THE PROHIBITION OF ARTIFICIAL INTELLIGENCE SYSTEMS THAT EVALUATE AND CLASSIFY PEOPLE BASED ON DATA THAT ARE UNRELATED TO THE CONTEXT IN WHICH THEY WERE GENERATED AND THAT LEAD TO DISCRIMINATION

Miguel Ángel Presno Linera[1]

Professor of Constitutional Law at the University of Oviedo

I. Introduction

The AIA includes a Recital 31 which explains that "AI systems providing social scoring of natural persons by public or private actors may lead to discriminatory outcomes and the exclusion of certain groups. They may violate the right to dignity and non-discrimination and the values of equality and justice. Such AI systems evaluate or classify natural persons or groups thereof on the basis of multiple data points related to their social behaviour in multiple contexts or known, inferred or predicted personal or personality characteristics over certain periods of time. The social score obtained from such AI systems may lead to the detrimental or unfavourable treatment of natural persons or whole groups thereof in social contexts, which are unrelated to the context in which the data was originally generated or collected or to a detrimental treatment that is disproportionate or unjustified to the gravity of their social behaviour. AI systems entailing such unacceptable scoring practices and leading to such detrimental or unfavourable outcomes should therefore be prohibited".

In other words, we are talking about the fact that after the entry into force of the Regulation, AI systems that generate qualifications or social hierarchies of people based on their behaviour or characteristics and that may give rise to discriminatory situations and, therefore, violate the principles of dignity and equality, will be prohibited in the European Union and may not be exported to other countries.

This is an extremely important issue because essential elements of the social and democratic rule of law are at stake, which would be seriously undermined if systems, such as those mentioned above, aimed at conditioning the behaviour of citizens and capable of generating, at the very least,

social and economic damage, if not physical and psychological harm, were allowed[2].

In the following pages we will develop the hypothesis of the certain risk posed by the systems we are dealing with[3] and the success, in our opinion, of their introduction in the European Regulation as a way of dealing with one of the growing manifestations of, in Shoshana Zuboff's words, "surveillance capitalism"[4].

Obviously, it is not a matter of excluding any type of personal "punctuation" that aims at behavioural modifications, since there are systems that are not only possible but surely necessary; for example, the points-based driving licence would be a good and well-known example: in the words of the Directorate General of Traffic, its objective "is to modify the behaviour and attitudes of offending drivers, to make them aware of the serious human, economic and social consequences of traffic accidents and to make them see the implication of their behaviour in accidents"[5].

In other contexts, the systems will not be prohibited but will be classified as "high risk"; thus, "AI systems used in employment, workers management and access to self-employment, in particular for the recruitment and selection of persons, for making decisions affecting terms of the work-related

---

[2] See Paquale, F. and Keats Citron, D., "The Scored Society: Due Process for Automated Predictions", *Washington Law Review*, vol. 89, 2014, pp. 1-33.

[3] On algorithmic risk San Martín Segura, D., *La intrusión jurídica del riesgo*, CEPC, Madrid, 2023, pp. 271 et seq.

[4] "Surveillance capitalism, m. 1. A new economic order that claims human experience as a free raw material to be exploited for a series of hidden commercial practices of extraction, prediction and sales. 2. Parasitic economic logic in which the production of goods and services is subordinated to a new global architecture of behavioural modification. 3. Unscrupulous mutation of capitalism characterised by vast concentrations of wealth, knowledge and power that are unprecedented in human history. 4. The fundamental framework for a surveillance economy. 5. As great a threat to human nature in the twenty-first century as industrial capitalism was to the natural world in the nineteenth and twentieth centuries. 6. The origin of a new instrumental power that imposes its domination on society and poses alarming contradictions for market democracy. 7. A movement that aspires to impose a new collective order based on absolute certainty. 8. Expropriation of crucial human rights that can perfectly well be considered a coup from above: an overthrow of the sovereignty of the people", *La era del capitalismo de vigilancia*, Paidós, Barcelona, 2022, 2nd edition, p. 9.

