

# Algorithmic transparency and bureaucratic discretion: The case of SALER early warning system

J. Ignacio Criado<sup>a,\*</sup>, Julián Valero<sup>b</sup> and Julián Villodre<sup>a</sup>

<sup>a</sup>*Department of Political Science and International Relations, Universidad Autónoma de Madrid, Madrid, Spain*

<sup>b</sup>*Department of Administrative Law, Universidad de Murcia, Murcia, Spain*

**Abstract.** The governance of public sector organizations has been challenged by the growing adoption and use of Artificial Intelligence (AI) systems and algorithms. Algorithmic transparency, conceptualized here using the dimensions of accessibility and explainability, fosters the appraisal of algorithms' footprint in decisions of public agencies, and should include impacts on civil servants' work. However, although discretion will not disappear, AI innovations might have a negative impact on how public employees support their decisions. This article is intended to answer the following research questions: *RQ1. To what extent algorithms affect discretionary power of civil servants to make decisions? RQ2. How algorithmic transparency can impact discretionary power of civil servants?* To do so, we analyze SALER, a case based on a set of algorithms focused on the prevention of irregularities in the Valencian regional administration (GVA), Spain, using a qualitative methodology supported on semi-structured interviews and documentary analysis. Among the results of the study, our empirical work suggests the existence of a series of factors that might be linked to the positive impacts of algorithms on the work and discretionary power of civil servants. Also, we identify different pathways for achieving algorithmic transparency, such as the involvement of civil servants in active development, or auditing processes being recognized by law, among others.

**Keywords:** Algorithms, algorithmic transparency, algorithmic governance, regulation, compliance, decision-making, discretionary power, Spain

## Key points for practitioners:

- SALER is an early warning system based on algorithms to promote transparency and good government.
- The existence of an early warning system based on the utilization of algorithms will prevent corruption.
- Algorithmic transparency is a critical point of public decisions based on artificial intelligence systems, addressing their accessibility and explainability.
- Challenges to civil servants' work autonomy and discretion of decision-making will increase as a result of the use of algorithms in public sector settings.
- Future studies need to leverage the implications of predictive algorithms to understand the full potential on public discretion and work autonomy.

## 1. Introduction

Artificial Intelligence (AI) and algorithms have the potential to transform critical dimensions of public

---

\*Corresponding author: J. Ignacio Criado, Department of Political Science and International Relations, Universidad Autónoma de Madrid, Building of Law, Politics and Economics, 1st floor, office 2, C/Marie Curie, 1, Ciudad Universitaria de Cantoblanco, Madrid, 28049, Spain. Tel.: +34 914972466; E-mail: ignacio.criado@uam.es.

sector organizations and people working in them. Governance based on the utilization of algorithms is at its first stage of development, whereas it is under implementation in different governmental sectors (Janssen & Kuk, 2016). Nonetheless, we have very limited knowledge about the governance opportunities and challenges that implementing public services based on algorithms will encompass for public agencies, people working in them and citizens (Criado & Gil-Garcia, 2019; Van der Voort et al., 2019). At this stage, this field of research has not overcome traditional deterministic (both utopian/positive or dystopian/negative) approaches on technology adoption and use in the public sector. Therefore, this article is intended to provide evidence about algorithmic transparency, regarding the accessibility and explainability of algorithms. Accordingly, we analyze the implications in decision-making processes supported by the use of an AI-based system and the effects in the discretionary power of public employees involved in its implementation.

Recently, different authors have studied the notion of algorithmic transparency. From a philosophical standing point, transparency of algorithms is a key dimension to build an ethical governance, as it would be difficult to state that an opaque government is ethical (Winfried & Jirotko, 2018). Here, the ethical dimension of algorithms is oriented to make organizations responsible for the algorithms they design and implement (Martin, 2018) and uncover the values and assumptions embedded in automated systems (Geiger, 2017). Besides, others have raised the concern about black box problems with algorithms by arguing that the “*lack of openness has meant that users and researchers would often consider the actual behavior of algorithms in thinking about governance*” (Introna, 2015: 12); and that while public agencies could make more efficient decisions, also they could disguise information inside “black boxes”, preventing citizens from the knowledge on the implications they might have in their own lives (Fink, 2017).

Methodologically, this article analyses the case study of SALER, an early warning system implemented at an emerging stage in the government of Valencian region (GVA) in Spain. Our article is intended to answer the following research questions: *RQ1. To what extent algorithms affect discretionary power of civil servants to make decisions? RQ2. How algorithmic transparency can impact discretionary power of civil servants?* Hence, SALER is studied to understand algorithmic transparency, based on the accessibility and explainability of this algorithmic system. At the same time, we study the implications in decision-making processes supported by the use of this AI system and the effects in the discretionary power of public employees (service inspectors), who work fighting against corruption in the GVA public administration. Also, this article identifies practical implications for public sector managers working with AI.

The remainder of the article is as follows. After this introduction, the second section develops a literature review working with different concepts and dimensions, including algorithmic transparency, regulation of algorithms in the public sector (mostly in Spain and the European Union), algorithmic transparency and the black box problem, and algorithms and discretion of public employees’ decisions. Our third section displays the analytical framework, case selection and methods. The fourth section exhibits the results of the analysis including the impacts of SALER algorithms on GVA public managers’ work and the perceptions on SALER’s algorithmic transparency and impact on discretion. Then, the discussion section presents the main findings and practical implications of this paper, before opening up the conclusion.

## 2. Literature review

AI in the public sector encompasses different aspects, including governance and regulative issues. This section will focus on reviewing the most important aspects about the use of AI technologies in public organizations. On the one hand, we address the notion of algorithmic governance and regulation of

algorithms. On the other hand, we present the black box problem, the transparency of algorithms and the actual debate on AI and autonomy and decision-making of employees in public sector organizations.

### 2.1. Artificial intelligence and algorithmic governance

Implications of algorithmic governance in the public sector come up hand in hand with expected situations derived from the utilization of autonomous systems to make decisions. The implications of algorithms and AI systems is growing in different areas of the public sector. Wirtz et al. (2018) distinguish at least ten areas of AI and algorithm-mediated governance with potential impact in the future of the public sector (i.e. AI Process Automation Systems, Virtual Agents, Predictive Analytics and Data Visualization, Recommendation Systems, or Speech Analytics). All of them are fostering disruptive consequences in different industries and sectors and it is supposed to occur in the same way in the public sector (Agarwal, 2018).

At the same time, the notions of AI, algorithm and algorithmic decision-making are not completely shared in the literature. Here, following the High-Level Expert Group on AI of the European Commission we broadly define AI as “*systems that display intelligent behaviour by analysing their environment and taking actions – which some degree of autonomy – to achieve specific goals*” (European Commission, 2019: 1), and even more precisely “*AI systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions*” (European Commission, 2019: 6). Accordingly, the idea of algorithm is very closely related to AI as it consists of the step-by-step processes and/or rules processing inputs into outputs in AI-based systems (Jansen & Kuk, 2016). Therefore, one of the key issues of AI and algorithms since the emergence of computing is the reasoning/information processing dimension.

