, except for AI systems used for the purpose of detecting financial fraud'.

One of the best-known use cases of AI in social environments is that of creditworthiness assessment through predictions. The existence of large amounts of data on past credit transactions and also the possibility to request data from credit applicants facilitates the development of creditworthiness profiles from this data and the ranking of credit applicants, allowing for highly nuanced and seemingly reliable creditworthiness reports. Such creditworthiness reports replace or accompany more traditional means of assessing creditworthiness, such as the use of well-known economic data or of subjective criteria (in the traditional way of working in the sector, these two methods are combined). The regulation in some states of people's credit history or of registers of 'defaulters' (to determine precisely what data can be taken into account in obtaining credit) can be circumvented or counterbalanced by the use of AI systems which, based on data apparently unrelated to creditworthiness (level of education, form of relationship with the bank, type of work), can be correlated with it (or at least that is what the past data indicate).

Recital 58 justifies the classification of these systems as high-risk because of the importance of the legal goods at stake, i.e. because of the relevance of access to credit for people's lives: 'AI systems used to evaluate the credit score or creditworthiness of natural persons should be classified as high-risk AI systems since they determine those persons' access to financial resources or essential services such as housing, electricity, and telecommunication services'. It is assumed that, precisely because of this, possible errors in these systems are highly relevant and should be avoided thanks to the requirements imposed on high-risk systems.

On the other hand, the possibility of systematic errors, i.e. biases, to the detriment of certain groups is also mentioned to justify the classification: 'AI systems used for those purposes may lead to discrimination between persons or groups and may perpetuate historical patterns of discrimination, such as that based on racial or ethnic origins, gender, disabilities, age or

sexual orientation, or may create new forms of discriminatory impacts'. Such biases can have different origins, including the reproduction of historical patterns reflected in the data that were specific to times when these groups were formally discriminated against or did not have accurate and effective access to credit.

Systems used to assess the creditworthiness of natural persons only are classified as high risk. This is in line with the general thrust of the Regulation and also of data protection law (which only protects natural persons), although damage and bias may also occur in the assessment of the creditworthiness of legal persons.

There is an exception since systems used 'for the purpose of detecting financial fraud' are not classified as high-risk, which Recital 58 specifies in these terms: 'AI systems provided for by Union law for the purpose of detecting fraud in the offering of financial services and for prudential purposes to calculate credit institutions' and insurance undertakings' capital requirements should not be considered to be high-risk under this Regulation'. There is no perfect correspondence between the preamble and the text found in Annex III.

Annex III speaks of 'detecting financial fraud' in general, which would also include possible fraud committed by consumers, in line with the possible use of AI systems to detect suspicious behaviour. However, Recital 58 limits the scope of this exception to fraud in the offering of financial services (i.e. corporate fraud) and also extends it to AI systems that are used for prudential purposes to calculate credit institutions' and insurance undertakings' capital requirements. The latter are not included in Annex III, so as they are not high-risk, they are unaffected by the exemption either. As far as fraud is concerned, given the prudential purpose of the Regulation, AI systems for the detection of the risk of consumer fraud, to the extent that they are part of AI systems for the assessment of consumer creditworthiness, should be considered as high risk.

#### 2.3.5.3. *Systems used for risk assessment and pricing in life and health insurance*

This point of paragraph 5 classifies as high risk 'AI systems intended to be used for risk assessment and pricing in relation to natural persons in the case of life and health insurance'.

The reference to life and health insurance was introduced by the European Parliament and has been incorporated, with nuances, in the final version.

Recital 58 justifies the inclusion by referring to the importance that life and health insurance can have for the economy and the life of individuals and the consequent risk in case of failure: 'Moreover, AI systems intended to be used for risk assessment and pricing in relation to natural persons for health and life insurance can also have a significant impact on persons' livelihood and if not properly designed, developed and used, can infringe their fundamental rights and can lead to serious consequences for people's life and health, including financial exclusion and discrimination'.

Insurance has been one of the most prominent use cases for AI systems. If insurance is trying to cope with risk, algorithmic predictions allow it to be measured more accurately, hence their use. By handling a multitude of previously unavailable data to insurers, it is possible to predict more accurately (at least in theory) the risk associated with each individual and make decisions about whether to offer insurance and at what price. This opens the way to greater individualisation of contract terms, with winners and losers. For example, in car insurance, there are people who will pay less (because the system predicts that their risk of accidents is minimal) and others who will be charged a higher premium (for the opposite reason). Moreover, it is often the case that the latter are the least well off (who are the young, or those who are forced, because of poor working conditions, to drive at night, or on more dangerous roads, etc.). When we talk about the 'risks' of such AI systems, we talk about several things. Of course, the risk that these predictions are wrong and make a person pay too high a premium for the risk they actually produce, or even exclude them from insurance. However, also the very fact that, because of this individualisation, people will pay for the risk they actually produce, and some will necessarily be harmed.

The inclusion of insurance is quite limited because it only includes life and health insurance, which are undoubtedly important, but leaves out some critical insurances, such as car insurance (essential, in turn, for many people to have an economic social life, as they need their own vehicle to work or move around) or civil liability insurance, also necessary in many activities, or even home insurance, which is necessary to take out a mortgage.

#### 2.3.5.4. *High risk systems for classification and prioritisation of calls to emergency services*

This point of Annex III(5) classifies as high risk 'AI systems intended to evaluate and classify emergency calls by natural persons or to be used to

dispatch, or to establish priority in the dispatching of, emergency first response services, including by police, firefighters and medical aid, as well as of emergency healthcare patient triage systems'.

The explanation given in Recital 58 is obvious: 'Finally, AI systems used to evaluate and classify emergency calls by natural persons or to dispatch or establish priority in the dispatching of emergency first response services, including by police, firefighters and medical aid, as well as of emergency healthcare patient triage systems, should also be classified as high-risk since they make decisions in very critical situations for the life and health of persons and their property'. In short, they are systems that are used in environments where decisions are made that may influence the life or death of persons or their ability to recover, which, in certain situations such as strokes, heart attacks, etc., depends on their receiving rapid attention. Naturally, such risks also occur when prioritisation decisions are made without using an AI system.

