Article

Automated controls against fraud and corruption in Spain

Oscar Capdeferro Villagrasa

PhD in Law (Universitat de Barcelona). Associate Professor of Administrative Law at the Universitat de Barcelona. He has served as coordinator of the Master's Program in Advanced Public Procurement at the Universitat de Barcelona and as Co-Director of the postgraduate course on Good Governance, Prevention of Maladministration, and Public Integrity at the same university.

He has undertaken research stays at several foreign universities, including the Università di Bologna, the Università degli Studi di Ferrara, the Università di Roma LUMSA, and the University of Oxford.

He has been the principal investigator in several research projects related to public integrity and currently leads the research group on the Regulation of Algorithmic Systems and Artificial Intelligence.

He is the author of the book *El derecho administrativo y la prevención de la corrupción urbanística* (Ferrer Eguizábal Legal Award), as well as editor of several books. He is also author of book chapters and academic papers mainly on the legal aspects of public integrity and anti-corruption, and on the application of automation and AI systems in public administration. E-mail: ocapdeferro@ub.edu.

Received 01 April 2025, Accepted 02 March 2025

KEYWORDS:

Artificial Intelligence; Automated control; Tax Agency; Inspection; Public Administration.

ABSTRACT:

This paper examines the current state of affairs concerning the regulation of automated administrative activities related to inspection, supervision, and control in Spain, drawing on existing legal frameworks and selected case studies. It critically reflects on three key challenges associated with the use of AI and algorithmic systems in oversight functions: the lack of transparency, the occurrence of errors, and the potential unreliability or insufficient robustness of system outputs. The study explores several possible solutions and concludes by emphasizing the need to incorporate specific legal safeguards into national legal systems, in light of the insufficient protections currently provided by the EU AI Regulation for most systems employed in supervisory functions.

PALABRAS CLAVES:

Inteligencia artificial; Control automatizado; Agencia Tributaria; Inspección; Administración pública.

RESUMEN:

El trabajo presenta el estado de la cuestión en relación con la regulación de la actividad administrativa automatizada de inspección, vigilancia y control en España, a partir de la normativa y de algunos ejemplos de casos seleccionados. Se reflexiona sobre tres destacados problemas que se aprecian en el uso de IA y sistemas algorítmicos en el control: la falta de transparencia, los errores y la poca fiabilidad o posible falta de robustez de los resultados de los sistemas. Se exploran algunas soluciones posibles, y se concluye la necesidad de incorporación de algunas garantías en los sistemas jurídicos nacionales, a la vista de la insuficiente cobertura de garantías que parece ofrecer el Reglamento de IA de la UE para la mayoría de sistemas empleados en funciones de control.

MOTS CLES :

Intelligence artificielle ; Contrôle automatisé ; Administration Fiscale ; Inspection ; Administration publique.

RESUME :

Cet article examine l'état actuel de la réglementation des activités administratives automatisées en matière d'inspection, de surveillance et de contrôle en Espagne, à partir du cadre juridique existant et de quelques études de cas sélectionnées. Il propose une réflexion critique sur trois principaux enjeux liés à l'utilisation de l'intelligence artificielle et des systèmes algorithmiques dans les fonctions de contrôle : le manque de transparence, la survenue d'erreurs et la fiabilité incertaine ou l'insuffisante robustesse des résultats produits par ces systèmes. L'étude explore plusieurs pistes de solution et conclut à la nécessité d'intégrer des garanties juridiques spécifiques dans les ordres juridiques nationaux, compte tenu de la protection insuffisante qu'offre, en l'état, le Règlement européen sur l'IA pour la majorité des systèmes utilisés dans des fonctions de supervision.

Contents:

1 INTRODUCTION

The integration of artificial intelligence (AI) technologies into public administration has emerged as a rapidly expanding phenomenon. Far from being a speculative or future-oriented prospect, the deployment of AI systems is now a consolidated reality across numerous public sector domains. This transition from potential to implementation is evidenced by recent empirical findings, as reported in the study by Grimmelikhuijsen and Tangi (2024, 4), which draws on insights from over 500 public managers across various European countries. These findings highlight a paradigm shift whereby AI is no longer perceived as an aspirational tool but as an operative component within contemporary public services.

Currently, there even appears to be a degree of competition among public administrations to position themselves as leaders in proactivity and technological innovation, with AI digital systems gaining visibility in the media and being recognized through awards and distinctions. Notable examples include the award-winning BRAVA and ERICCA systems[1], developed to detect anti-competitive practices in public procurement (Jiménez, N. 2022), as well as VeriPol, an AI tool designed to identify false police reports. However, the latter, despite having received an award in 2017[2], ceased to be used in 2024 on the grounds that its results lacked validity in judicial proceedings.

Moreover, it is quite common in citizens' daily lives to interact with administrative chatbots (Aoki, N. 2020), or to receive a fine automatically issued for speeding after being caught by a radar (Carnis, L. 2008; Snow, A. 2017), to name just two frequent examples of AI usage in the public sector[3].