[5] "The points balance can change: being a good driver and/or taking awareness courses earns you points. Committing offences subtracts points, until you reach zero. If you reach this point, your licence will be revoked and you will not be able to drive any vehicle, although before this happens, you can recover points", available at https://www.dgt.es/nuestros-servicios/permisos-de-conducir/tus-puntos-y-tus-permisos/como-funciona-el-permiso-por-puntos/ (as of 18 March 2024).

relationship, promotion and termination of work-related contractual relationships, for allocating tasks on the basis of individual behaviour, personal traits or characteristics and for monitoring or evaluation of persons in work-related contractual relationships, should also be classified as high-risk, since those systems may have an appreciable impact on future career prospects, livelihoods of those persons and workers' rights. 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. 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. AI systems used to monitor the performance and behaviour of such persons may also undermine their fundamental rights to data protection and privacy" (Recital 57 of the European Regulation).

Similarly, "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. 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. 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. " (Recital 58).

As is well known, and as explained in more detail in other sections of this collective work, the classification of a system as high risk implies a series of obligations; among others:

"High-risk AI systems shall be accompanied by instructions for use in an appropriate digital format or otherwise that include concise, complete, correct and clear information that is relevant, accessible and comprehensible to deployers." (Article 13.2);

"High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which they are in use. 2. Human oversight shall aim to prevent or minimise the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse..." (Article 14.1 and 2);

"High-risk AI systems shall be designed and developed in such a way that they achieve an appropriate level of accuracy, robustness, and cybersecurity, and that they perform consistently in those respects throughout their lifecycle." (Article 15.1);

"... High-risk AI systems that continue to learn after being placed on the market or put into service shall be developed in such a way as to eliminate or reduce as far as possible the risk of possibly biased outputs influencing input for future operations (feedback loops), and as to ensure that any such feedback loops are duly addressed with appropriate mitigation measures" (Article 15.4.3).

II. The Chinese social credit system

The most talked-about social scoring system, which began to develop even before the current AI boom, is the Chinese social credit system (hereafter CSCS); as Lauren Yu-Hsin Lin and Curtis J. Milhaupt[6] explain, planning for a comprehensive social credit programme to complement China's weak legal system began in the 1990s with the more ambitious goal of addressing widespread fraud in the country's transition from central planning to a fledgling market economy. Those efforts culminated in 2014 with the joint publication by the Chinese Communist Party Central Committee and the Chinese State Council of the *Planning Outline for Building a Social Credit System (2014-2020)*, a comprehensive programme to assess the social credit of individuals, enterprises, government entities and other organisations.

Today, the social credit system is also the centrepiece of China's digital governance strategy, marking a shift towards a self-regulating market, i.e., one

---

[6] "China's Corporate Social Credit System: The Dawn of Surveillance State Capitalism?", *The China Quarterly*, Cambridge University Press, 2023, pp. 1-19; in particular, pp.2-4; available, as of 18 March 2024, at https://www.cambridge.org/core/journals/china-quarterly/article/chinas-corporate-social-credit-system-the-dawn-of-surveillance-state-capitalism/EC80AC0C-C9AE60D3D3C631A707A5CE54 (as of 18 February 2024); see also Rogier CREEMERS "China's Social Credit System: An Evolving Practice of Control", 9 May 2018, available, as of 18 February 2024, at https://ssrn.com/abstract=3175792 and http://dx.doi.org/10.2139/ssrn.3175792 (as of 18 March 2024).

in which actors are pressured or incentivised to conform their behaviour to party-State norms beyond the ordinary channels of law and regulation.

In the private sphere, Alibaba introduced its own personal credit scoring system*, Sesame Credit*, as early as 2015, to collect information on personal identity, credit history, contractual reliability, and social behaviours and relationships. Based on this information, participants are assigned social credit scores that are visible to others, and those with high scores are offered advantages, such as faster loan approval[7].