The scope of application of this paragraph does not create too many problems, as the concept of emergency call handling is clear.

### 2.3.6. *Law enforcement*

Law enforcement activities are among the most frequent use cases for AI systems, as they can make predictions that help public authorities use their resources more efficiently.

As Recital 59 indicates, these AI systems are classified as high risk because public authorities may, in order to prevent crimes, take measures that have a substantial impact on citizens, such as 'surveillance, arrest or deprivation of a natural person's liberty as well as other adverse impacts on fundamental rights guaranteed in the Charter'. Possible errors can cause serious harm to those affected.

On the other hand, reference is made to the fact that the difficulties in reviewing or challenging the use of these systems, due to their lack of transparency, may affect 'the exercise of important procedural fundamental rights, such as the right to an effective remedy and a fair trial as well as the right of defence and the presumption of innocence'. In this sense, it is recalled that 'The impact of the use of AI tools on the defence rights of suspects should not be ignored, in particular the difficulty in obtaining meaningful information on the functioning of those systems and the resulting difficulty in challenging their results in court, in particular by

natural persons under investigation'. In the same vein, it points to the risk that errors are not individual, but systemic and have the effect of exacerbating previously recorded patterns of discrimination: 'The use of AI tools by law enforcement and other relevant authorities should not become a factor of inequality, or exclusion'.

Precisely because of the sensitivity of these cases of use of AI systems, the caution is introduced that these systems are high-risk 'in so far as their use is permitted under relevant Union or national law', thus preventing anyone from understanding (in my opinion, mistakenly) that classification as high-risk systems automatically determines the admission of their use. Such admission will depend on whether European or national law so provides.

Several types of AI systems are classified as high-risk for law enforcement use. It must be said that there is a double filter or delimitation of the scope of application, subjective and objective. The objective is clear, and we will now turn to it. The subjective filter refers to the use of these systems by 'law enforcement authorities' (or on their behalf or in support of them), which is, logically, a concept that will have to be defined internally but which basically refers to police forces and similar bodies. The use in the judicial field, which is a borderline field, is dealt with in section 8.

If a private entity uses AI systems for these purposes (calculating the likelihood of a person committing a crime or being a victim of a crime), for example, for the purposes of such as pricing an alarm service or burglary insurance, it seems that Annex III accepts that such AI systems are not classified as high risk, probably because such entities cannot use the same powers as a public authority, which is not always justified.

The concept of law enforcement seems clearly linked to the prosecution of criminal offences, not administrative offences or other types of conduct in breach of the law. In virtually every letter in this section of Annex III, reference is expressly made to criminal offences. Furthermore, Recital 59 expressly states that systems used in tax or money laundering investigations should not be considered high-risk: 'AI systems specifically intended to be used for administrative proceedings by tax and customs authorities as well as by financial intelligence units carrying out administrative tasks analysing information pursuant to Union anti-money laundering law should not be classified as high-risk AI systems used by law enforcement authorities for the purpose of prevention, detection, investigation and prosecution of criminal offences'. This does not seem to be a specific exclusion for this particular type of system (as opposed to others such as, for example,

trafficking systems), but rather an exclusion of 'administrative' rather than criminal systems. Logically, this raises the logical and usual doubts in the case of offences punishable by sanctions, which, because of their seriousness, are comparable to penalties. This has also been the case in other areas, such as the right to a second hearing (Saquetti judgment of 30 June 2020 of the European Court of Human Rights, application 50514/13).

It is important to relate this paragraph to Article 5(1)(d), which prohibits the use of AI systems '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'.

AI systems that predict the risk of offences based on profiling alone are prohibited. Those not based solely on profiling are considered high risk (paragraph d). What kind of systems are these that are not based solely on profiling? In addition to referring to the commentary on Article 5, it should be understood that these are systems which also take into account specific data on the act committed and the involvement of the subject.

Interestingly, profiling systems are also classified as high risk. This does not seem to fit entirely with Article 5, where such systems are prohibited. The clash probably occurred because this prohibited conduct was introduced during the drafting of the Regulation and was not in the original text.

When we talk about the risk of a person committing or re-offending, we are not so much talking about the investigation of offences but about the adoption of measures concerning this person (e.g. granting or not granting prison leave), depending on this risk profile.

In addition to the risk of committing a crime, systems that assess the risk of a person being a victim of crime are also classified as high risk. Initially, these systems were regulated in the same paragraph that also refers to the assessment of the risk of committing a crime, but this has since been split into several paragraphs to take into account the distinction between profiling and the use of other mechanisms. The reason for this division is that when a person is predicted to be at risk of committing an offence, the consequence is the adoption of measures to protect that person, which, even if errors occur, does not seem serious for that person, whereas the prediction that a person may commit offences may lead to the application of

negative or burdensome measures for that person, which is something that requires greater precautions.

Although the use of AI systems to assess the risk of a person being a victim of a crime is less well known, there are very typical cases, such as systems to protect women from the risk of suffering crimes of gender-based violence. In cases such as this one, it can be seen that the difference with systems for assessing the risk of committing crimes is not so great because the deprivation of protection measures can also be severe and harmful to the affected person and also because, sometimes, these protection measures affect third parties because they are protection measures against specific persons.

In addition to AI systems aimed at assessing the risk of a person committing a crime or being a victim of a crime, this section of the Annex classifies as high-risk AI systems used for two other purposes: polygraphs and those used to evaluate the reliability of evidence during the investigation or prosecution of criminal offences. In fact, the two types of systems are very similar because polygraphs also serve to lend credibility to evidence, in this case, to a statement (usually of the investigated person). It seems clear that this section refers to the use of these systems in the investigation and pre-trial phase of criminal proceedings because the trial phase is discussed in section 8.