Also, we present the idea of decision-making based on AI-systems and algorithms, as it is one of the critical issues regarding organizational (public administration) studies. Here, we argue that the term “decision” should be considered broadly regarding AI systems, including “*any act of selecting the action to take, and does not necessarily mean that AI systems are completely autonomous. A decision can also be a selection of a recommendation to be provided to a human being, who will be the final decision maker*” (European Commission, 2019: 3). At the same time, decision-making based on AI-systems implies that once the action has been decided, the AI system might be ready to perform it or leave that actuation to humans (Harris & Davenport, 2005). Therefore, decision-making based on AI-systems may, or may not, convey final actuation of humans.

Algorithmic governance encompasses deep transformation in the political and public administration realm. From the political side, some claim that AI and algorithms will reshape the global order (Wright, 2018) or change interactions among countries (Lee, 2018). From the perspective of public organizations, AI and algorithms may transform processes (Brynjolfsson & McAfee, 2017), be used in the analysis of citizens behavior to understand their needs and deliver public services better suited to those needs (Vogl et al., 2020), verification of eligibility for public services and detection of fraud, or better operational performance of service providers (Margetts & Dorobantu, 2019). At the same time, different authors have recently identified several cases in the public sector implementing AI and algorithm-mediated governance coming from different public service sectors (Desouza, 2018; Sun & Madaglia, 2018) (i.e. health, education, social services, migrations, customer care, tax, social security, police, emergencies,

transportation, water management, waste recycling, air control, among others). Besides, it is likely that these service areas may be improved by AI and algorithms regarding the different stages of the policy-cycle process (Pencheva et al., 2020; Valle-Cruz et al., 2020).

At the organizational and service level, governance of algorithms and decisions based on them in public agencies raise different concerns, including data quality, social and economic biases, and opacity. First, the quality of data is a key factor to avoid noise, heterogeneity, etc. Using algorithms that are nurtured by data from different organizational areas requires that quality is controlled. Second, different potential biases derive from the design and training of data, in particular, regarding to racial, gender or economic features (Cath, 2018). Hence, public organizations could produce negative outcomes using social and economic profiling based on previous biased administrative systems and data records. Also, opacity is a potential hidden burden in public agencies using AI as a result of the black-box effect and potential lack of accountability over decision-making. Here, the limits of transparency during the design and implementation of algorithms might erode public governance and reduce control and even trust in public institutions. Also, there is another key point having an appropriate understanding of regulation of algorithms in their context of application.

## *2.2. Regulating algorithms in the public sector*

The current doctrinal discussion on the legal aspects related to the use of the AI in the public sector has mainly focused on the analysis of the requirements, conditions and guarantees arising from the use of this technology as regards, on the one hand, the guarantee of the citizens' rights and, on the other hand, the adequate protection of the general interest. From the latter perspective, the fact that there is a close relationship between the use of AI in the public sector and the principles of good administration in dealing with cases of corruption has been highlighted (Ponce, 2018), especially in the area of public procurement (Aarvick, 2019). The link with such principles is an essential requirement when facing the adaptation of the legal framework to this emerging technology, particularly if we take into account the singularity of the legal risks and challenges that emerge from the technological innovation approach (Valero, 2019).

Indeed, one of the areas of major concern has been the need to adapt the guarantees provided by the general regulation on transparency and access to public sector information to the algorithms (Cerrillo, 2018), in the light of the fact that the regulatory framework is rather outdated in terms of the IA's requirements. Therefore, the proposals based on the preventive application of general principles in the design of the applications are particularly thought-provoking (Loi et al., 2019). One of the main topics of discussion concerns the transparency of the algorithms and, in particular, when private entities contracted by public administrations participate in their design. In these cases, public procurement legislation may involve restrictions on transparency that should be proactively addressed to ensure adequate control (Capdeferro, 2020: 9–10).

On the other hand, the requirements of the principle of good government are projected onto the discretion of administrative decisions, especially with regard to their motivation when algorithms are being used. In the case of the Spanish legislation, there is no specific legal obligation in this regard despite the fact that the regulatory framework is certainly recent (2015). However, the need to reserve the adoption of discretionary decisions to human decision has been discussed (Ponce, 2018), particularly given the need to weigh up certain aspects that are difficult to assess adequately by a computer system. In this context, the discussion about whether or not the decision-makers have the discretion not to use the result of computer applications based on the use of AI or big data seems to be particularly relevant, insofar as other values beyond mere technology have to be considered (van der Voor et al., 2019; Ponce, 2019).

### 2.3. Algorithmic transparency and the black box problem

Transparency is one of the challenges for public agencies operating algorithms to deliver public services. Algorithms are a set of rules or procedures to solve a problem or to find a solution for an administrative procedure. Then, transparency of algorithms (or algorithmic transparency) is a key factor for the understanding of their results and potential problems derived from the decisions made using autonomous systems. To put it differently, algorithmic transparency, defined here as accessibility and explainability of public decisions made by algorithms (Grimmelikhuijsen, 2019), needs to be addressed in order to appraise the impact of algorithms in public agencies and divert black box concerns in public administrations during the time of open government.

Mostly, the black box problem of algorithms has much in common with two dimensions of transparency: accessibility and explainability. Here, Grimmelikhuijsen (2019) or Grimmelikhuijsen and Meijer (2014) argue that, in the first case (accessibility), transparency is accomplished by the availability of information in different dimensions of the policy process or managerial functions of public agencies. Making the information available and observable by external users is essential to assure the accessibility of government processes (Lepri et al., 2018). However, accessibility in the case of algorithms also means that external audit agents or agencies could access algorithms assessing their compliance with ethical, legal, and governance standards, regulations and procedures.

The second dimension of transparency (explainability) of algorithms have much in common with the capacity to explain what they do and how a decision has been made in a determined context. As an example, the European Parliament (2019: 2) states that “*explicability should also be available when, on the basis of profiling, citizens are the object of a set of micro-decisions that individually are not very important but which, on the whole, may have a substantial impact on them*”. Burrell (2016) insists on the complexity of AI systems (i.e. machine learning) making it difficult to decode the learning capabilities of algorithms or to explain “*how an algorithm rendered a certain outcome in an individual case*” (Grimmelikhuijsen, 2019: 8). This idea is close to the executability of the code or the assumption that AI systems based on algorithms can act automatically (Introna, 2016). Here, explainability has much more in common with the capacity to monitor the behavior of algorithms during their operation and ensuring human intervention afterwards (according with the General Data Protection Regulation, articles 13–15).