Zuboff explains that the *Sesame Credit* system generates a "holistic" assessment of a person's "character" through algorithmic learning that assimilates much more than whether they pay their bills and loans on time. Algorithms evaluate and rank purchases (for example, whether they are video games rather than children's books), educational titles, and things like the quantity and "quality" of friendships. Well-scored individuals receive distinctions and rewards from *Sesame Credit* customers in their behavioural futures markets. They

---

[7] On its website *Sesame Credit* explains: "The concept of a credit score may feel complicated, but in essence it looks simply at your payment history, amount of debt, how long you have had debt and how many recent applications you have made for credit accounts. Information about these items are reported to the three credit bureaus, Experian, TransUnion and Equifax, who compile your credit report. The information on your credit report is used to calculate your credit score. Your three-digit credit score captures your experiences with credit and debt and can help you track changes in your financial history over time, from the very first debt you encounter-such as the credit card you opened in college-up to the present. Credit score is a powerful tool that signals to prospective lenders your ability to make payments in a timely manner. This number is unique to you but publicly available under federal law to lenders considering you as a borrower. Your score can be a point of personal pride for good financial management and a point of public documentation. A credit score is an easy way to explain to another person or prospective lender that you can honor your commitment to make timely payments on outstanding debts. In turn, higher scores might lead a lender to extend interest rates lower than they would for consumers with less-favorable credit scores. You can get your credit score as part of a request for a credit report or independently of a credit report. A comprehensive solution is to open a free Credit Sesame account. This provides you with fast access to everything you need to know about your credit history, including your credit score. It includes helpful supporting information that makes sense of your score and report....

Legally, a variety of entities and people can request a copy of your creditreport, which is the information that feeds into your credit score. According to the Consumer Financial Protection.

Bureau (CFPB), this list includes: Businesses to whom you owe money, Government agencies.

Landlords, Employers, Insurance providers, Banks and financial providers, Legal entities (in the event of court orders, for example), Others you have authorised in writing to receive a copy"; available, as of 18 March 2024, https://www.creditsesame.com/knowledge-hub/what-is-credit-score/.

may be able to rent a car without paying a deposit, or receive more favourable terms on that loan or flat rental they apply for, or have their visa application expedited, or receive more prominent exposure on dating applications, and so on. However, some testimonies suggest that the privileges associated with a high personal reputation can suddenly turn into penalties for reasons completely unrelated to a person's behaviour in their role as a consumer: for example, if they have cheated on an exam at university[8].

Turning to the CSCS, it has two main features: the first is the collection of nationwide data from a wide range of regulatory bodies, central and local governments, the judiciary and private platforms. When fully operational, the system will collect two basic types of information: public credit information, generated by a company's interactions with government bodies and regulatory agencies (fines, judgements, business licences...), and market credit information, generated by a company's interactions with other market players (consumer complaints, data generated by credit rating agencies...). The data will be used in scoring systems run by local administrations, most of which are under construction.

The second main element of the CSCS is a regime of rewards and punishments (in the form of "red lists" and "black lists") maintained by government agencies. Some lists have a broad scope, such as non-compliance with court rulings, while others apply to specific sectors of the economy, such as food or medicine.

The inclusion in a red or black list is public; in the former case, it may entail various benefits, ranging from increased access to loans to reduced frequency of inspections or increased opportunities in public procurement processes and access to funding, especially for small and medium-sized entities. Blacklisting creates market barriers, such as restrictions on obtaining government approvals, increased frequency of inspections and prohibitions on obtaining funding. When an entity is blacklisted, its legal representative and the persons directly responsible for the violation will also be blacklisted[9].

---

[8] *Ob. cit.*, pp. 520 and 521.