The fact that these AI systems are classified as high risk does not mean that they can be used (we have already seen that the Regulation does not decide this question), nor does it mean that the result produced by the system must be accepted as valid. On the contrary, it will be subject to criticism or review at the request of whoever is harmed, but precisely the fact that the system must comply with the requirements of high-risk systems is a guarantee of transparency and that progress is being made as far as possible in opening up possibilities for review and control.

### 2.3.7. *Migration, asylum and border control*

Annex III also states that, to the extent that AI systems are to be used for border control and in migration or asylum matters (which will have to be authorised by European or national law, without Annex III itself providing a sufficient legal basis for this), the relevant systems will be considered high-risk and will have to meet the requirements set out for them.

Recital 60 clarifies something that is almost self-evident, namely that the

reason for this qualification is the vulnerable position of the persons to whom these systems apply (i.e. the high stakes) and the risk of errors. 'AI systems used in migration, asylum and border control management affect persons who are often in particularly vulnerable position and who are dependent on the outcome of the actions of the competent public authorities. The accuracy, non-discriminatory nature and transparency of the AI systems used in those contexts are therefore particularly important to guarantee respect for the fundamental rights of the affected persons, in particular their rights to free movement, non-discrimination, protection of private life and personal data, international protection and good administration'. Therefore, and unlike in other cases, the risk is not in the use of AI itself or in the individualisation of decisions that should be taken with equal criteria but in the system making mistakes in assessing the circumstances it seeks to evaluate.

In delimiting the four cases of systems classified as high risk, there is a parallelism with the previous section (law enforcement), with some of the systems coinciding.

The first case (not in the order in which they are listed) is that of systems 'intended to be used (...) for the examination of applications for asylum, visa or residence permits and associated complaints with regard to the eligibility of the natural persons applying for a status'. We are, therefore, talking about persons who, from outside the territory or already in the territory, apply for a certain status or recognition, and AI systems are used for the examination and resolution of this application. Once again, very different realities are included, ranging from systems for reading and processing mass applications to others aimed at intervening in the substance of the matter and assessing compliance with some of the requirements for obtaining that status, which may be indeterminate or, in any case, attainable through predictive tools.

This assumption is related to the employment assumption in the sense that in both cases, systems are used for the evaluation of applications, which are often massive.

Closely related to the previous one is the case of systems used 'to assess a risk, including a security risk, a risk of irregular migration, or a health risk, posed by a natural person who intends to enter or who has entered into the territory of a Member State'. The scenario is similar to the previous one because normally, entry into the territory is linked to obtaining a legitimising title, although this also includes systems that are used to apply the grounds for expulsion legally foreseen for persons who are in the territory but who

may be expelled in the event that their stay causes a risk to basic interests (security, health, etc.). We are talking about algorithmic predictions, profiling and finding correlations. In any case, if a person cannot be treated as a suspected criminal just because it is found that, statistically, he or she is more likely to end up committing a crime than other people, neither can a person be expelled from the territory, or be considered a risk, just because he or she fits a specific profile. In combination with Article 5, it seems clear that indications of risk that an AI system may point to are to be taken as indications, not evidence, and that they are insufficient on their own to expel a person or refuse entry.

Another type of systems included in this section and classified as high risk are those used 'for the purpose of detecting, recognising or identifying natural persons, with the exception of the verification of travel documents'. This brings us back to the concept of biometric identification (as distinct from biometric verification, not considered as high risk). Bearing in mind that biometric identification systems, when not prohibited, are normally qualified as high risk, it seems logical that biometric identification systems are also qualified as high risk for such sensitive uses as border control, on which citizens' ability to enter a given territory depends, which has a decisive influence on their lives and the realisation of their projects.

Finally, and in parallel with the section on law enforcement, other instrumental systems, such as polygraphs or those used to assess the reliability of evidence, are included as high-risk systems. The precise relationship with analogous law enforcement scenarios makes further considerations unnecessary.

### 2.3.8. *Administration of justice and democratic processes*

Annex III progresses with a list of cases of the use of AI systems that are gradually penetrating into the heart of the State and, therefore, raise greater doubts and the need for precaution. The maximum level is that of this section 8, which includes two clearly separate areas that have their maximum relevance in common.

Curiously, this section does not include the formula, which, in other cases, warns that these systems will be classified as high risk if they can be used following the relevant European or national regulations. This does not mean, of course, that the inclusion of these systems in Annex III and their subjection to the requirements of high-risk systems should be understood as

recognition of these systems. Therefore, the fact that in some cases, this proviso is made and in others it is not is of very little relevance.

#### 2.3.8.1. *Administration of justice*

The first paragraph, concerning the Administration of Justice refers to 'AI systems intended to be used by a judicial authority or on their behalf to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts, or to be used similarly in alternative dispute resolution'.

The Regulation could say that AI systems used in the Administration of Justice are high risk in any case but has preferred to use this formulation, which lists a number of different uses, leaving out (possibly) other different uses, which will not, therefore, be prohibited, but, on the contrary, will not even be high risk.

This description basically includes two issues: interpreting and investigating facts and the law (on the one hand) and applying the law to a concrete set of facts. The former is something that is already done routinely and, in many cases, without being recorded. AI as a form of access to factual or legal information. In fact, legal information companies (which is what traditional legal publishers have become) already incorporate AI utilities that enrich the results or performance of their databases. It seems clear that these AI systems will have to be classified as high-risk and subject to the requirements established for them if they are to be used in the judicial field. Many of these systems are intended for use by lawyers, judges or other legal professionals. In theory, it would not always be necessary to qualify as a high-risk system, but the system should certainly be qualified from the outset so that it can be used in all cases.

As in other cases, a subjective delimitation has been incorporated in the drafting of the Regulation, in the sense that systems are high risk when used for that purpose *and by certain authorities*. It is questionable whether their use by lawyers should be considered to fall within this definition (and be considered high risk). On the one hand, lawyers are not judicial authorities, but their role is to assist judicial authorities as collaborators of justice, so it could well be argued that they should be considered high-risk AI systems. In practice, this seems to me to be more of a theoretical problem because lawyers and judges will generally use the same legal AI systems, which will have to be adapted to the most demanding standard (that of judges) and meet the requirements of high-risk AI systems.