### 2.4. Algorithms and discretion of public employees’ decisions

Decision-making based on algorithms and discretion of public managers and employees is the other dimension that we explore in this article. Up to now, there is limited empirical research about decisions of bureaucrats based on algorithms (Meijer & Wessels, 2019). AI and algorithms in public agencies seem to evoke the Weberian iron cage, now transformed into a digital fortress, where rules are not readily understood or even available for study (algorithmic opacity) and automated decisions might lead to unintended effects (algorithmic biases) (Young et al., 2019). Besides, it is also expected that algorithms will change the way knowledge-based public employees work and take decisions, so that all aspects of individuals’ performance are quantified, compared to others, and managed against algorithm models, that will carry on important changes requiring deeper understanding (Orlikowski & Scott, 2016). Then, we focus on decisions where AI may either augment or replace human discretion. Following Young et al. (2019), we argue that the relationship of AI to discretion is unique because of three AI’ design features: (a) *it is built for automating learning and decision-making processes through abstract mathematical representation of problems*; (b) *it can utilize input data with speed and dimensionality that vastly outstrip*

*human cognition; and (c) as more data become available, it can “learn” and adjust its behavior by updating its decision heuristics’* (Young et al., 2019: 2).

Therefore, discretion in decision-making is one of the emerging issues regarding algorithms and public managers and employees. At first sight, algorithms can copy the ways in which tacit knowledge is acquired by public employees in knowledge-based organizations, including public administrations, and this may affect personal discretion of the worker to take his/her own decisions (Faraj et al., 2018). Discretion of street-level bureaucrats is one of the traditional topics in public administration (Lipsky, 1971), and ICTs developed the idea of an emerging new system-level bureaucracy (Bovens & Zouridis, 2002). Challenges to discretion go bottom-up affecting street-level bureaucrats (i.e. social workers, doctors, judges, etc.) and public managers (from middle to top), in all cases, working in intensive knowledge-based organizations and units surrounded by algorithms (Barth & Arnold, 1999). Then, decisions based on AI systems seem to encompass rules outperforming traditional limitations of human bounded rationality (Simon, 1991). However, this is not yet clear in all policy and service domains.

Consequently, decision-making based on the use of algorithms might impact the capacity of public employees not only to organize their jobs and autonomy, but also to get fair results or to be accountable for the decisions made by those autonomous systems. At this stage, the study of implications for public officials’ discretion of algorithmic decision-making is critical to discern to what extent their work environment will be challenged by the generalization of algorithms and the potential side effects of their use (Meijer & Wessels, 2019). At the same time, studying algorithmic decision-making in the public sector is important to assess the interaction of public employees with AI systems and whether they understand how algorithms operate and the consequences of their opacity/openness. Also, this might suggest that we are facing the advent of an algorithmic bureaucracy (Vogl et al., 2020), linked to the imbrication of algorithms with traditional forms of public sector organizations and the socio-technical relationships between workers and their work tools.

### 3. Analytical and methodological design

In this section, we present the case selection and methods approach of our article. First, we introduce SALER case, summarizing its origins and evolution to date, and the technological, organizational, and legal dimensions in the implementation of this project in the *Generalitat Valenciana* (GVA). This is a Spanish regional government in the Mediterranean Sea area, with more than 150,000 public employees, 22,000.000 k € annual budget, for a total population of about 5 million (4<sup>th</sup> out of 17<sup>th</sup>, in Spain) (National Statistics Institute, 2019). Also, we explore actors (human dimension) involved in the development and use of SALER. Here, we are following other studies suggesting the potential of AI systems to explain and predict corruption (Melo Lima & Delen, 2020). On the other hand, we introduce semi-structured interviews as principal research technique used in this article, as well as the operationalization of concepts on which these interviews are based on. Also, a documentary analysis was developed using official documents, laws and regulations, framing the implementation of the system. Here, the study operates to understanding algorithmic transparency and the implications of SALER system in decision-making of public employees.

This study is based on two research questions that come up from our literature review: *RQ1. To what extent algorithms affect discretionary power of civil servants to make decisions? RQ2. How algorithmic transparency can impact discretionary power of civil servants?* From these research questions, we expect to develop significant findings that will improve our understanding of the impacts of algorithms on civil servants’ daily work and their discretionary power to make decisions over their activities. In the same line, we will explore how more transparent algorithms could positively affect civil servants work.

### 3.1. SALER. Algorithms to prevent irregularities and promote good government

SALER emerged from the opportunity to take advantage of big data and AI systems to fight fraud and corruption in public administrations. Cases involving corruption practices hit several Spanish regional and local public administrations in the past decade, and particularly in the Valencian region. This required public officials to take measures to prevent these events from happening again (Puncel, 2019), as a new government took office in 2015. This environment included the attention to fiscal consolidation policies and the growing social preoccupation against governmental corruption (Parrado et al., 2018; Ruvalcaba-Gomez et al., 2017; Villoria et al., 2014). Therefore, one could say that in the mid 2010s, corruption arose as a key issue in the Spanish political landscape.

The first change in office at the government of GVA (*Generalitat Valenciana*, governing body of the Valencian region), after 20 years under the same ruling party, took place in 2015 elections. This situation opened a policy window for the improvement of several internal management control mechanisms. Among others aspects, it was essential to provide the *Services Inspectorate* (whose functions are detailed below) with new data and technological tools to improve the instruments that were in operation, and that did not allow for easy follow-up contracts or grants procedures (which are the most likely to lead to bad practices) (Puncel, 2019). Thus, SALER system (named from its Spanish acronym “*Sistema de Alertas*”, in English “Early Warning System”), also known as “SATAN” by Spanish news media, was developed and implemented in the GVA as a new technological framework for the *Services Inspectorate*. The original purpose of SALER was to enable service inspector’s data analysis on contracts, subsidies, grants, among others, in order to detect, prevent and even anticipate, conflicts of interest and corruption practices by using its potential machine-learning and descriptive-analysis capabilities (Martínez et al., 2018). Figure 1 shows a global scheme of SALER system, whose dimensions are explained below. Next paragraphs describe the technological, organizational, human, and legal dimensions of this system.

#### 3.1.1. Technological dimension

One of the key components of SALER is based on the logical and computerized system. The first prototype was developed in 2017–2018 in collaboration with the *Universidad Politécnica de Valencia* (Polytechnic University of Valencia), taking into account other initiatives such as the ARACHNE program (European Commission), or the zIndex (Czech Republic) (Martínez et al., 2018). The logical system is based on a set of algorithms that act as independent processes. Each of them is responsible for answering different questions. The algorithms were designed based on natural language questions, extracted from the knowledge and needs of the GVA service inspectors (*Services Inspectorate*).

The algorithm is a process that handles data collection, modelling and visualization. Data collection occurs through different types of databases: a) internal data from the GVA, based on own data about public contracts, grants, subsidies, etc.; and b) external data from public sources, including the commercial register or public notaries. The algorithm crosses and merges these data, facilitating descriptive statistical operations on them, graphic mapping using Social Network Analysis for individuals/organizations, as well as risk metrics that evaluate the case according to whether it has a low, medium or high risk regarding to possible irregularities. At the end of 2018, the system was used within the organization, still in a testing phase, and its activities were limited to the following set of risk areas: contracts, grants, subsidies, allowances, conflicts of interest, human resources, and permits.