[9] Yu-Hsin Lin and Curtis J. Milhaupt, ob. cit., pp. 3-4; more extensively, Schaffer, K., "China's social credit system: context, competition, technology and geopolitics." *Trivium China*, 16 November 2020, available, as of 18 March 2024, at https://www.uscc.gov/sites/default/files/2020-12/Chinas_Corporate_Social_Credit_System.pdf See also Lam T. "The People's Algorithms: Social Credits and the Rise of China's Big (Br)other", in Mennicken, A. Salais, R. (eds) *The New Politics of Numbers. Executive Politics and Governance*, Palgrave Macmillan, 2022; pp. 71-95; especially pp. 78 ff; Xu XU, Kostka, G. and Cao, X. "Information Control and Public Support for Social Credit Systems in China", The Journal of Politics, Vol. 84, no. 4, 2022, pp. 2231-2245, https://www.journals.uchicago.edu/doi/10.1086/718358 (as of 18 March 2024).

III. The development of qualification systems as a way of expanding surveillance capitalism

In a note at the beginning of these pages we collected the definitions of "surveillance capitalism" proposed by Zuboff, and the first two meanings, with some qualifications, seem to encompass practices such as those that characterise the Chinese social credit system: they would be, firstly, part of a new economic-political order that claims for itself human experience as free raw material exploitable for a series of hidden political, social, and commercial practices of extraction, prediction, and sales; secondly, they would be presided over by a parasitic logic in which the production of goods and services is subordinated to a new global architecture of behavioural modification.

It does not appear that the AIA's provision prohibiting AI systems that provide social ratings of natural persons for general use is intended to address the implementation or use in Europe of systems such as the Chinese social credit system: in EU countries and other democratic states, privacy and personal data enjoy a high level of legal protection and there is a higher degree of social concern about the threats that tools of this nature pose to these rights and to the free development of individual personality; as a result, practices typical of totalitarian societies have not developed, such as the so-called "*dang'an*, the personal file of multiple and varied aspects of each of the hundreds of millions of urban inhabitants that is updated from their childhood and for the rest of their lives. This "Mao-era system for recording the most intimate details of life" draws on up-to-date information provided by teachers, Communist Party officials and employers. Citizens have no right to check the contents of their own files, let alone challenge them"[10].

Notwithstanding these differences between the European and Chinese "ecosystems", it should be noted for the sake of nuance that, firstly, the so-called "privacy paradox" is present here: while individuals claim to be concerned about their privacy and value it highly, their decisions are significantly inconsistent with the value they profess, as they do little or essentially nothing to protect their personal data and thus their privacy[11].

---

[10] Zuboff, *ob. cit.*, p. 524.

[11] Artigot Golobardes, M. "Mercados digitales, inteligencia artificial y consumidores", *El Cronista El Cronista del Estado social y democrático de Derecho*, n.º 100, 2022, pp. 130 and 131; more extensively, Barth and De Jong, "The privacy paradox -Investigating Discrepancies between expressed privacy concerns and actual online behavior -A sytematic literature review", *Telematics and Informatics*, 34(7) (2017); Norberg, P. A. and Horne D. A. "The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors", *Journal of Consumer Affairs*, 41 (1), 2007, pp. 100-126.

And, secondly, although the consolidation of an authoritarian state surveillance capitalism such as China's does not appear to be forthcoming in Europe, this does not mean that there are not already practices of corporate surveillance capitalism which, to paraphrase Zuboff again, use human experience as a free raw material for a series of hidden commercial and labour practices of extraction, prediction, and sales, presided over by a parasitic logic in which the production of goods and services and labour relations are gradually subordinated to a new global architecture of behavioural modification.

In Creemers' words, this "tendency to socially engineer and "nudge" individuals towards "better" behaviour is also part of the Silicon Valley approach, which holds that human problems can be solved once and for all through the disruptive power of technology. Human beings are reduced to a set of numbers that indicate their performance on pre-set scales, in their eating habits, for example, or in their physical exercise regime, which they are then challenged to improve. The mere fact that information exists means that companies and governments will seek to exploit it for their own purposes, whether political or commercial. In that sense, perhaps the most shocking element of the story is not the Chinese government's agenda, but how similar it is to the path technology is taking elsewhere"[12].