Among the various applications of AI in the public sector, one that stands out, particularly in the Spanish context, is its use in administrative functions related to oversight, inspection, and regulatory enforcement. These functions aim to detect potential infractions and ensure the application of appropriate sanctions. Within this domain, the development of AI systems for anti-fraud and anti-corruption control has been especially prominent. A comparative study across 30 European countries (including EU member states, the United Kingdom, Norway, and Switzerland) identified 14 examples of national-level initiatives employing AI technologies specifically designed to prevent and detect fraud and corruption (van Noordt & Misuraca, 2022, 9).

This paper focuses specifically on such systems, which will be referred to as automated administrative control systems, with particular attention given to cases related to fraud and corruption prevention. Most of them are AI-based systems. In this field, the advancements made by the Spanish Tax Agency and the Labour Inspectorate in modernizing their control mechanisms over recent years are especially noteworthy. These systems vary considerably in nature and exhibit different levels of automation. For the purposes of this paper, all systems that contribute to the automation of control tasks—whether fully or partially automated—will be considered under the category of automated systems. This includes the fully-automated systems that can detect infractions on their own -without human intervention- and draw a sanctioning proposal, but also includes what might be more

[1] See https://acco.gencat.cat/ca/detall/noticia/20221103_premis_alfons_ortuno_ericca_acco and https://blog.cnmc.es/2024/04/11/premio-a-brava-cnmc-contra-el-fraude-en-contratacion-publica/.

[2] See https://www.interior.gob.es/opencms/es/detalle/articulo/El-secretario-de-Estado-de-Seguridad-ha-presidido-la-entrega-de-los-Premios-de-Investigacion-de-la-Fundacion-Policia-Espanola/ and

https://civio.es/transparencia/2025/03/19/la-policia-nacional-deja-de-usar-veripol-su-ia-estrella-para-detectar-denuncias-falsas/

[3] A more exhaustive list can be found in the study by Brizuela, A. Et al. (2024), which starts from the study of 940 AI use cases in Europe.

accurately described as semi-automated systems (Muñoz, D. 2025, 502), which are used primarily to prioritize inspection tasks based on risk indicators or behavioral patterns, with final verification still carried out by human inspectors.

Regarding the structure of the paper, I will begin by outlining the main legal and regulatory frameworks applicable to AI systems used for control functions, particularly those targeting fraud and corruption. I will then highlight several illustrative cases. Following this, I will examine key challenges associated with the design, implementation, and governance of these systems, and propose several recommendations to mitigate such challenges and promote their fair use and regulation.

2 THE REGULATION OF AI IN SPAIN: FRAMEWORK FOR AUTOMATED ADMINISTRATIVE CONTROL ACTIVITIES

When examining the Spanish legal framework for references related to the use of AI and automated systems in administrative control activities, the references are limited. There is a single article that regulates automated administrative actions in a transversal manner for public administrations, as well as a law on equality of treatment and non-discrimination, which includes some safeguards for AI systems, specifically aimed at minimizing biases and ensuring transparency. Additionally, there are specific provisions related to automated control in certain sectors where oversight is central, such as taxation and labor. Furthermore, there are some very specific regulations governing tools that perform control functions, as well as legislation from the Autonomous Communities, both general[4] and concerning specific AI-related issues[5]. Let us review these references, except those referring to the autonomous communities, which are excluded except in the case that they refer specifically to automated control tools, which would be the case of the Valencian Law 22/2018.

Currently, the general and foundational regulation of automated administrative action in Spain is found in Article 41 of Law 40/2015, of October 1, on the Legal Regime of the Public Sector. This article establishes substantive regulatory safeguards that apply specifically to automated administrative action, including the requirement to designate the competent authority for the purpose of challenging automated decisions, as well as provisions for defining the specifications, programming, maintenance, supervision, quality control, and, where appropriate, auditing of the information system and its source code. However, these safeguards are tied to the definition of automated administrative action, which is understood, within the regulatory framework of the law, as "any act or action carried out entirely through electronic means by a Public Administration within the framework of an administrative procedure and in which no public employee has directly intervened". As is evident, this definition of fully automated action is too narrow to encompass the full range of AI applications in the public sector, since not all AI systems meet these criteria (Cerrillo, A. 2024, 539–540; Gamero, E. 2023: 404). For instance, it excludes systems that require subsequent human validation (Valero 2019, 87), such as AI tools used to prioritize cases or files for investigation. These are sometimes considered part of the preliminary inspection activities prior to the initiation of formal administrative procedures, as set out in Article 55 of Law 39/2015, of October 1, on the Common Administrative Procedure of Public Administrations, as noted by several scholars (Miranzo, J. 2023, 112; Ponce, J. 2019; Capdeferro, O. 2019, 19).

[4] Such as the Galician Law 2/2025 (Ley 2/2025, de 2 de abril, para el desarrollo e impulso de la inteligencia artificial en Galicia).

[5] For example, article 16.l) of Valencian Law 1/2022 (Ley 1/2022, de 13 de abril, de Transparencia y Buen Gobierno de la Comunitat Valenciana) includes obligations regarding AI-systems' transparency.