#### 3.1.2. Organizational dimension

SALER has been applied in the GVA, as well as other public agencies linked to this administrative body, including public enterprises, autonomous bodies, and other semi-public agencies. Internally, the *Services*

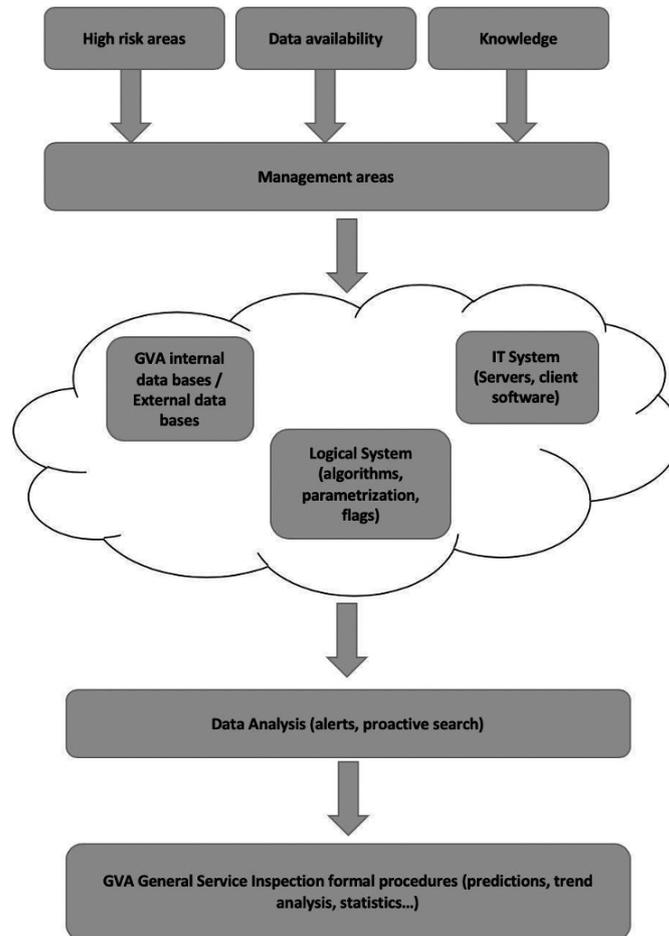


Fig. 1. The SALER early warning system as described in the SALER deployment and consolidation plan (2019–2023).<sup>1</sup>

*Inspectorate* is the core user of the system. This is an administrative body leading the internal control and oversight of GVA departments and units, namely responsible for monitoring regulatory enforcement and services compliance (Giménez et al., 2018). By using the SALER toolkit, the *Services Inspectorate* leads defining and establishing specifications and indicators that size up alerts, and the verification of these alerts in order to prevent, evaluate, and anticipate risks. On the other hand, the IT department within the GVA (known as DGTIC), also offers computer support to service inspectors. At the same time, this technological unit has been endorsed with the responsibility of sustaining the future development of SALER within the GVA.

At the organizational level, it seems especially important the role of risk evaluation maps and individual self-evaluations plans of SALER. Risk evaluation maps “are tools aimed at identifying the activities or processes subject to risk, quantify the probability of these events and measure the potential damage associated with their occurrence” (Giménez et al., 2018: 8). These tools allow the discovery of new risk factors, some of which could be added into the logical algorithmic system. Individual self-evaluations

<sup>1</sup>Internal GVA document provided by interviewees.

plans fulfilling a similar preventive function within some specific departments of the organization that report high risks. Particularly, this aspect is key within the organization giving each unit and department a potential assessment tool that needs to be validated by the *Services Inspectorate*, fostering future transformations in the organizational dimension of this regional government.

### 3.1.3. Legal dimension

Another important aspect of SALER is the legal and regulatory support to this system. The general Spanish regulatory framework for administrative action does not require for an obligation to formally authorize applications used for the exercise of administrative powers (Valero, 2018). However, in the case of SALER it was considered convenient to proceed with a legal regulation that would rule the inspection procedures in the GVA, a decision that gave rise to Law 22/2018, of 6 November. Although this decision was not the result of a predefined strategy, the adoption of a law that specifically envisaged the tool was of enormous importance in the further development of the project. Among other reasons, this is especially remarkable when it came to imposing collaboration obligations (art. 24) or ensuring greater access to data, specifically when it referred to individuals.

In fact, beyond the general impact on public opinion and the administrative organization itself at internal level, the fact that it is a rule with legal status has allowed the difficulties and reticence with regard to the protection of personal data to be resolved. Therefore, given this circumstance, the need to adopt, as a preventive measure, those measures that will ensure the regulatory compliance of the tool acquires special importance. In particular with regard to the fundamental rights potentially affected (Martínez, 2019), which on occasions may require the approval of a formal Law for the purposes of the provisions of article 6.1.e) GDPR. In this respect, specific regulations have been established on this particular aspect (third additional provision), thus complying with the requirements imposed by the Spanish constitutional system when it comes to affecting fundamental rights.

Finally, and as far as the object of this work is concerned, even though the legal norm could have been used to impose the use of SALER for those cases in which the service inspection activity was discretionary, the fact is that its usage is conceived as a simple support tool that generates preventive alerts (art. 17). However, the Spanish legislation that generally regulates administrative action requires that discretionary decisions be justified. Taking into account the Spanish general regulation, it has been proposed (Capdeferro, 2020) to consider that it is an obligation when the action is discretionary. It could be said, therefore, that in these cases the inspection should justify its action against the result initially offered by the algorithm, especially if it decides to act in a manner different from the criterion suggested by the algorithm.

### 3.1.4. Human dimension

Different actors and groups interact with the SALER system inside the regional government. The main actor and user of the system are *service inspectors*. This is a multidisciplinary group of civil servants whose main duties (as stated by law) can be summarized in the following: a) prevention and analysis of risks to avoid irregularities and bad practices in public management; b) evaluation, control and analysis of the effectiveness and efficiency in public management; c) investigation of irregularities and infractions, proposing the adoption of any necessary measures to correct them; and d) conducting evaluations and studies on the organization in relation to different aspects, such as the administrative structure or concrete procedures. They work under the organizational umbrella of the *Services Inspectorate*.

Regarding to the particular duties and tasks that we have specified, service inspectors represent an unusual type of civil servant. Unlike the majority of studies on street-level bureaucracy and discretion,

with civil servants directly contacting with citizens and being accountable to them (Tummers & Bekkers, 2014; Lipsky, 1971), service inspectors are fundamentally performing internal duties across organizational operations, mostly the supervision of contracts, grants, etc. Then, their direct connection with citizen is limited. In fact, their consideration as street-level bureaucrats is complicated by not finding that direct link. Service Inspectors provide internal services regarding to other co-workers of the organization. Although they do not implement policies nor services to external users, they are in charge of preventing corruption, conflicts of interest or fraud in public funds assignment. Therefore, since we focus on detecting whether a certain set of algorithms influences the autonomy of their work, our study did not consider this particularity as a part of the design.