And to mention a specific example in Spain of the use of data that a company has been using with the aim of avoiding the requirements of a dependent employment relationship and at the same time to behaviourally "push" workers to be available as long as possible in order to obtain more orders and, in short, higher pay, it is worth recalling, even if it is a bit lengthy, what was said by the Social Chamber of the Spanish Supreme Court in its ruling of 25 September 2020 on the status of GLOVO delivery drivers as salaried workers:

> "Factual background nº7: The company has established a rating system for "glovers", classifying them into three categories: beginner, junior and senior. If a delivery driver has not accepted any service for more than three months, the company can decide to downgrade him (Clause four of the service contract). The ranking system used by GLOVO has had two different versions: the *fidelity* version, which was used until July 2017, and the *excellence* version, used from that date onwards. In both systems, the delivery driver's score is based on three factors: the final customer's assessment, the efficiency demonstrated in the com-

---

[12] *China's chilling plan to use social credit ratings to keep score on its citizens*, CNN, 27 October 2015, https://edition.cnn.com/2015/10/27/opinions/china-social-credit-score-creemers/index.html (as of 18 March 2024).

pletion of the most recent orders, and the performance of services during peak hours, known by the company as "diamond hours". The maximum score that can be obtained is 5 points. There is a penalty of 0.3 points each time a delivery driver is not operational in the time slot previously booked by him. If the unavailability is due to a justified cause, there is a procedure for communicating and justifying this cause, avoiding the penalising effect... The delivery drivers who have the best score have preferential access to the services or errands that are coming in...

Eighteenth legal basis: ... In practice, this system of rating each delivery driver conditions his freedom of choice of timetables because if he is not available to provide services in the time slots with the highest demand, his rating decreases and with it the possibility of being assigned more services in the future and achieving the economic profitability he is seeking, which is equivalent to losing employment and remuneration. In addition, the company penalises delivery drivers by not assigning them orders when they are not operating in the reserved slots, unless there is a justified cause that is duly communicated and accredited.

The consequence is that delivery drivers compete with each other for the most productive time slots, with economic insecurity resulting from commission-based pay with no guarantee of minimum orders, which encourages drivers to try to be available for as long as possible in order to get more orders and higher pay.

Twenty-first legal basis – Glovo is not a mere intermediary in the procurement of services between shops and delivery persons. It does not merely provide an electronic intermediary service consisting in bringing consumers (the customers) and genuine self-employed workers into contact with each other, but coordinates and organises the production service. It is a company that provides courier and messenger services by setting the price and terms of payment for the service, as well as the essential conditions for the provision of the service. And it owns the essential assets for the performance of the activity.... The company has established instructions that enable it to control the production process. Glovo has established means of control that operate on the activity and not only on the result by means of algorithmic management of the service, the valuations of the delivery drivers and constant geolocation... To provide these services, Glovo uses a computer programme that assigns the services according to the valuation of each delivery driver, which decisively conditions the theoretical freedom to choose schedules and to refuse orders. In addition, Glovo has the power to sanction its delivery drivers for a variety of different behaviours, which is a manifestation of the employer's managerial power. Through the digital platform, Glovo carries out a real-time control of the provision of the service, without the delivery person being able to carry out his task without being linked to that platform...".

Other examples in the workplace can be mentioned; Todolí Signes explains, in an extensive quote, that "work in a *call centre* is one of the most affected by this high level of monitoring. Algorithms control the number of calls attended, their duration, pauses, even the content of the call through the detection of key words, tone of voice and intonation... The company *CallMiner* announces that its software can evaluate and score -and rank workers- in terms of professionalism, courtesy and empathy in the attention shown during calls... In the same way, supermarkets can measure how fast each cashier scans the products in the shopping basket and compare them with the rest of the workers for the purposes of remuneration, assigning work shifts, dismissing those who are less fast and making cashiers compete with each other to speed up the pace of work. Computer work, whether in the office or teleworking, is another area subject to absolute control of working times and subsequent evaluation by algorithms through productivity indexes. The company Crossover offers a tool called *WorkSmart* to monitor computers. This programme counts keyboard and mouse clicks, the computer screen, emails sent and even takes a picture every ten minutes via the computer's webcam. In this way, every second of inactivity with the computer -which does not mean that the worker is not thinking or working with a notebook- is penalised...