Moreover, the set of safeguards on article 41 can be viewed as insufficient to address the complexities and risks associated with AI. Indeed, the inadequacy of existing administrative regulation in the face of technological innovation has been noted some time ago by scholars (Valero, J. 2019, 84–85). Additionally, as previously mentioned, the narrow definition set forth in the first section further limits the scope of these already scarce and imprecise safeguards to only a subset of AI systems used in the public sector.

Subsequently, and without significant innovation, the regulatory framework set forth in Article 41 of Law 40/2015 was further developed in Article 13 of Royal Decree 203/2021, of March 30. This article specifies that automated administrative actions must be authorized by a resolution issued by the head of the competent authority, and it establishes the obligation to publish that resolution, including information on available remedies and the requirement to implement appropriate safeguards for the rights, freedoms, and interests of affected individuals. This publication requirement can be seen as a partial solution to the current lack of transparency surrounding the use of AI in the public sector. Furthermore, the inclusion of the obligation to adopt adequate safeguards is undoubtedly a positive development. However, it is important to note that these obligations apply only to the General State Administration and do not extend to regional or local administrations (see Article 13.2 of Royal Decree 203/2021).

One year later, Law 15/2022, of July 12, on comprehensive equality of treatment and non-discrimination, introduced for the first time in Spain a regulatory framework applicable to AI on all areas of administrative activity. Notably, this law explicitly references the term "artificial intelligence," although only in two articles, of which just one provides substantive regulation. Article 3 affirms that the law applies to the field of "Artificial Intelligence and the large-scale management of data, as well as other spheres of analogous significance." The key provision is Article 23, which specifically addresses the issue of AI. This article is particularly noteworthy because it introduces the need to consider criteria such as bias minimization, transparency, and accountability in the algorithms used by public administrations in decision-making processes. It also sets forth the goal of promoting ethical, trustworthy, and fundamental rights-respecting uses of AI.

I will now focus on the provisions adopted in specific areas related to automated control activity. In terms of control, the two paradigmatic areas with the most regulatory and practical development in Spain are the tax and labor sectors.

The Spanish Tax Agency, for its part, already has extensive experience with AI-based tools and has even published the guidelines for its approach to the use of AI in its Strategic Plan for 2024–2027. Several authors have focused in other works on the use of AI by the Tax Agency, particularly in its efforts to combat fraud (Pontón 2024: 70–71). A clear indication of the Tax Agency's early interest in developing automated administrative control tasks is Law 58/2003 of December 17, the General Tax Law, which was enacted in 2003 and already included, in its original 2003 version, provisions for automated tax procedures in Articles 96 and 100[6].

In the labor sector, regulations allowing for the use of automated decision-making in certain administrative actions have existed since 2009[7]. Regarding inspection functions in this area, Article 53 of Royal Legislative Decree 5/2000 of August 4, which approves the revised text of the Law on Infractions and Sanctions in the Social Order, provides, following its 2021

[6] An analysis of these specific articles can be found in Oliver (2021, 7-8).

[7] In particular, in the area of unemployment benefits, this possibility was introduced by Royal Decree-Law 10/2009, of August 13, which regulates the temporary unemployment protection and insertion program, in its first final provision 1. On this topic, with reference to other developments in the labor and social security sector, see Goerlich (2024, 230-233).

amendment (Goerlich 2021: 22–25), that infraction reports issued by the Labor and Social Security Inspectorate may be automatedly generated.

This approach is further developed in Royal Decree 928/1998 of May 14, which approves the General Regulation on procedures for imposing sanctions for infractions in the social order and for proceedings concerning Social Security contributions. The regulation includes specific provisions on automated actions in Articles 43 and following, incorporating notable safeguards, such as the possibility of human intervention in cases where the automated system's resolution is contested, which will be discussed in more detail below.

Finally, there are specific legal provisions that regulate specific tools used to assist in control functions, such as the law governing the SALER system and the ministerial order regulating the MINERVA system. These two Systems, along with others not explicitly addressed in legal rules, are the subject of the next section, which focuses on cases of use of AI tools for automated control in Spain.

However, before addressing the specific use cases, it is important to note that European regulations also apply domestically. In this regard, particular attention must be given to Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act, hereinafter EU AI Act).

This legal instrument prohibits certain uses of AI (Art. 5 EU AI Act) and sets out safeguards for the permitted uses, including additional requirements for systems classified as high-risk (Art. 6 and Annex III EU AI Act). In general, AI systems used by public administrations for control purposes would not fall under the high-risk category, unless their use could lead to the withdrawal of a public service or an essential benefit[8], which can lead to the necessity of thinking about legal safeguards for these systems on a national-based scope.

3 BRIEF OVERVIEW OF SOME AUTOMATED CONTROL SYSTEMS IN SPAIN

Information about the use of AI systems by the public sector in Spain is limited, despite existing transparency obligations for public authorities. In fact, it could even be said that opacity is widespread and extends beyond control-related activities. As L. Cotino has pointed out, “In the case of the public sector, it is particularly concerning that we do not even know what AI systems exist and how they impact government action, service delivery, and our rights” [translated by the author] (2023, 24).