Furthermore, it is important to highlight other actors that have interacted in one way or another with SALER. On the one hand, political leaders who have been policy entrepreneurs supporting the development of the tool. On the other hand, computer engineers, both internal (DGTIC) and external (*Universidad Politécnica de Valencia* and private companies), who have been in charge of the design, development and implementation of SALER. Also, GVA public managers have acted as another key group of individuals who are directly involved in the implementation of this alert system. Finally, here we might include private contracting companies and individual citizens who may be affected, in one way or another, by the results of the use of SALER. However, this final group is out of our scope within this article, as we focus on the internal functioning and impact of this system.

### 3.2. *Analytical framework and methods*

We have selected semi-structured interviews to gather data and information about this case study. Relying on a case study research strategy in organizational settings (Eisenhardt, 1989; Yin, 2004), using interviews helped us to develop a comprehensive assessment of the story behind actors' experience with SALER, pursuing in-depth information around the case (Creswell, 2009). Therefore, we have used the semi-structured interview for the following reasons: a) our research design is based on a case study; b) it is an emerging phenomenon, for which we have little information and empirical data; c) we seek to maintain an exploratory logic to foster explicability based on the case study analysis, that is, to understand the experience and obtain more information about the case and the studied concepts.

The interviews were carried out as follows. First, the interviewees selection procedure was based on the chain sampling technique (Guest et al., 2006): initial interviewees recommended other relevant people inside the organization, generating a snowball effect. This technique was appropriate since we initially had few reference contacts within the organization. In total, 6 interviews were carried out (although a few more were scheduled, they could not take place due to the COVID19 crisis). The interviews targeted 3 different types of actors: 1 politician, 2 developers, and 3 service inspectors. Although the number of interviews seems low, the interviews targeted the most relevant people regarding to the development, implementation and usage of the system: the key political champion of SALER the service inspectors who use the system more often, and the IT developers who created the initial prototype. This particular selection criteria have allowed us to get different perceptions, from technical (developers), to political (policy entrepreneur), and non-technical end-users (service inspectors). All interviewees have been active participants during the process of design, adoption, implementation and use of SALER. In the case of service inspectors, they worked in the GVA and the *Service Inspectorate* even before the system was launched and in some cases, they have been working within this public administration for more than 20 years, keeping organizational memory of changes.

The interviews were recorded in Spanish and manually transcribed and coded by the authors. Furthermore, some interviewees provided us with additional official documentation that has also helped us to

Table 1  
Operationalization of conceptual and analytical aspects

Concept	Definition	Operationalization
SALER	Application framework that comprises a toolkit for analyzing data on contracts, subsidies or conflicts of interest, among others, with the aim of detecting, accompanying, preventing and even anticipating the consequences of bad practices in public administrations (Puncel, 2019; Martínez et al., 2018)	Using the main components of the SALER system: a) the <i>logical cross-data processing algorithms</i> ; b) the <i>risk evaluation maps</i> and the <i>individual self-evaluation plans</i> (Martínez et al., 2018; Giménez et al., 2018)
Algorithmic transparency	A design and implementation component by which the decision-making process and logic result of algorithms should be stated clear, visible and comprehensible in order to prevent information manipulation, power-asymmetry and discrimination (Lepri et al., 2017; Diakopoulos & Koliska, 2016)	Using two dimensions of algorithmic transparency (Grimmelikhuijsen, 2019): a) <i>accessibility</i> , that is, the source code can be audited by external experts; and b) <i>explainability</i> , that is, decision-making results are humanly understandable
Discretionary power	The freedom perceived by public employees in relation to the possibility they have to make decisions about different areas of their daily work (Tummers & Bekkers, 2014; Evans, 2010)	<i>How “free” the civil servant perceives (s)he is to perform its duties and formal tasks</i>

Source: Own elaboration.

understand the case. An interview protocol listing the different questions asked to our respondents can be found at the end of the article (appendix). Each interview was manually analyzed (we did not use specialized software as they were few interviews) using the text processor “comment” function to highlight and code relevant information regarding the dimensions and operationalization. These dimensions and the operationalization of conceptual aspects are summarized in Table 1.

Our subsequent manual coding for the interview questions has been designed around the scheme presented in Table 1. Here, we denote the conceptual elements that we have identified regarding to SALER System, including the notions of algorithmic transparency and decision-making discretion. Each dimension of the interview questions/topic lists was intended to identify and link some of the operationalization indicators as specified by our research questions. Thus, our dimension 1 asked for general questions, identifying the interviewee based on their position inside the organization and his/her relationship with SALER system. Our dimension 2 was intended to gauge the impacts of SALER on the daily work of the *Services Inspectorate* unit requesting their components and making connections with the functions and perceived discretion of these civil servants to make their own decisions. Our dimension 3 was linked to perceptions about SALER’s level of algorithmic transparency and how they impact on discretionary power. This way of organizing the coding process is connected to the interview protocol (see Appendix). Also, we have intended to somehow reproduce this workflow in the results section.

#### 4. Results

In this section we present the results of this study. They are in line with our exploratory qualitative research design assessing a set of algorithms that have not yet fully deployed their potential. In the first subsection, we describe the impacts of SALER components on Generalitat Valenciana (GVA) service inspectors daily work and perceived autonomy to perform their duties within the *Services Inspectorate* unit. In the second section, we depict SALER’s perceived levels of algorithmic transparency, and how they impact the perceived autonomy of civil servants about their own decisions and jobs.

#### 4.1. Impacts of SALER algorithms on public managers' work

All the interviewees agreed that one of the basic characteristics of SALER algorithms is their “decision support” nature. This conclusion implies that this system helps GVA service inspectors to use and analyze big data, but that the algorithm results should never be considered as the final decision to be made by the services inspectors. Hence, GVA service inspectors should maintain their supervisory functions. Therefore, they should review the results of the algorithm, considering them as a supplementary indicator to inform their final decisions, and not as the final result of the process itself. Thus, service inspectors have an ample sphere of freedom to take into consideration the indicators displayed by the algorithm, to interpret the findings and to take actions accordingly. To put it differently, SALER system is not making direct decisions, at this stage, it is a starting point to develop further steps into a decision. As exposed by one of the interviewees:

*“In no case, an information that we obtain from SALER will be conclusive. This information will be a starting point for us to carry out an investigation. I will give you an example: if we obtain information of SALER that tells us that a certain department has had a percentage of 50% of its contracts with a final price lower than the bid price. Well, based on this information, that 50% seems to be an important number. But we are not going to directly say they are acting reckless. No, what we will do is go in to analyze, see the detail of what those contracts are, see how the algorithm works, and see what type of contract and what kind of procedure is ...”*

(Interview 6)

Since the beginning, the fact that SALER is based on a set of algorithms oriented to support decision-making has promoted a general opinion that this system has not deeply impacted service inspectors' ability to work autonomously. GVA service inspectors feel that SALER has provided (and will provide in the future) a new direction for the *Services Inspectorate* unit, and the possibility of carrying out its functions with greater efficiency and efficacy. At the same time, SALER is regarded as an additional source of data and information about detect and verify symptoms of fraud and corruption. In all cases, the involved actors highlight the human dimension behind the decisions made by the system:

*“This is one more way of entry, one more way of verification. So, we do not care if this information comes from the algorithmic system, or if it has been reported by a public official, or a union, or that we have discovered it by doing an ordinary investigation. Because we always do not accept the alert until we do a check ...”*

(Interview 5)

Correspondingly, in the case of other GVA service managers, the impact on their discretion does seem to be perceived negatively. Different interviewees commented that other departments within the GVA have seen SALER's logical and algorithmic system as a potential threat to their daily work, as some sort of “watchdog” or “big brother”. Despite this fact, service inspectors have strongly stated that SALER algorithmic functions are not related to intrusive monitoring and accusation, but accompaniment, prevention, and help to avoid risks in other departments. One of the interviewers put it as: *“initially, part of the civil servants, including the unions, perceived this system as “I am going to persecute the official who is doing it wrong. Until we explained that this was not intended, but rather an accompaniment.”* (Interview 1).

One of the most interesting parts of SALER is the risk evaluation maps and self-evaluation plans. These tools are developed with a double perspective. On the one hand, there are risks that have to do with detecting irregularities. On the other hand, there are others based on prioritization factors that have to do

with management efficiency. The construction of these maps is covered by the ISO 31000 standard. Again, this tool has raised some concerns among GVA public managers, although the *Services Inspectorate* unit has tried to explain the role of these tools for support the decisions made by their people (humans) at a later stage of the algorithm operation:

*“Because many people believe that you are going to be controlling everything to persecute them. So, of course, people say: hey, let’s see, if I am managing a contracting procedure, are you going to be chasing me with SALER? And we say: no, the other way around. With the risk map we are going to make it easier for you to have the quite position that you are not going to see any irregularity.”*

(Interview 5)

In addition to make public managers active participants of risks detection, service inspectors are trying to establish a self-evaluation mechanism, allowing each department of the organization to report its own risks and prioritize efficiency. To put it in other words, the purpose is once again to improve the understanding of the system as a whole and as a complementary tool. This might help other public managers to prevent the risks in their own units/departments before letting unethical practices to take place. Besides, SALER will be used to define possible lines of action to mitigate those risks.

Finally, a fundamental element for the protection of autonomy seems to be directly linked to the existence of legal support. Regarding the contents of the regional Law 22/2018, which we have already commented on previous sections, the interviewees stated that it was important for the SALER project, as it provided legal security, both internally and externally, by protecting their functions and appropriately establishing technical and functional mechanisms for SALER. Therefore, it is a law that has a very programmatically and explanatory nature, in addition to its technical-regulatory elements. As one of the interviewees commented:

*“And then we passed Law 22/2018, which gives legal support to both the Service Inspection itself and to third parties. In fact, I believe that the sanctioned part is more directed to those who use the system rather than to third parties. And we realized that it was important to have a law, to know to whom the information was passed, who could use this information, how the information could be used, what the algorithms are, how the algorithms report results . . .”*

(Interview 1)

#### 4.2. Perceptions on SALER’s algorithmic transparency and impact on discretion

This subsection presents the results of our study regarding to algorithmic transparency and discretion over decisions of public managers in the *Services Inspectorate*. Here, we have operationalized algorithmic transparency based on two dimensions (Grimmelikhuijsen, 2019): *accessibility*, that is, source code of the algorithms can be audited by external agents; and *explainability*, that is, algorithms results are humanly comprehensible. Regarding to the first of these dimensions (*accessibility*), GVA service inspectors recognize that SALER algorithms can be audited in a simple way, both by independent agents and by actors within the organization. One of these positive impacts over the service inspectors, and other GVA public managers, is based on the fact that the audits are formally recognized in Law 22/2018 itself, so that a bi-annual report has to be carried out. This adds security to both groups, service inspectors, and other civil servants within this regional administration, as the functionalities of the system could be reviewed. As indicated by one of the interviewees:

*“In the law itself, it is indicated that a report on system’s audit has to be made every two years. And we can count on external help. It is not exactly defined, but we had even thought to collaborate with*

*the anti-fraud agency, or with some other administrative bodies that have knowledge on these types of questions. To have their opinion on how the algorithms are being developed and how the system is used.”*

(Interview 4)

Unfortunately, at the time of writing this article no audit had been performed yet, because SALER has been recently implemented (less than two years). The greater or lesser difficulty to carry out these audits does not seem to depend on the code. Here, as SALER is based on simple algorithms, one might expect that it will not be difficult to audit them since they are based on understandable and easily explainable rules. In fact, *“It would be quite easy to audit the code. Because, as I told you, the code was not very complex. The interface part, the graphics . . . were more complex than the warning rules. [. . . ] In the end, the tool it is just a rule application. So, it is pretty easy to audit”* (Interview 2).

Finally, one of the audit concerns was the potential existence of biases in the code of this software. This fact could act as a barrier during the implementation of the algorithm, in general, and the audit process, in particular. However, the developers of SALER system highlighted that biases primarily could appear in the data, and not in the algorithm itself. Particularly, different interviewees (Interview 3) suggested that frequently data includes biases, and then you have to study it, as biases usually appear in terms of social groups, most affected groups, least affected groups . . . Then, algorithms are not biased, all that they do is transferring biases engendered in the data manipulated by human.

In relation to the other dimension (*explainability*) of SALER, interviewees shared the unanimous opinion that this tool produces humanly comprehensible responses. This is a critical aspect of the system for several reasons. The first, and probably one of the most important, is the fact that GVA service inspectors actively participated by generating the “questions” embedded in SALER’s logic system. So this algorithm really responds to their work needs and they can easily handle it and better understand the results:

*“The questions we are asking the system are humanly understandable, because it is a translation into a programming language of a question like: “look for contracts made by a person in a specific period of time.” It is not applying strange formulas.”*

(Interview 4)

On the other hand, the nature of the algorithm itself makes it easier to understand the results and how they have been produced. SALER relies on descriptive models and it based on evidence from data. Therefore, SALER’s operations are properly documented, and there are no “black boxes” in the data crossing and analysis processes. This contrasts with other types of algorithmic models: predictive algorithms, which can be divided into intelligible models (you can trace the prediction) on unintelligible models (it is difficult to trace the prediction as algorithms learn in the process). As explained by one of the developers:

*“They are simple algorithms for data analysis. In Artificial Intelligence we talk about “black boxes” or models with very low comprehensibility or intelligibility when using predictive models like neural networks: that would be non-intelligible models. And then we would also have intelligible predictive models, such as decision trees, etc., in which you can trace the reason for a decision. Our case is not like this, we do not use predictive models. They are descriptive models that can always be explained based on evidence.”*

(Interview 3)