Face-to-face jobs are not spared from such productivity checks and rankings. They exist in transport, cleaning, hospitality, etc. The best known example is Amazon's monitoring of warehouse workers by measuring the number and speed of boxes packed, the number of steps taken in a day in the warehouse, bathroom breaks, or socialising, etc. Thus, by means of smart bracelets or chips in the boots, an exhaustive count is made of the work done and, together with other variables, a productivity index is drawn up which is used to generate automatic warnings (the bracelet vibrates or a message is sent to it) or to automatically dismiss people who do not reach a minimum productivity level. According to the data, 10% of Amazon's warehouse workers in the US have been fired because of the productivity index"[13].

Finally, and to briefly approach a different area such as insurance contracts, a classic example is the use of the credit rating of the insured to set

---

[13] "Artificial Intelligence will not steal your job, but your salary. Retos del Derecho del Trabajo frente a la dirección algorítmica del trabajo", *El Cronista del Estado social y democrático de Derecho*, no. 100, 2022, pp. 155 and 156; more extensively, and by the same author, *Algoritmos productivos y extractivos. Cómo regular la digitalización para mejorar el empleo e incentivar la innovación*, Aranzadi, 2023.

the premium in motor insurance, which, as María Luisa Muñoz Paredes recalls, gave rise to a rejection movement in the United States, following the finding by the Consumer Reports Association in 2015 that this factor was taken into account more than other more influential factors in risk, such as the driving record of the insured[14]. In this regard, Recital 37 of the AIA recalls that AI systems intended to be used for risk assessment and pricing in relation to individuals for health and life insurance can also have a significant impact on people's livelihoods and, if not properly designed, developed and used, can violate their fundamental rights and lead to serious consequences for people's lives and health, including financial exclusion and discrimination.

With the provisions contained in the Regulation, some of these tools, as mentioned above, will be considered "high risk" systems if the data used come from the context in which the results of the evaluations are applied, and may be prohibited if they come from different contexts and generate discrimination.

IV. The prohibition of certain systems that evaluate or classify natural persons

Article 5.1(c) of the Regulation has had the following course from the Commission's proposal of 21 April 2021 to the final wording, before the Common Position ("general approach") of the European Council on the AIA of 6 December 2022 and the amendments formulated by the European Parliament on 14 June 2023.

The following AI practices are prohibited

---

[14] " Big Data, AI y seguro: riesgos de inasegurabilidad y discriminación entre asegurados", *El Cronista del Estado social y democrático de Derecho*, n.º 100, 2022, p. 122; more extensively, and by the same author, ""Big Data" y contrato de seguro: los datos generados por los asegurados y su utilización por los aseguradores", in Huergo Lora, A. H (dir.): *La regulación de los algoritmos*, Aranzadi, Cizur Menor, 2020, pp. 129-162; "El "Big Data" y la transformación del contrato de seguro", in Veiga, A. B. *Dimensiones y desafíos del seguro de responsabilidad civil*, Cizur Menor (Aranzadi), 2021, pp. 1017-1051; on the use in insurance contracts of what Caty O'neil calls "weapons of mathematical destruction" see her book of the same title, Capitán Swing, Madrid, 2017, pp. 199 ff.