It is difficult to determine what is meant by law enforcement assistance systems. The term is broad and covers many different uses, some of which are probably difficult to imagine at this stage. The key here is twofold: any system so used must meet the requirements of high-risk systems, and its admissibility must be decided in the light of the applicable rules, without compliance with those requirements being sufficient.

In any case, recital 61 reminds us of something important: 'The use of AI tools can support the decision-making power of judges or judicial independence, but should not replace it: the final decision-making must remain a human-driven activity'.

It is also important to note that AI systems that are used for similar purposes in alternative dispute resolution are classified as high risk. Arbitration is clearly included, so the fact that its bodies are not state but private is irrelevant in this respect. Other means of dispute resolution, such as mediation, where the law is not strictly applied, may not be as strongly affected by this rule.

Although it is the task of the following paragraphs of Article 6 to exclude from the high-risk classification those systems which, although included in Annex III, do not have relevant effects in terms of risk production, here it is Annex III itself which leaves out certain systems: 'The classification of AI systems as high-risk should not, however, extend to AI systems intended for purely ancillary administrative activities that do not affect the actual administration of justice in individual cases, such as anonymisation or pseudonymisation of judicial decisions, documents or data, communication between personnel, administrative tasks'. These are all very low-profile tasks in terms of information handling. Not even research applications or access to information are included.

#### 2.3.8.2. *Democratic processes*

The possibility of electoral processes being subject to external influences is a major concern that has been accentuated recently, given the evidence that external powers (Russia, Iran, and individuals such as Elon Musk) may try to favour certain options in an attempt to weaken member states or the EU itself. In this sense, this attempt at manipulation by third parties is more worrying than the actual use of malfeasance by electoral contenders, which is an ever-present danger but seems to be relatively under control in European democracies. If this can be said, the manipulation of electoral processes by third parties means that democracy becomes a weakness, a

vulnerability of a given political system. Democracies, led by short-lived leaders and subject to harsh electoral processes, would be at a disadvantage compared to regimes led by long-lived leaders and totally or partially exempt from such processes.

In any case, this paragraph of Annex III does not only refer to such attempts to influence by third parties but to the use of AI systems in any way and any form, provided that it is aimed at influencing the outcome of an electoral or referendum process. In contrast to the other paragraphs of Annex III, no distinction is made here according to who uses the AI system. In fact, political parties, consultants, advertising agencies, etc. are probably being targeted.

AI systems intended to be used to influence the outcome of an election or referendum or the voting behaviour of natural persons in the exercise of their vote in elections or referenda' are considered high risk. Although the idea of 'influencing' the outcome of an electoral process seems to refer to attempts at manipulation by third parties, it is normal for political parties and, in general, participants in elections or those who favour one of the positions in a referendum to try to influence the outcome, and this is the meaning of advertising activity, which is perfectly legitimate.

AI can bring, for example, greater efficiency to this advertising activity and election campaigns, by profiling voters and target audiences and knowing which message to address individually to each of them (based on their inclusion in one of the previously identified groups whose preferences are known and which can be better influenced through specific messages than through common messages for the whole population). Such profiling and targeted advertising have already been the subject of Regulation 2024/900 on the transparency and targeting of political advertising, which imposes transparency and data protection obligations.

AI systems with a purely instrumental role and are concerned with rationalising or improving the management of electoral campaigns from an organisational point of view, but which do not affect the message to citizens, are excluded from the high-risk classification. As this paragraph of Annex III states, 'AI systems to the output of which natural persons are not directly exposed, such as tools used to organise, optimise or structure political campaigns from an administrative or logistical point of view' are excluded. This is also stated in Recital 62.

### 2.4. *The possible exemption of AI systems despite their inclusion in Annex III*

The Regulation contains several elements of flexibility in an attempt to avoid undesired effects. Among these undesired effects is the qualification as high-risk systems of some systems that do not actually produce relevant risks, resulting in excessive rigidity and increased costs.

Annex III is a tentative standard, entering almost uncharted territory and attempting to prevent harm. As a result, reality may show that it does not adequately 'capture' the risk scenarios it intends to include and includes systems that are in fact innocuous or of little relevance. Moreover, the systems included in Annex III are identified by the area in which they operate, not by what they do, the latter being the most important determinant of whether they do (or do not) produce risks.

Therefore, Article 6, in paragraphs 3-8, establishes a mechanism that allows systems included in any of the sections of Annex III to be considered as not being high risk and, therefore, not needing to comply with the requirements established for this type of system.

This provision did not exist in the Commission's original proposal. Both the Parliament and the Council incorporated flexibility mechanisms in their proposals, and, in the end, the solution that now appears in the final text was adopted, which, as is logical with such an origin, has some adjustment difficulties.

Unlike Article 7, which operates in both directions and allows new cases to be included in Annex III or deleted, paragraphs 3-8 of Article 6 only operate in the direction of excluding high-risk status specific systems included in Annex III.

The possible exclusion can only apply to AI systems that are high risk by virtue of their inclusion in Annex III, not to those that are classified as high risk under their relationship with products subject to one of the risk prevention rules listed in Annex II. The latter cannot be exempted from this classification. This is probably because Annex III is approximate in nature (it refers to an as yet untested reality), whereas the standards in Annex II are already well established and refer to products (vehicles, lifts, halls, etc.) which are undeniably dangerous. There is no question that an AI system 'attached' to one of these products could have such an accessory use or function that its classification as a high-risk system is unjustified.

### 2.4.1. *Material criteria determining exclusion*

Paragraph 3 has a complex structure. It begins by stating that 'an AI system referred to in Annex III shall not be considered high-risk where it does not pose a significant risk of harm to the health, safety or fundamental rights of natural persons'. This is the principle that determines that an AI system is not considered to be high-risk despite being included in Annex III.

How is it determined whether an AI system listed in Annex III poses a significant risk or not? The paragraph goes on to say 'including by not materially influencing the outcome of decision making'. Systems that do not materially influence the outcome of decision making are not considered to produce a significant risk, but that is only one of the possible cases where an AI system does not create a significant risk.