It is therefore not surprising that few AI-based digital systems used by public administrations are publicly disclosed. Some systems become known because their implementation is approved through legal provisions, and, as such, their adoption is made public through official publication of the law or regulation. This is the case, for example, of the SALER system, which was approved by Valencian Law 22/2018 of 6 November on the General Inspection of Services and the alert system for preventing misconduct within the administration of the Generalitat and its instrumental public sector. This was the first

[8] “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” (Annex III.5.a EU AI Act). This provision could be interpreted in such a way that automated systems for detecting infractions that may result in the termination of a subsidy as part of the sanctioning response, would indeed be considered high-risk systems. Examples might include systems used by the Labor Inspectorate to detect fraud in subsidies or other unemployment benefits. I have addressed the potential application of the EU AI Act to this type of control systems in greater detail in another work, to which I refer the reader for further discussion on this issue (Capdeferro, O., 2025).

algorithmic control system in Spain to be approved by law. It is part of an innovative risk-based anti-corruption control framework. The algorithmic system generates alerts concerning potential misconduct, which are then reviewed and prioritized, if appropriate, for further verification or follow-up by the human inspection team. Valencian Law 22/2018 also includes a sanctioning regime to ensure the proper use and functioning of the system.

The use of another system has been approved through regulation in order to prevent corruption cases. It is the MINERVA system, governed by Order HFP/55/2023 of 24 January, concerning the systematic risk analysis of conflicts of interest in procedures executed under the Recovery, Transformation, and Resilience Plan. The MINERVA AI system is a preventive anti-corruption tool aimed at the early detection of potential conflicts of interest. It is a data mining system managed by the Spanish Tax Agency that generates red flags in situations where a public decision-maker may have a personal, family, or business interest in public procurement procedures or the awarding of grants involving Next Generation EU funds.

Without normative character but also subject to publication, there are resolutions that approve the use of automated systems in the tax sector. One recent example is the Resolution of September 23, 2024, by the Directorate General of the Spanish Tax Agency, which approves new software applications for automated administrative procedures.

Another example of automated technology from the Spanish Tax Agency of interest that combines AI technology with behavioral sciences (specifically, nudging) is a system developed by the agency in 2021 for the annual income tax declaration campaign[9]. This system is capable of predicting errors in the preparation of personal income tax self-assessments. Upon detecting one of these errors, the system is designed to automatically send a message to the taxpayer, warning them that there is a high likelihood that their declaration is incorrect when any data provided or modified by the taxpayer does not match the information available on the Tax Agencs databases. Additionally, this tool is also designed for identifying taxpayers at high risk of having submitted an incorrect declaration by altering the data available to the Tax Agency.

Finally, some other tools can be indirectly known through the public accountability reports published by public agencies and institutions. For example, the 2023 Annual Report of the Labor and Social Security Inspectorate, published in December 2024, refers to the modernization of ITSS management through specific software Applications, stating that it has also marked a milestone with the creation of the Fraud Detection Tool, which is capable of identifying signs of irregularities to direct inspection efforts through the mass processing of data (p. 10), referring to the automated system used by the Inspectorate since 2015 (p. 45).

In addition to these channels, some systems are mainly made known to the public through news reports, generally highlighting the innovation they bring to the areas in which they are employed, such as the VERIPOL police system or the ERICCA and BRAVA systems for detecting collusive practices in public procurement, which are used by competition control authorities.

These tools, of course, streamline control tasks by enhancing administrative efficiency (Coglianese, C. & Ben Dor, L., 2021, 827) and enable large-scale, indiscriminate oversight, as they can potentially monitor all subjects or activities subject to control. However, they are not without risks. In the following section, I reflect on some of these risks or issues that should be taken into account when designing appropriate regulations for the use of automated systems in public administration control activities.

[9] See Resolution of January 19, 2021, of the Directorate General of the Spanish Tax Agency, approving the general guidelines of the Annual Tax and Customs Control Plan for 2021.

4 SOME ISSUES FOR WHICH LEGAL SYSTEMS SHOULD ESTABLISH ADEQUATE SAFEGUARDS

Up to this point, I have outlined the general and specific regulatory framework concerning AI and automated systems, particularly those used for control activities, along with a few notable practices. However, over the years, AI-based automated systems used for control purposes have undoubtedly revealed several problems. In order to correct or minimize these issues, it is essential to ensure that proper safeguards and mitigation mechanisms are in place to address the unintended impacts of AI, as well as to provide effective oversight and monitoring of AI systems, and adequate protection for individuals affected by AI-driven outcomes.

In the following sub-sections, I will highlight some of these problems and suggest measures that could help reduce the negative impacts of automation in administrative activities related to inspection and the sanctioning of unlawful behavior.

4.1 SAFEGUARDS AGAINST ALGORITHMIC OPACITY

The first issue I want to highlight is the lack of transparency in the use of these systems.

Algorithmic transparency is often discussed as a goal or desired outcome, since, unfortunately, what predominates is algorithmic opacity. In addition to the inherent problems of explainability and transparency found in the design and functioning of so-called black box AI systems (Brożek, B. et al., 2024), control-related activities involve an added layer of opacity: authorities responsible for oversight are generally reluctant to disclose information about the AI systems they use.