Commission	European Council	Parliament	Regulation
The placing on the market, putting into service or use of AI systems by or on behalf of public authorities for the purpose of assessing or classifying the reliability of natural persons over a given period of time on the basis of their social conduct or known or predicted personal or personality characteristics, in such a way that the resulting social ranking results in one or more of the following situations: (i) prejudicial or unfavourable treatment of particular individuals or entire groups in social contexts which are unrelated to the contexts in which the data were originally generated or collected; (ii) prejudicial or unfavourable treatment of certain individuals or entire groups which is unjustified or disproportionate to their social behaviour or the gravity of the latter.	The placing on the market, putting into service or use of AI systems for the purpose of assessing or ranking natural persons over a given period of time on the basis of their social behaviour or known or predicted personal or personality characteristics, in such a way that the resulting citizen score results in one or more of the following situations: (i) prejudicial or unfavourable treatment of particular natural persons or groups of natural persons in social contexts which are unrelated to the contexts in which the data were originally generated or collected; (ii) prejudicial or unfavourable treatment of certain natural persons or groups of natural persons which is unjustified or disproportionate to their social behaviour or the gravity of the latter.	The placing on the market, putting into service or use of AI systems for the purpose of assessing or ranking natural persons or groups of natural persons for social rating over a given period of time on the basis of their social behaviour or known, inferred or predicted personal or personality characteristics, in such a way that the resulting citizen score results in one or more of the following situations: (i) prejudicial or unfavourable treatment of particular individuals or entire groups in social contexts which are unrelated to the contexts in which the data were originally generated or collected; (ii) prejudicial or unfavourable treatment of certain natural persons or groups of natural persons which is unjustified or disproportionate to their social behaviour or to the gravity of the latter.	The placing on the market, putting into service or use of AI systems for the purpose of assessing or ranking natural persons or groups of persons over a given period of time on the basis of their social behaviour or known, inferred or predicted personal or personality characteristics, in such a way that the resulting citizen score results in one or more of the following situations: (i) prejudicial or unfavourable treatment of particular individuals or entire groups of individuals in social contexts which are unrelated to the contexts in which the data were originally generated or collected; (ii) prejudicial or unfavourable treatment of certain natural persons or groups of persons which is unjustified or disproportionate to their social behaviour or to the gravity of the latter.

Table prepared by the authors.

Although this is not one of the provisions that has undergone most changes between the Commission's proposal and the amendments adopted by Parliament, it is worth highlighting those that have been made and, first of all, one of the most important is the one relating to the person prohibited from introducing these systems: whereas the Commission's proposal mentioned "public authorities" or anyone acting "on their behalf", the Council's common position, as well as Parliament's amendment and the final wording resulting from the interinstitutional agreement remove this specification and the prohibition will affect both public authorities and private individuals, whether physical or legal, including, therefore, companies.

This modification seems very positive because the risks to be combated can come from both public and private parties and, as we have already seen, we find examples of the use of scoring systems by very important companies.

Secondly, the Commission's proposal referred to the assessment or classification of the 'trustworthiness' of natural persons, whereas the Council's common position, Parliament's amendment and the final text refer to 'assessing or classifying natural persons or groups of persons', i.e., the analysis is not limited to the "trustworthiness" of a person but extends to the person as such and, moreover, Parliament's amendment includes persons "or groups of persons" (e.g., consumer groups, workers, insured persons, etc.).

Thirdly, the Commission and Council texts, although not identical -the former refers to 'social conduct or personal haracteristics or personality traits' and the latter to 'social behaviour or personal characteristics or personality traits'- refer to 'known or predicted' characteristics, whereas the Parliament's amendment and the final wording of the Regulation also include 'inferred' characteristics, which is relevant because inferences are conclusions drawn from data processing and this is one of the properties of AI systems: the ability to extract new information from existing data.

Fourthly, while the Commission's proposal speaks of "social ranking", the Council and the Parliament use the term "citizen score", which will finally be included in the "Regulation", although it does not seem that the idea to which they refer is different: the ranking of people on the basis of known, predicted or inferred data.

The fifth issue to comment on is the generation of one or more of the situations described below that would justify the prohibition, the first of which is that it results in detrimental or unfavourable treatment of specific individuals or entire groups in social contexts unrelated to those in which the data were originally generated or collected. The score resulting from the processing of the data is considered to result in discrimination or, in the words of the texts under consideration, "detrimental or unfavourable treatment".