Recital 53 is consistent with this interpretation: a system included in Annex III may not produce a high risk, either because it does not materially influence the content of the decision or because it does not harm those protected interests: 'there may be specific cases in which AI systems referred to in pre-defined areas specified in this Regulation do not lead to a significant risk of harm to the legal interests protected under those areas because they do not materially influence the decision-making or do not harm those interests substantially'.

However, specific assumptions are then established in which the AI system does not materially affect the content of the decision (we will see below what these are), and a procedure is also established so that new assumptions can be included or some can be removed, so that, in the end, it seems that in the final wording, the way for an Annex III system to be exempt from complying with the requirements of the high-risk AI systems is by verifying that it fits into one of the assumptions that article 6.3 itself establishes of systems that do not materially affect the content of the decision.

The criterion that the AI system does not materially affect the content of the decision is a very good one. It relates to (but does not correspond to) the concept of 'AI system' as defined in Article 3 (in conjunction with Article 2) and the requirement that the system has 'autonomy'. AI systems can condition the content of decisions affecting the rights of individuals. This is the common element of Annex III systems. For example, they can influence whether a person is admitted to an educational establishment or whether their emergency call is dealt with more quickly than others. Nevertheless, this is relevant to the extent that the system alone establishes

criteria for that decision from the mathematical analysis of the data. If the system merely applies the criteria that the programmer has previously established, or if it merely corrects errors, then it does no influence on the content of the decision.

When the AI system influences the content of the decision, this means that, in order to explain why the decision has been taken, reference must be made to the system because the decision is derived from it. Hence, the need for the system to be 'explainable', otherwise, it will be impossible to give reasons for the decision. On the other hand, when the system does not influence the content of the decision, the system will have been used to elaborate the decision, but it will not be necessary to mention it in order to justify it, and its explicability will be irrelevant (just as it is not necessary to know how a computer or a word processor works in order to control or give reasons for a decision that has been drafted with their help).

### 2.4.2. *The specific cases in which it is understood that the system does not influence the content of the decision*

Paragraph 3 sets out 4 cases in which the AI system is understood not to influence the material content of the decision (and would, therefore, be exempt from the requirements of high-risk systems). In addition, a counter-exception is included, i.e. a case in which, although we are in one of these four cases, the system is understood to be high risk.

The fact that paragraphs 6-8 set out a formalised procedure for increasing or decreasing the list may indicate that the list is very substantial, i.e. not exemplary but exhaustive. In other words, an AI system will only be exempted when one of these four cases is fulfilled. However, recital 53 seems to go in another direction: 'An AI system that does not materially influence the outcome of decision-making *could include* situations in which one or more of the following conditions are fulfilled'. This might indicate that it leaves open the possibility that there are other cases in which it can be argued that an AI system does not affect the content of the decision.

In any case, the first of these assumptions is that 'the AI system is intended to perform a narrow procedural task'. Recital 53 expands on this in these terms: 'such as an AI system that transforms unstructured data into structured data, an AI system that classifies incoming documents into categories or an AI system that is used to detect duplicates among a large number of applications. Those tasks are of such narrow and limited nature

that they pose only limited risks, which are not increased through the use of an AI system in a context that is listed as a high-risk use in an annex to this Regulation'.

The second assumption is that 'the AI system is intended to improve the result of a previously completed human activity', something that seems clearly not to have a relevant influence on the outcome of the decision and which the Recital amplifies in these terms: 'Considering those characteristics, the AI system provides only an additional layer to human activity with consequently lowered risk. That condition would, for example, apply to AI systems that are intended to improve the language used in previously drafted documents, for example, in relation to professional tone, academic style of language or by aligning text to a certain brand messaging'. Again, this is a purely instrumental application. In fact, this kind of support can also be achieved even with generative AI models.

The third of the assumptions listed in this sub-paragraph is that 'the AI system is intended to detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment without proper human review'. The Recital explains it this way: 'The risk would be lowered because the use of the AI system follows a previously completed human assessment which it is not meant to replace or influence, without proper human review. Such AI systems include for instance those that, given a certain grading pattern of a teacher, can be used to check ex post whether the teacher may have deviated from the grading pattern so as to flag potential inconsistencies or anomalies'. Therefore, we are dealing with systems that do not seek to make any decisions, but rather to analyse human decisions in order to detect biases or inconsistencies. As opposed to the clichéd image of the biased AI system, here, the system allows us to detect biases in human performance. This makes much sense as long as we are talking about sustained human performance over time, not specific decisions. A decision may be 'legal' in the sense that it conforms to the regulatory framework, i.e. it is one of the decisions allowed within that framework and meets the requirements of procedure and motivation, but a systematic analysis of all decisions of the same type taken by that human operator reveals that he or she behaves inconsistently and does not act in the same way in similar scenarios. The AI system, as this sub-paragraph says, does not make decisions on its own but simply triggers an alarm when it detects such biases or discontinuities so that the human operator (the same one who has made the decisions, or

often a different one) checks whether the difference is justified or needs to be corrected.

The fourth condition is that 'the AI system is intended to perform a preparatory task to an assessment relevant for the purposes of the use cases listed in Annex III'. Recital 53 explains it as follows: 'The fourth condition should be that the AI system is intended to perform a task that is only preparatory to an assessment relevant for the purposes of the IA systems listed in an Annex to this Regulation, thus making the possible impact of the output of the system very low in terms of representing a risk for the assessment to follow. That condition covers, inter alia, smart solutions for file handling, which include various functions from indexing, searching, text and speech processing or linking data to other data sources, or AI systems used for translation of initial documents'.