This trend is, unfortunately, widespread, as explained by Rachovitsa, A. & Johann, N., (2022, 11), based on the case of the well-known SyRI system from the Netherlands: “the Court found that the legislation provided little, if any, insight into the risk model and risk indicators used, the objective factual data that could justifiably lead to the inference of an increased risk or the data processed in SyRI projects. Crucial information concerning the algorithm’s use was deliberately kept secret—an instance of intentional opacity. The Netherlands’ refusal to disclose additional information, on the grounds that citizens would otherwise ‘game the system’, is an argument invoked by many countries with regard to different uses of algorithmic systems in different areas”.

The underlying reason, well known and discussed for years, is the belief that understanding how control systems function would allow offenders to adapt their behavior in order to evade detection by the automated system (Kroll, J. A., 2017, 658). Thus, it may be understandable, to some extent, the opacity regarding the exact functioning of the system, particularly in terms of the risk indicators used, the specific data considered relevant in that context, the suspicious behavior patterns being targeted, or the exact variation threshold that triggers the alert.

However, I believe it is important to consider that the list of infringements is, of course, public and outlined in the relevant laws and regulations. Therefore, the irregular behaviors that the system may detect could be made public without jeopardizing the system's effectiveness in identifying potential infringements. In fact, on a case-by-case basis, it could be assessed whether it would be feasible to publish, in a summary form, the indicators and data used to achieve the detection objective intended for the algorithmic system, as this information is unlikely to reveal specific details that could be used to circumvent the algorithm's control.

For example, consider the case in which it is made public that a system from the Tax Agency seeks potential cases of companies' tax fraud based on data comparisons of income declarations and company revenue. It could be stated that these data are cross-referenced with the national average revenue and income levels of similar companies in order to detect suspicious cases of fraud due to low declared levels. With this information, a company would hardly find any reference to alter its fraudulent behavior in order to avoid the algorithm. However, it could potentially change the behavior of the suspected offender if the public information also includes the algorithm's reference index. For example, it could be publicly stated that the system detects companies declaring revenue lower than 30% of the average of similar companies in the province. In conclusion, the specific alert threshold could be kept confidential, while the more general information mentioned earlier could be published, which would not, in principle, offer any indirect guidance for evading the intended control.

A final element of transparency that could be considered is the ability to determine whether the algorithmic system is achieving the expected results. That is, determining whether its functioning truly represents an efficient way to modernize controls or whether, instead, it has been a waste of economic resources that fails to meet the anticipated outcomes. It is not common to know the functioning of these tools, as the usual practice in Spanish administrations is that the activity reports of institutions and agencies using them remain silent on the matter. Nevertheless, on occasion, some relevant information can be found, as in the case of the reports from the Valencian General Services Inspectorate, in relation to the SALER system. These reports show some indications of how SALER operates, and it is clear that, at the moment, it faces some challenges in functioning effectively, particularly due to insufficient data or errors in the data[10].

In this regard, it may be necessary to strengthen the right of access to public information recognized for citizens.

In general, in Spain, the independent authorities responsible for safeguarding the right of access to public information tend to be favorable to recognizing that the digital systems used by public administrations are public information that can be accessed to.

Some resolutions from the Transparency and Good Government Committee (Consejo de Transparencia y Buen Gobierno) have upheld, either fully or partially, the right to access information about various digital systems used by public administrations. For example, the resolution of August 23, 2023 (ref. 551-2023), which upheld the claim regarding the right to know how the VioGén system, used to predict the risk of gender-based violence, operates. In relation also to this system, the resolution of November 29, 2024 (ref. 1449/2024) upholds the right to access data from the VioGén register concerning gender-based violence fatalities in 2023, as a means of controlling the effectiveness of the digital system. Additionally, the resolution of January 16, 2024 (ref. 1659-2023) recognizes the right to access the use cases, functionalities, and technical specifications that allow understanding the operation of the MAX system, employed by the Labour Inspectorate to monitor overtime hours.

However, the most relevant case to date in Spain regarding algorithmic transparency is the BOSCO case, which has been the subject of a highly publicized litigation and is still pending resolution after an appeal was filed before the Supreme Court, in which the Civio Foundation's right to access the source code of the BOSCO system has already been denied twice on the courts of justice[11].

Beyond the specific cases mentioned, it is worth noting that transparency legislations in Spain often limit access to public information precisely in the areas where the use of AI and

[10] See Inspección General de Servicios (2023, 33-35).

[11] This system is not related to controls in the sense used in this paper. A detailed explanation of the case could be seen in De la Cueva (2025).

automated systems is most expanding in the public sector, which are those related to the control exercised by the administration through inspections or other forms of prevention, detection, and sanctioning of offenses. Taking as an example the Transparency, Access to Public Information, and Good Governance Act 19/2013, of December 9, Article 14 specifies that, among other reasons, the right to access public information may be limited if granting access would result in harm to "administrative functions of surveillance, inspection, and control" and to the "prevention, investigation, and sanctioning of criminal, administrative, or disciplinary offenses."