In this respect, and as we have seen at the beginning, the final wording of Recital 31 explains that "AI systems providing social scoring of natural persons by public or private actors may lead to discriminatory outcomes and the exclusion of certain groups. They may violate the right to dignity and non-discrimination and the values of equality and justice."

A significant qualification, to which we have already referred to above, is that the data generating such unfavourable treatment must have been obtained in contexts other than the one in which they would cause the detriment, but nothing would prevent their use in the context of origin; In this respect, it seems that data obtained in the context of an employment relationship could be used to carry out a scoring of those who work in that company or data obtained in a contractual relationship for the provision of services (for example, electricity supply) to establish a hierarchy of different prices to customers in different situations because one thing is the difference in prices and another discrimination; in this line, Law 3/1991, of 10 January, on Unfair Competition, in article 16.1 establishes that "discriminatory treatment of the consumer in terms of prices and other conditions of sale shall be considered unfair, unless there is a justified cause", i.e., different treatment for which there is justification would not be unfair, nor would the mere difference in prices[15].

However, the absence of discrimination or detrimental treatment contrary to the prohibition of Article 5.1(c) does not exclude that the data used are used without the knowledge or even the consent of the person concerned, which may place him in a position of particular vulnerability in digital markets. For this reason, "it is necessary to create mechanisms to prevent such vulnerability from materialising in the form of an expropriation of the contractual surplus that the consumer expected to obtain from the transaction and that only purely contractual instruments will not be able to recover"[16].

On the other hand, and as also noted above, the fact that the system in question is not subject to prohibition does not exclude that it can be qualified as "high risk" in the terms already seen.

Finally, and as has already been pointed out, what would be unacceptable is the use of someone's data to carry out evaluations or classifications in a context other than the one in which they were generated or obtained and

---

[15] See in this regard Muñoz Paredes, M. L. "Big Data, AI y seguro: riesgos de inasegurabilidad y discriminación entre asegurados", *El Cronista del Estado social y democrático de Derecho*..., p. 123.

[16] Golobardes, A. "Mercados digitales, inteligencia artificial y consumidores", *El Cronista El Cronista del Estado social y democrático de Derecho*...p. 135.

which would entail prejudice or unfavourable treatment[17]; thus, for example, a person's higher or lower credit rating should not be a conditioning factor for promotion within a company[18].

The second scenario that would justify the prohibition of an AI system is if it leads to "detrimental or unfavourable treatment of certain natural persons or groups of persons that is unjustified or disproportionate to their social behaviour or the seriousness of their behaviour". What is taken into account here is the way in which a natural person interacts with and influences other natural persons or society, resulting in unfavourable treatment that is either unjustified or the consequences are disproportionate to its severity; for example, that political opinions or ideological, religious, social, or cultural manifestations expressed on a social network generally imply a cause for exclusion from employment or expulsion from an educational establishment or that ratings of a worker's friendliness by customers are sufficient cause for dismissal or an unreasonable financial penalty.

[17] Obviously, comments or behaviour that are in breach of contractual good faith or offensive to the employer can have repercussions on the employment relationship (54.2 (c) and (d) of the Workers' Statute).

[18] Cathy O'NEIL provides numerous examples of the perverse results of the use of, among others, credit rating criteria in the labour and consumer spheres in *Weapons of Mathematical Destruction. How big data increases inequality and threatens democracy...* pp. 181 ff.

Lorenzo Cotino Hueso and **Diana-Urania Galetta**

(editors)

THE EUROPEAN UNION ARTIFICIAL INTELLIGENCE ACT

A Systematic Commentary

THE EUROPEAN UNION ARTIFICIAL INTELLIGENCE ACT

A Systematic Commentary

Lorenzo Cotino Hueso and Diana-Urania Galetta

(editors)

Editoriale Scientifica
Napoli