The fact that the task entrusted to the AI system is 'preparatory' explains, by itself, that the risk involved is very low and that, consequently, the system is exempted from meeting the requirements of AI systems. In fact, this assumption (as well as, to some extent, the first one) is likely to be very broad, as it is not linked, like most of the previous ones, to a specific content. Therefore, the discussion as to whether this list of cases is exhaustive or exemplary is, to a certain extent, unnecessary. We concluded that discussion with a certain dissatisfaction because initially, it seems that it must be an exemplary enumeration, while later, the structure of the paragraph seems to indicate that it is exhaustive (and, therefore, a formal procedure is established for adding or removing paragraphs). However, this is of little consequence because this fourth assumption (and also, to some extent, the first) can be interpreted broadly.

### 2.4.3. *The counter-exception: profiling*

Paragraph 3 closes with a clause stating that AI systems performing profiling are always high-risk, which is to be understood as meaning that such systems are high-risk even if they do not materially influence the content of the decision because they fall under one of the four exceptions (or those to be included in the future): 'Notwithstanding the first subparagraph, an AI system referred to in Annex III shall always be considered to be high-risk where the AI system performs profiling of natural persons'.

As Recital 53 explains, the definition of profiling is that of Article 4 (4) of

the GDPR, Article 3 (4) of Directive 2016/680 or Article 3 (5) of Regulation 2018/1725 (the three fundamental European rules on data protection): they are overlapping definitions: "'profiling' means 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, interests, reliability, behaviour, location or movements'. The AI Regulation does not contain a definition of profiling.

Profiling is one of the typical AI techniques. The analysis of data on a given subject (consumption of a product, voting for a party, compliance or non-compliance with contractual obligations) allows us to establish correlations and establish groups (for example, people who have certain economic or social characteristics have a higher degree of compliance), which, In turn, when a new subject appears for whom we know certain data, we can include him in one of these groups and predict his behaviour in the specific aspect of interest (for example, predict whether or not he will repay the loan, which is impossible to know, based on his economic or social characteristics, which we do know). Profiling involves risks because it leads to different treatment of individuals, which is one of the common features of most Annex III systems. In principle, systems that fit any of the assumptions of sub-paragraph 2 of paragraph 3 should not include profiling, but it is understood that the paragraph concludes that if there is profiling, there is always high risk.

### 2.5. *The application of the exceptions in Article 6(3)*

Paragraph 4 states: 'A provider who considers that an AI system referred to in Annex III is not high-risk shall document its assessment before that system is placed on the market or put into service. Such provider shall be subject to the registration obligation set out in Article 49(2). Upon request of national competent authorities, the provider shall provide the documentation of the assessment'.

It is, therefore, up to the provider to assess, on his own responsibility, whether any of the exceptions apply. The system must nevertheless be registered, which gives the authorities information about its existence and may enable them to request information from the provider. The provider must give reasons why any of the exceptions apply. This is a so-called compliance system identical to the one used to prove compliance with the

requirements of high-risk systems. The provider must reason why it considers that they do not apply (in the cases of article 6.3) or why it considers that the solutions it has adopted comply with them.

Logically, although it is the provider who must assess whether an exception exists (other parties cannot do so if the provider considers the system to be high risk), this judgement affects or binds other parties, who will also not have to comply with the obligations imposed by the Regulation on high-risk AI systems.

The provider's judgement is not infallible. Moreover, it may be biased in the sense that it is in the provider's interest to claim that the system is not high risk, in order to facilitate its commercialisation. It may be the case that the system is indeed high risk, and this is revealed by the competent authority, e.g. following a complaint. It will have to be seen to what extent the provider's justification, even if incorrect, is sufficiently motivated to determine whether or not a sanction can be imposed. In any case, the system should be subject to the requirements of high-risk systems.

It may happen that one of the parties to whom the Regulation is addressed (e.g. the deployer) has doubts as to whether an AI system is actually covered by an exception, even though the provider believes that it is covered. In that case, although it seems reasonable to assume that the deployer cannot be sanctioned because it can rely on the provider's judgement, it is prudent for the deployer to make an enquiry or take some kind of action to corroborate the provider's judgement.

### 2.6. *Approval of guidelines by the Commission*

Paragraph 5 obliges the Commission to adopt guidelines on the application of this article by 2 February 2026, i.e. before the obligations of high-risk AI systems become applicable. The guidelines are set out in article 96, to which I refer to the commentary. They are the lightest and easiest instrument for the implementation of the Regulation to be adopted by the Commission.

Within the development of Article 6, the point where the guidelines will be most useful is the application of the exception in paragraph 3 and its use cases.

The guidelines should be updated and adapted to the state of the art. This includes not only technological developments but also the perception of risk or lack of risk.

The guidelines and their amendment cannot introduce amendments to

Article 6 (or Annex III) but only make explicit or interpret their content. Such amendment is regulated in paragraphs 6-8 of Article 6 and Article 7.

### 2.7. *The Commission's power to add or delete cases of AI systems exempted from high-risk status (paragraphs 6-8)*

In a further attempt to facilitate the adaptation of the Regulation to circumstances and to prevent the text from becoming obsolete, the Commission is delegated the power to amend it in specific respects. Paragraphs 6 and 7 specifically delegate the power to add or delete paragraphs from the list in Article 6(3). The aim is to add other cases in which it is understood that an AI system, despite being included in Annex III, does not pose a significant risk because it does not materially influence the content of the decision. Conversely, it also allows the deletion of cases, which means that certain systems that until now were understood not to pose a significant risk and were exempt from the high-risk condition are now subject to it.

This power of amendment is very limited, because it is bound by the concept of 'not materially influencing the outcome of the decision'. In addition, the Regulation imposes a strict standard for introducing such amendments since there must be 'concrete and reliable evidence' to support the amendment, both in one direction and in the other.

This amending power is entirely voluntary for the Commission (unlike the approval of guidelines, which is mandatory). It is possible that the Commission may consider that the list of cases in which an AI system, despite being included in Annex III, does not create a significant risk, is sufficient and appropriate, with no missing or excess paragraphs.

In addition, and thinking, above all, of downward amendments, i.e. those that exclude systems from high-risk status in situations that were not initially listed in paragraph 3, it is required that they do not 'decrease the overall slight of protection of health, safety and fundamental rights provided for by this Regulation', as well as logically ensuring consistency with the delegated acts in Article 7 (those amending Annex III) and with the state of the market and technological developments.