These types of clauses, which justify opacity regarding information related to administrative control, can, in practice, represent a significant limitation to understanding the existence and functioning of these systems.

4.2 SAFEGUARDS AGAINST ERRORS OF THE AUTOMATED SYSTEM

The second issue I want to highlight is the errors. Automated systems can generate erroneous results, producing false positives or false negatives[12]. A false positive may lead to unjustified sanctions, while a false negative could allow illicit conduct to go unnoticed.

In particular, the false positive is problematic for those being monitored, as they may be subject to invasive inspection actions, which could even affect fundamental rights, such as if a home entry and search is carried out, or they may be unfairly sanctioned based on indicators or facts that were mistakenly identified or misinterpreted by the algorithmic system. In this regard, an example that clearly illustrates the system's error in interpreting reality, leading to widespread false positives, occurred in the automated vehicle access control in low-emission zones.

Indeed, since it was put into operation in the metropolitan area of Barcelona, the image-capturing technology allowed an algorithmic System (Chronos-Eco) to monitor vehicle license plates, cross-referencing the plates identified in the images with the list of vehicles authorized to access the area. Soon, unjust sanction proposals began to proliferate against owners of older, more polluting vehicles that were not permitted to enter the low-emission zone, alleging that they had entered the area. However, in fact, those vehicles had not entered under their own power but were being towed by tow trucks, on their way to repair shops or even to scrapyards. These kind of errors of the Chronos-Eco system could be easily seen and fixed if human-supervision was allowed to control the results of the automated system, as claimed by the municipal ombudsman (Sindicatura de Greuges de Barcelona, 2024).

In response to this phenomenon, I propose at least two safeguards that should be taken into account and promoted. These two safeguards are: providing information about the automated nature of the control activity, and the possibility of appealing the automated decision promptly and to a human being.

Regarding both safeguards, in Spain, we can find some indications in the legislation, although, discouragingly, these protective or safeguarding measures against the automation of public administration are regulated in a sector-specific rather than a general manner. As a result, their application is limited to specific sectoral control activities or to the use of particular AI-based tools.

[12] A false positive occurs when the system alerts from a fact that is not real; and, on the contrary, a false negative is the failure of the system to alert from a fact that should have been detected (Baader, G. & Krcmar, H., 2018, 1).

Firstly, a positive example can be found in Royal Decree 928/1998, of May 14, following its 2022 amendment. With that amendment, a specific regulation was introduced regarding the automated control activities carried out by the Labour and Social Security Inspectorate. As for safeguard elements, the two I previously mentioned are included: the provision of information about the automated nature of the control, and the free appeal to a human public officer.

Firstly, infringement reports generated automatically must explicitly indicate that they are the result of an automated administrative action, and must also specify the means used to verify the facts that support the report (Article 45). Secondly, it is also established that any objections made against the infringement reports must be reviewed by inspection staff (Article 47.3), that is, by a human public officer. These objections can be submitted by the concerned party without needing a lawyer, and the objection process is free of charge, which means that this human oversight or review may come at no cost to the individual. Naturally, if the person chooses to obtain legal counsel, this may incur costs for them, which cannot be recovered from the administration.

Knowing that the detection of alleged infractions (and even the imposition of the corresponding sanction, in some cases) has been carried out automatically could be very useful when preparing the objection, which could be based on simply arguing that the system made an error, attaching basic information or documentation that supports the claim that no infraction was committed[13]. Furthermore, that objection must be resolved by a human inspector, which opens the door to recognizing the false positive, annulling the imposed sanctions, and, ideally, even promoting an internal review of the AI system to identify the source of the error. It should be added, however, that one element is notably absent: a specific time limit for resolving objections to automated system errors, which in some cases can be very obvious, as demonstrated by the fines in the low-emissions zone case. It could also be argued that the workload reduction achieved through automation might free up inspectors' time to carry out this kind of review more promptly.

But this regulation is not the only one in Spain that includes measures to mitigate the effects of false positives. At least two other systems include provisions in their governing regulations to limit or correct false positives before the system's error produces definitive consequences.

In particular, I refer to the MINERVA system, which I have already mentioned. According to Order HFP/55/2023, if the automated system generates a positive by signaling a red flag that the affected person disagrees with, they can submit a reasoned objection, refusing to acknowledge the validity or accuracy of the result produced by the automated control system. If an objection is filed against the identification of a conflict of interest flagged by the system, a procedure is initiated that must be resolved by the hierarchical superior of the public decision-maker affected by the red flag. This person is ultimately responsible for validating or discarding the result provided by the MINERVA system.

Similarly, the SALER system assigns to human inspectors the task of manually classifying all alerts generated by the automated system, allowing them to decide in each individual case whether to initiate an investigation or consider the alert as a false positive (Article 31 of Valencian Law 22/2018).

Despite these good practices, these safeguards against errors in automated systems should be explicitly included in the cross-cutting or general regulation governing the functioning of

[13] Requiring a higher burden of proof would be totally disproportionate, since access to the AI systems of the administrative control authorities is generally prohibited or, at least, highly limited. As I said before, usually opacity about these systems is maintained voluntarily, so only authorized personnel would have access to check if the system is working properly.

public administrations, since errors can occur in any control action carried out by a public administration through automation—not only in a specific sector or region.