Finally, SALER *explainability* is also achieved when the results of the algorithms are displayed responding to specific knowledge about each question. Thus, all algorithms’ questions are related to

administrative processes. And, therefore, civil servants who are regularly working with contracts, grants, subsidies . . . will be able to interpret results and acting accordingly. However, if algorithmic complexity increases (for example, with more complicated questions), more expert knowledge on administrative processes will be required, which might imply that certain actors will no longer be able to understand the results:

*“At the level we are now, probably yes. Because we are still asking very simple questions, due to the limited availability of data that we have. In the future, it will probably be more complicated. Because certain conclusions will require expert knowledge, not in computer matters, but knowledge of the processes and the rules that support these procedures . . . The algorithm is obviously essential, but if you do not know the (administrative) process well and the regulations that support these processes, you may think that the algorithm works in one way when it really works the other way round.”*  
(Interview 6)

## 5. Discussion

Based on the results, this section now describes the main contributions of the paper, looking for connections with our theoretical approach, and discussing the analytical framework of the study. First, we look at how our results answer the research questions and what theoretical insights can be set up from our analysis (findings). Secondly, we discuss practical implications regarding SALER’s technological, organizational, legal and human dimensions. All in all, this section is the result of the debate with the literature that we initially presented in the first part of the article. Also, we intend to understand how our results might impact public organizations from the side of public managers involved in algorithmic governance implementation.

### 5.1. Findings

SALER algorithms are having (and probably will continue to have in the future) noticeable impacts on the work autonomy and decisions of civil servants who are actually using them. The use of SALER in the Generalitat Valenciana (GVA) regional government is still at an emerging stage, and its potential as an AI tool has not yet been fully developed. However, this case study facilitates our understanding about the impacts of this kind of algorithmic governance systems on civil servants’ work freedom to perform their duties, and discretion to make their decisions.

#### 5.1.1. Factors for positive effect of SALER algorithms on discretionary power

Our interviews support the existence of a series of factors that might be linked to the positive impacts of algorithms on the work and discretionary power of civil servants. Table 2 lists some of these factors in relation to the impacts of the two basic components of the SALER system. On the one hand, SALER’s logic system seem to have a positive impact in service inspectors work due to *being initially defined as a decision support system* rather than as a decision-making system. This means that the results of the algorithms are not conclusive. Inspection employees review the results and are completely free to decide whether to take them into account and how it helps them making decisions. SALER system does not make a direct decision about a specific procedure. At this state, this approach makes civil servants perceive that their decision-making capacities remain unaltered, which will probably reduce side or non-desired effects (Young et al., 2019). In the long term it is expected that results of algorithms might gradually be taken into

Table 2  
Factors for positive effect of SALER algorithms on discretionary power

Dimension	Factors for a positive effect
Logic system	<ul style="list-style-type: none"> <li>– Implemented as supplementary tool by definition (system to support decisions made by services inspectors)</li> <li>– Functions limited and regulated by law</li> </ul>
Risk evaluation maps	<ul style="list-style-type: none"> <li>– Supported by ISO international standards</li> <li>– Involve other public managers and civil servants with self-evaluations of their services and units</li> </ul>

Source: Own elaboration.

account in decisions as SALER system become more widespread and its functions more comprehensive, influencing the knowledge acquisition process and approach of public managers (Faraj et al., 2018), and even changing the nature of their relationship with work tools and organizational tasks (Vogl et al., 2020). On the other hand, the fact that the *functions and limits of these algorithms are clearly defined by law* provides regulation certainty to the members of the *Services Inspectorate* unit, and also the rest of public managers of the organization. In this case, law enforcement ensures that algorithms integration is developed according to the needs of service inspectors and the GVA (and citizens). Therefore, more than a generic product, it is a system specifically designed for the needs of this organization, addressing public values and good governance objectives.

Regarding to the second component of SALER, the risk evaluation maps and individual self-evaluation plans, some lessons can be drawn about their positive implications. These components have been identified as critical in detecting risk areas and integrating them within the logical elements of the warning system. To strengthen these mechanisms, GVA authority has count on *ISO standards* (ISO 31000. *Enterprise Risk Management for the Professional*). The usefulness of this standard lies on the fact that it provides a series of risk detection techniques bringing together a trusted and systematic methodology to evaluate practices and processes, allowing its integration with other internal systems. On the other hand, through these self-evaluation plans, the intention is *to involve other management units in the alert system*, so that public managers do not see SALER algorithms as a persecutory tool, but rather as a mechanism to help them in achieving their own goals.

### 5.1.2. Pathways for achieving algorithmic transparency

In this same line, interviews showed different pathways to achieve *algorithmic transparency*, that are expected to have positive effects on discretionary power of public managers. Table 3 points out some of the “carriers” that we have identified in our case study that could be important to accomplish such positive effects of algorithmic transparency. According to the abovementioned literature, we have identified two dimensions of algorithmic transparency (Grimmelikhuisen, 2019) and some carriers connected to them: *accessibility* (external audits, clear rules, openness in design, and access in implementation) and *explainability* (data control to reduce biases, participation of civil servants, intelligible algorithms, algorithms results aligned with civil servants’ knowledge).

On the one hand, to achieve *accessibility external audits of public algorithms are protected and recognized by law*. SALER case shows that recognizing the audits by law can foster accessibility, by granting legal certainty that external controls will be carried out by other agencies. Thus, this condition supports the need to be able to evaluate the use and operation of algorithms in public organizations. In our case, this institutional feature grants that any misuse or negative impact can be reviewed and corrected.

Another pathway to ensure *accessibility* is related to the simplicity of public algorithms. As public *algorithms are based on simple rules*, they will be more accessible and their internal processes more

Table 3  
Pathways for achieving algorithmic transparency

Dimension	Pathway for algorithmic transparency
Accessibility	<ul style="list-style-type: none"> <li>– Audits recognized by law</li> <li>– Algorithms based on simple and clear rules</li> <li>– Openness during the design process</li> <li>– Granted access during the implementation</li> </ul>
Explainability	<ul style="list-style-type: none"> <li>– Control and study of “data” to reduce biases</li> <li>– Involved civil servants as active participants since the beginning of algorithms development</li> <li>– Descriptive model based on machine-learning algorithms that are intelligible</li> <li>– Algorithm results displayed based on civil servants’ administrative knowledge</li> </ul>

Source: Own elaboration.

transparent. This view may lead to think that, as complexity increases, the algorithms will be inevitably less transparent. However, we argue that while the complexity of processes might increase – for example, because there are more rules and the administrative processes involve more situations – it could be possible to keep these rules simple and comprehensible. This condition will be supported by the openness of the algorithms during the design process and by granting access to the system during the implementation.

On the other hand, we have identified four paths to achieve the *explainability* dimension of algorithms in the public sector. First, it is necessary to *keep an eye on the data nurturing the algorithms*. During the first stage of the design it will be a more straightforward process to detect biases with a potential negative impact on people who will use the algorithms in the organization and to whom they could have an impact during their future operation. Secondly, regarding to the case of SALER, the assumption that *civil servants who are going to use the system are involved in certain areas of their development since the beginning* seems to be very important for the understanding of the results of algorithms. This entails that the more the feedback between developers and civil servants who will be final users of the algorithms, the better these civil servants will understand the results of algorithms, and the more this will help them with their duties.