Delegated acts are subject to the conditions of Article 97, which implies time limits for the exercise (subject to renewal and revocation) and also control by the Parliament and the Council of the acts resulting from the delegation before they enter into force.

## 3. *Commentary to Article 7 (Amendments to Annex III)*

The AI Regulation regulates a constantly changing subject matter. In a sense, it is a tentative regulation, a trial run. To prevent it from quickly becoming obsolete, certain mechanisms are established to allow for its rapid reform by the Commission (under the control of the Council and the Parliament) to adapt to technological changes or to lessons learned from experience in applying the Regulation itself.

Article 7 allows for additions or deletions to Annex III so that the number of AI systems classified as high risk increases or decreases. The amendment can only affect the sub-sections of Annex III, the use cases within the areas listed in Annex III, as it is necessary that the system(s) to be added 'are intended to be used in any of the areas listed in Annex III'. No new areas can be added (their deletion will be referred to below).

### 3.1. *Relationship between Article 7 and Article 6(3)*

It is important to understand the difference between two different flexibilities, that of Article 6(3) and Article 7.

Both refer only to AI systems in Annex III. For AI systems that are high risk, because they are related to products subject to risk prevention rules, it will be in these rules that the scope can be adjusted to include or exclude any system.

Article 6(3) serves to exclude systems that are already included in Annex III. Its ability to exclude will be greater or lesser depending on the Commission's use of the delegated powers it receives *ex* Articles 6(6) and 6(7). However, Article 6.3 cannot ensure that an AI system not included in the initial wording of Annex III is classified as high risk. Instead, Article 7 serves to expand Annex III by including AI systems that were outside it (something that 6, para. 3 could never do) and also reduces Annex III by excluding systems initially included (something that coincides, in its result, with Article 6, para. 3).

Another difference is that Article 7 has a material or thematic focus (it will include or remove AI systems that are used in certain domains), while Article 6.3 has a formal or cross-cutting focus (it includes systems that perform a certain function or task, regardless of the domain in which they are used, and may be used in several domains).

Article 6(3) operates on its own, with the effect that systems listed in

Annex III are exempted from high risk classification when their provider concludes that they do not create a significant risk because they do not materially influence the decision outcome. In addition, the Commission may extend or reduce the list of cases, but it is always bound by the concept that the system does not materially influence the outcome of the decision. Article 7, on the other hand, only serves to allow the Commission to expand or reduce the list in Annex III. It may never be applied. It is also bound by material criteria to ensure that the general level of protection is maintained and that the level of protection is not thereby reduced or the objectives of the Regulation changed. Article 7 serves to ensure that Annex III is aligned with the principles of the Regulation.

In other words, Article 6 incorporates two different flexibility mechanisms. One is in the hands of the providers of AI systems, who must assess that the system does not create a significant risk, as it fits into one of the scenarios foreseen for safe systems. The other is in the hands of the Commission, which can (if it wishes) increase or reduce the list of scenarios, opening or closing the window left to the providers. Article 7, on the other hand, can only be applied (if it wishes) by the Commission by increasing or reducing the content of Annex III.

Both Articles 6(3) and 7 use the same concept as a criterion for the exclusion of systems or the increase or reduction of Annex III scenarios: the creation (or not) of a risk to health, safety or fundamental rights. But here, too, there are interesting differences. Article 6.3 only covers cases where the system does not produce a significant risk (since the effect of the article is to exempt high-risk systems from the requirements). In contrast, the criterion in Article 7 is that the system produces risks similar to those of the systems already included (when it comes to adding new cases), while the criterion for deleting systems is the same as in Article 6.3: that the system does not create a significant risk.

On the other hand, Article 6(3) and Article 7 can be applied cumulatively, in the sense that if Article 7 is used to increase the number of Annex III scenarios, Article 6(3) can be used to exclude any of those AI systems from the high-risk classification, if it is demonstrated that it does not create a significant risk because it does not materially affect the outcome of the decision by being included (in turn) in one of the scenarios listed in Article 6(3) (increased or reduced by the Commission using its delegated powers). On the other hand, if the Commission uses Article 7(3) to remove paragraphs from Annex III, it will no longer be necessary to use Article 6(3) to exclude any particular scheme, as all schemes will have been excluded.

Another important difference is that systems falling under Article 6(3) (including any modifications by the Commission using delegated powers) remain subject to the registration requirement (but not to the other requirements for high-risk systems), which makes it easier for authorities to check whether they are in fact eligible for this exemption. On the other hand, systems falling into categories in Annex III, which are removed under the Commission's Article 7 powers, will not be high-risk systems for any purpose.

### 3.2. *Criteria to be taken into account by the Commission for the inclusion in Annex III of AI systems not initially included therein*

The general criterion is that new paragraphs may be added to Annex III in the case of systems which pose a similar or comparable (or greater, of course) risk to those systems already included. It is, therefore, a matter of correcting gaps in Annex III.

The Regulation assumes that the fact that a potential area of AI use does not appear in Annex III does not mean that there was a conscious intention to exclude it. It may be assumed that there has been inadvertence or that it is a system or use case that appeared later. Annex III is conditioned by the material or principled rules of the Regulation (notably the concept of 'high risk to health, safety or fundamental rights'), and should therefore be amendable to ensure consistency with that concept. The fields or areas of Annex III are not subject to upward immovability (no new areas can be added through the fast track of Article 7, but require an amendment of the Regulation), but the Commission can increase or decrease the use cases.

This may be surprising in view of the tough negotiations that took place to fix the content of Annex III, but the fact is that, under Article 7, the Commission can include new sub-paragraphs (not new areas), provided there is no objection from the Parliament or the Council under the conditions of Article 97.

In order to assess whether a system or use case creates a similar or higher risk than systems or use cases already listed in Annex III, a number of criteria are mentioned in paragraph 2. However, these are not specific risk or lack of risk scenarios (as in Article 6.3, second sub-paragraph), but aspects in which the risk or lack of risk of an AI system manifests itself. It is important to note the difference. Therefore, unlike in Article 6.3, in order to include a new system or use case, it is not necessary to point to one of the 11 criteria in Article 7.2, but usually how the system scores on several of them will be taken into account.