In the Spanish case, this gap could be addressed through an amendment to Article 41 of Law 40/2015, which has already been mentioned. This article, which generally regulates automated administrative activity, could be modified to stipulate that automated administrative acts or decisions must always be subject to human review if objections or appeals are submitted, and that such human supervision must be issued promptly. Additionally, Article 64 of Law 39/2015 could be amended to require that, in decisions initiating sanctioning procedures, all infraction reports or other acts that identify violations or impose penalties must state that the facts were established by automated means, when that is the case.

4.3 Safeguards Against the Lack of Robustness in the Results Generated by Automated Systems

Lastly, I would like to raise a problem that is by no means minor: How much trust do automated systems inspire in judges?

The SyRI case in the Netherlands is well known, which was annulled by the courts (Rachovitsa, A. & Johann, N., 2022). But beyond that case, I believe there are also rulings in Spain that suggest there is still skepticism within the judiciary regarding the validity of the results produced by automated systems.

It is therefore worth asking whether the system can generate results that can be considered valid evidence to materially justify the initiation of investigative actions, especially those that are more invasive or detrimental to fundamental rights. This issue has arisen particularly before courts in relation to requests for home searches submitted by oversight bodies, such as the Tax Agency or competition authorities.

Thus, the Supreme Court, in its ruling of November 15, 2021[14] held that the entry into a home for the purpose of conducting verification actions to be unjustified in a case where the tax administration's justification was based on "mere conjectures or inferences from purely statistical data—i.e., generic ones." It was criticized the fact that the appealed ruling "remained silent on the seriousness or reliability of such evidence and the necessity of entering the home as the only means of verifying its conformity with reality".

In this case, the Tax Agency had requested judicial authorization to enter the home of a company for verification activities without prior notice to the inspected subject (*inaudita parte*), suspecting that an infringement might be occurring (e.g., by concealing undeclared sales). This suspicion was based on the fact that, when comparing the company's billing data with the national average for that sector, an excessively low profitability was detected.

It is worth noting that, in this case, the initiation of the inspection procedure had not been notified to the inspected party, which contradicts the established case law, which the Supreme Court reiterates in this ruling. The case law states that "judicial authorization for entry and search in a constitutionally protected home must be connected to the existence of an inspection procedure already initiated, and whose initiation has been notified to the inspected party, indicating the taxes and periods that the investigation affects, as this requirement derives from Articles 113 and 142 of the General Tax Law".

It seems, therefore, relevant to ensure that the algorithmic systems used by inspection bodies employ relevant risk indicators that allow for obtaining meaningful evidence.

[14] ECLI:ES:TS:2021:3502.

Furthermore, after obtaining this initial evidence, it would be necessary to correctly initiate formalized inspection actions that allow for a more precise determination of the potential infringement and its scope before proceeding with more extensive inspection actions, such as those that notably affect fundamental rights, as in the case described. In short, the reliability and explainability of the results generated by the automated system are crucial if they are used to justify inspections or to support the imposition of sanctions, and it should be possible to guarantee their accuracy and validity.

5 CONCLUSIONS

As previously noted, the AI Act does not appear particularly well-suited to extending additional protective measures in relation to automated control systems, as these systems are difficult to classify within the high-risk categories currently set out in Annex III.

For this reason, it is essential that individual States undertake reforms within their domestic legal frameworks in order to mitigate the challenges arising from the use of automated control mechanisms. This study, based on Spanish regulations and experiences, has highlighted some issues that should be addressed by domestic law, including: the need for information on the automation of administrative activities, particularly where such processes may have adverse effects on individuals (such as leading to a subsequent sanction); the establishment of swift and accessible appeal mechanisms before human beings to correct false positive situations at no cost to the affected party; and the need to ensure that system outputs are reliable enough so as to inspire confidence and be accepted by the judiciary.

6 REFERENCES

Aoki, N. (2020). An experimental study of public trust in AI chatbots in the public sector. Government Information Quarterly, Volume 37, Issue 4, 2020, p. 1-10.

Baader, G. & Krcmar, H. (2018). Reducing false positives in fraud detection: Combining the red flag approach with process mining. International Journal of Accounting Information Systems, vol. 31, 1-16

Brizuela, A. et al. (2024). Public Sector Tech Watch. Mapping Innovation in the EU Public Services. Publications Office of the European Union, Luxemburg, 2024, doi:10.2799/4393.

Brożek, B. et al. (2024)The black box problem revisited. Real and imaginary challenges for automated legal decision making. Artificial Intelligence Law, vol. 32, 427-440. https://doi.org/10.1007/s10506-023-09356-9

Capdeferro Villagrasa, O. (2019). Las herramientas inteligentes anticorrupción: entre la aventura tecnológica y el orden jurídico. Revista General de Derecho Administrativo, num. 50, 1-28.