The criteria are as follows:

*(a)* the intended purpose of the AI system. This must always be a purpose linked to one of the areas in Annex III. The inclusion of a new system in Annex III will require consideration of whether its purpose is likely to affect people's lives in a significant way.

*(b)* the extent to which an AI system has been used or is likely to be used. This criterion is not the clearest, as it is the severity of the risk rather than its frequency that is important. It cannot be applied in isolation. If the risk is high, the rarity of use is of little significance.

*(c)* the nature and amount of the data processed and used by the AI system, in particular whether special categories of personal data are processed. The use of specially protected data always generates a special risk.

*(d)* the extent to which the AI system acts autonomously and the possibility for a human to override a decision or recommendations that may lead to potential harm. This criterion is an 'old acquaintance' because it connects us to the 'fully automated decisions' concept in Article 22 of the GDPR. The passage of time has shown that an AI system may generate risks even if the decision is not taken automatically by the system but must be validated by a human operator if, due to various circumstances, there is a high probability that the human operator merely systematically validates the decisions suggested by the system, without subjecting them to a meaningful control or having a sufficiently viable alternative. In assessing this circumstance, it will be necessary to take into account many circumstances, including the workload of the human operator, the consequences for the human operator of deviating from the decision suggested by the system (starting with an increase in workload) and the availability (or otherwise) of the means for the human operator to make an independent decision.

*(e)* the extent to which the use of an AI system has already caused harm to health and safety, has had an adverse impact on fundamental rights or has given rise to significant concerns in relation to the likelihood of such harm or adverse impact, as demonstrated, for example, by reports or documented allegations submitted to national competent authorities or by other reports, as appropriate. Although the Regulation is a preventive rule, and it should not be necessary that harm has already occurred for it to be triggered, in the event that the application or use of a type of System produces harm, it would be almost obligatory, in addition to the entry into play of corrective measures (including civil or criminal liability), to apply preventive measures in the future.

*(f)* the potential extent of such harm or such adverse impact, in particular in terms of its intensity and its ability to affect multiple persons or to disproportionately affect a particular group of persons. Again, we find the element of the seriousness of the risk. It is not only the likelihood that matters but also the severity. A doubtfulrisk situation may be relevant and require the system to be classified as high risk if the potential harm is high.

*(g)* the extent to which persons who are potentially harmed or suffer an adverse impact are dependent on the outcome produced with an AI system, in particular because for practical or legal reasons it is not reasonably possible to opt-out from that outcome. This is another defining element of risk used to explain several of the sections of Annex III: those affected by decisions taken with the help of AI systems are highly dependent because they have no other way of accessing relevant goods or services.

*(h)* the extent to which there is an imbalance of power, or the persons who are potentially harmed or suffer an adverse impact are in a vulnerable position in relation to the deployer of an AI system, in particular due to status, authority, knowledge, economic or social circumstances, or age. Again, this paragraph of Article 7 acts as a kind of motivation (in an unexpected place) for Annex III. AI is an instrument for exercising powers (always subject to a more or less strict legal framework). The greater the difference in position between the holder of the power and those subject to it, the greater the need to qualify AI systems as high-risk.

*(i)* the extent to which the outcome produced involving an AI system is easily correctable or reversible, taking into account the technical solutions available to correct or reverse it, whereby outcomes having an adverse impact on health, safety or fundamental rights, shall not be considered to be easily correctable or reversible. This criterion is related to point (*d*). The possibility of correcting the results of the AI is a requirement of the Regulation, which provides for specific remedies, but depending on the type of action, this possibility will be greater or lesser

*(j)* the magnitude and likelihood of benefit of the deployment of the AI system for individuals, groups, or society at large, including possible improvements in product safety. The positive effects of an AI system are a reason to justify its acceptance and ultimately to assume that the associated risks may be outweighed by these benefits. However, the fact that the system is classified as high risk does not prohibit it, so these

benefits should not be an obstacle to its inclusion in Annex III if other criteria or reasons are pushing in that direction.

*(k)* the extent to which existing Union law provides for:

*(i) effective measures of redress in relation to the risks posed by an AI system, with the exclusion of claims for damages;*

*(ii) effective measures to prevent or substantially minimise those risks.*

This criterion is illustrative, although it should probably not tip the balance as to whether new use cases or systems should be included in Annex III. The existence of mechanisms to compensate for possible damage (beyond civil liability, which always operates) is complementary, not necessarily alternative, to the implementation of preventive mechanisms (which is precisely what the classification as high-risk systems entails). Precisely what defines or characterises the type of preventive legal rules, such as those established by the Regulation for high-risk systems, is the will to prevent damage, considering that it may be too serious to assume the risk and rely exclusively on repressive measures (such as civil or criminal liability).

### 3.3. *The Commission's power to remove AI systems from Annex III*

As with Articles 6(6) and 6(7), the powers granted to the Commission in Article 7 are two-way, to add systems to Annex III and to remove them.

This removal means that such systems will no longer be classified as high risk and they will no longer be subject to its requirements, but also to the duty to register.

For an AI system to be removed from Annex III, it must be justified that it does not create a significant risk (taking into account the dimensions or aspects listed in Article 7(2)) and that, taken as a whole, the removal does not lead to a reduction in the overall level of protection or health, safety and fundamental rights under Union law.

The question is whether it is possible to delete whole paragraphs from Annex III or only individual cases or even parts of them. In other words, can we delete Arabic numerals, letters or just parts of them? We have seen that Article 7(1) says that the Commission may not, in the use of these powers, add new areas (which we can identify with the Arabic numerals in Annex III). Firstly, because the rule speaks of removing high-risks 'AI systems', which points to concrete systems or use cases. Moreover, it seems practically impossible that any of the AI systems used in any of the large areas identified with Arabic numerals in Annex III could be considered to pose a significant risk.