Capdeferro Villagrasa, O. (2025) (forthcoming). Red flags Systems against corruption. Spanish AI Systems Tools and the Impact of EU AI Act. In Ponce Solé, J. And Cerrillo i Martínez, A. (eds.), The EU Artificial Intelligence Act and the Public Sector. Humans and AI Systems in Public Administration in the light of the European Regulation on Artificial Intelligence of 2024, EPLO.

Carnis, L. (2011) Automated Speed Enforcement: What the French Experience Can Teach Us. Journal of Transportation Safety & Security, 3:1, p.15-26.

Cerrillo i Martínez, A. (2019). "Com obrir les caixes negres de les Administracions públiques? Transparència i rendició de comptes en l'ús dels algoritmes". Revista Catalana de Dret Públic, num. 58, p. 13-28.

Cerrillo i Martínez, A. (2024). "Lección 21. Actuación automatizada, robotizada e inteligente". VVAA. Manual de Derecho administrativo (2 ed.). Marcial Pons, p. 531-547.

Coglianese, C. & Ben Dor, L. M. (2021). AI in Adjudication and Administration, Brooklyn Law Review. Vol. 86:3, 791-838.

Cotino Hueso, L. (2023). Qué concreta transparencia e información de algoritmos e inteligencia artificial es la debida. Revista española de la transparencia, num. 16, p. 17-63.

De la Cueva González-Cotera, J. (2025) (forthcoming). Transparencia y acceso al código fuente de los programas informáticos. Ponce Solé, J. & Villora Mendieta, M. (eds.), Anuario del Buen Gobierno y de la Calidad de la Regulación 2024. Fundación Democracia y Gobierno Local, Madrid.

Gamero Casado, E. (2023). Las garantías de régimen jurídico del sector público y del procedimiento administrativo común frente a la actividad automatizada y la inteligencia artificial. Gamero Casado, E. (dir.). Inteligencia artificial y sector público. Retos, límites y medios. Tirant lo Blanch. Pp. 397-464.

Goerlich Peset, J. M. (2024). Reglamento de inteligencia artificial e intervención pública en las relaciones laborales. Labos, vol. 5, p. 228-242.

Goerlich Peset, J. M. (2021). Decisiones administrativas automatizadas en materia social: algoritmos en la gestión de la Seguridad Social y en el procedimiento sancionador. Labos, vol. 2(2), p. 22-42.

Grimmelikhuijsen, S. and Tangi, L. (2024). What factors influence perceived artificial intelligence adoption by public managers. Publications Office of the European Union, Luxembourg, https://data.europa.eu/doi/10.2760/0179285, JRC138684.

Inspección General de Servicios (2023). Informe de evaluación del plan de la Inspección General de Servicios 2022-2023, Anualidad 2022. IGS-GV.

Jiménez Cardona, Noemí. (2022). La intel·ligència artificial en la detecció de les pràctiques de bid rigging: el paper capdavanter de l'ACCO. Revista Catalana de Dret Públic, num. 65, 129-145. https://doi.org/10.2436/rcdp.i65.2022.3876

Kroll, J. A. et al. (2017). Accountable Algorithms. University of Pennsylvania Law Review, vol. 165, 633-705.

Miranzo Díaz, J. (2023). Inteligencia artificial y Derecho administrativo. Tecnos: Madrid.

Muñoz Vicuña, D. (2025). Modelos de evaluación de impacto algorítmico para las decisiones administrativas automatizadas. Vaquer Caballería, M. (dir.), La actuación administrativa automatitzada: sus claves jurídicas, Tirant lo Blanch, Valencia, 2025, 501-537.

Oliver Cuello, R. (2021). "Big data e inteligencia artificial en la Administración tributaria". Revista IDP, num. 33, p. 1-13.

Ponce Solé, J. (2019). Inteligencia artificial, Derecho administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico. Revista General de Derecho Administrativo, num. 50.

Pontón Aricha, T. (2024). "Aproximación al uso de la inteligencia artificial en la administración tributaria española". Hinojosa, J.J. & Sánchez-Archidona, G. (eds.). Inteligencia artificial, cumplimiento voluntario y los derechos y garantías de los contribuyentes. Atelier. P. 61-80.

Rachovitsa, A. y Johann, N. (2022). The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch SyRI Case. Human Rights Law Review, Vol. 22:2, 1-15.

Sindicatura de Greuges de Barcelona. (2024). Resolució de la Sindicatura de Greuges retrieved from: https://www.sindicaturabarcelona.cat/wp-content/uploads/202405_Dret-bona-administracio_infraccio-a-un-vehicle-que-no-pot-circular-per-la-ZBE.pdf

Snow, A., (2017), Automated Road Traffic Enforcement: Regulation, Governance and Use: A review, RAC Foundation, London.

Valero Torrijos, J. (2019). Las garantías jurídicas de la inteligencia artificial en la actividad administrativa desde la perspectiva de la buena administración. Revista Catalana de Dret Públic, n. 58, pp. 82-96.

van Noordt, C., & Misuraca, G. (2022). Artificial intelligence for the public sector: results of landscaping the use of AI in government across the European Union. Government Information Quarterly, 101714.