Thirdly, our analysis confirms that *explainability* is very closely related with the *existence or absence of black-boxes*, as the literature has previously determined (Martin, 2019; Geiger, 2017; Introna, 2015). SALER was based on an intelligible descriptive algorithmic model, allowing decision traceability. Other algorithms are considered as “non intelligible”, and as such, it is difficult to manage the traceability of decisions based on them. Also, SALER case shows that when algorithms are intelligible and there are no “black boxes”, the impact of these algorithms are usually positive on the work of civil servants, to the extent that the results and the processes that have been carried out to get these results are easy to follow. Finally, since the *results shown by the system are based on administrative procedures fully acknowledge by civil servants*, their interpretation is simpler, which could also have a positive impact on the efficiency of their work. In public sector settings this feature is critical as embedded knowledge in administrative processes is clearly attached to the discretion of public managers making decisions.

## 5.2. Practical implications

From this point, our article also raises some practical implications for public managers involved in the design and implementation of algorithmic governance systems. Overall, algorithmic governance will be a key issue for public agencies in the coming years. Public managers need to understand the potential implications and consequences of algorithms and AI systems implemented in their public services and

policies. Our case study supports this idea and it also highlights the potential cross-departmental nature of AI and algorithmic governance systems. Particularly, SALER system has a direct impact on corruption prevention, transparency, trust, and good governance with comprehensive results across this regional public administration.

Decision-making and discretion of public employees and managers in their span of responsibility within the public sector will be transformed in the near future. Automated decision-making based on AI systems promises to keep challenged some of the basic assumptions in public sector management, with some ideas already described since the seminal age of Simon (1973). The pervasive impacts of AI in public administration will be correctly governed if public officials and managers design adequate policies, foresee organizational impacts, and anticipate operational implications in different functional areas. Surely, how decisions are made will be part of this general reflection on how to adapt public organizations and services to the disruptive capabilities of AI. Here, better decision-making is not only about being more efficient or productive, but also how to create public value for the citizens served and for the good of the society.

At the same time, SALER system bolsters the window of opportunity for transparency and open government in contexts with high perception of corruption in public settings. As it was advanced in previous sections, this concern has been one of the motives of this system from its inception. Fiscal austerity and recovery grants, loans, or transfers will be part of the landscape of future financial mechanisms after COVID-19; hence public administrations with early warning systems and algorithms oriented to fight against corruption and provide trust to lenders, businesses, and society as a whole, will be in a more favorable position in the coming years to design and deliver better public policies and services. AI and algorithmic systems represent an opportunity to innovate public sector management, whereas they also entail challenges and risks for the future of the public sector.

Finally, algorithms like SALER can potentially change the relationships between civil servants, their tools and the way duties and tasks are organized inside public organizations. Ultimately, we may be talking about the evolution towards an algorithmic bureaucracy (Vogl et al., 2020), although in the case of SALER this statement has not been fully accomplished, regarding the emergent nature of this system. Whatever the case, the more algorithms-based systems mature, the more practitioners should encourage the identification of groups potentially impacted by their implementation, inside (and outside) public organizations.

## 6. Conclusions

This article has presented the case study of SALER, an early warning system implemented at an emerging stage in the government of Valencian region, Spain. Our study is closely attached to the scholar attention that Artificial Intelligence (AI) in the public sector has attracted in different contexts, regarding the implications of algorithmic transparency and discretion in decision-making. Based on an exploratory case study analysis and qualitative semi-structured interviews, our article has preliminarily answered our research questions (*RQ1. To what extent algorithms affect discretionary power of civil servants to make decisions? RQ2. How algorithmic transparency can impact discretionary power of civil servants?*). Among different findings, we have identified different pathways for achieving algorithmic transparency, based on the accessibility and explainability of algorithms. Also, we have identified practical implications for public sector managers working with AI systems and technologies.

This article has also some limitations. First, this study has low external validity and the results are difficult to generalize into different contexts. Second, SALER case is an undergoing project that will

develop its full potential during the coming years. At the same time, the algorithms studied in our case study are not predictive, which exerts some limitations about the conclusions made concerning algorithmic transparency in decision-making. Third, our study has focused on a specific type of public managers (services inspectors) who, unlike street-level bureaucrats, do not have direct contact with external clients (citizens) in their daily work, as they oversee internal clients in their organization (public employees and managers). This fact is distinctive from classical studies on street-level bureaucracy and discretion, and it should be taken into account to understand our results. Finally, the number of service inspectors' interviews may turn out to be somewhat limited and should be increased – if possible – in future explorations, although this article has made a selection based on the most active and representative users of SALER algorithms.

All in all, this study could have future lines of development in different aspects. On the one hand, SALER case will have further developments as the GVA government has planned its full development during the next years. At the same time, different governments around the world are fostering similar cases, at an earlier stage, regarding to the utilization of AI techniques and algorithms to fight against corruption practices and promote better decision-making processes. Besides, this approach to the transparency of public algorithms and discretion in public decision-making needs to be expanded to different public sector settings and areas of activity. In fact, the alignment of algorithmic transparency with other ethical concerns of AI systems opens up one of the most promising areas of research in the future of public administration.

## Acknowledgments

This study was partially supported by the Research Grant Prometeo 2017/064, Generalitat Valenciana Research Agency, and the Research Grant H2019-HUM 5699 (On Trust), Madrid Regional Research Agency and European Social Fund.

## References

- Aarvik, P. (2019). Artificial Intelligence a promising anti corruption tool in development settings? CMIU4 AntiCorruption Resource Centre.
- Agarwal, P.K. (2018). Public administration challenges in the world of AI and bots. *Public Administration Review*, 78(6), 917-921.
- Barth, T.J., & Arnold, F. (1999). Artificial Intelligence and administrative discretion: implications for public administration. *American Review of Public Administration*, 29(4), 332-351.
- Bovens, M., & Stavros, Z. (2002). From street-level to system-level bureaucracies: how information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review*, 62(2), 174-184.
- Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, 358(6370), 1530-1534.
- Bucher, T. (2017). The algorithmic imaginary: exploring the ordinary effects of Facebook algorithms. *Information, Communication & Society*, 20(1), 30-44.
- Burrell, J. (2016). How the machine 'thinks': understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1-12.
- Capdeferro, O. (2020). La inteligencia artificial del sector público: desarrollo y regulación de la actuación administrativa inteligente en la cuarta revolución industrial, IDP. *Revista de Internet, Derecho y Política*, 30, 1-14.
- Cath, C. (2018). Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133).
- Cerrillo, A. (2019). How to open the black box of public administration? Transparency and accountability in the use of algorithms. *Revista Catalana de Dret Públic*, 58, 13-28.
- Creswell, J.W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. Los Angeles: Sage.

- Criado, J.I., & Gil-Garcia, J.R. (2019). Creating public value through smart technologies and strategies: from digital services to artificial intelligence and beyond. *International Journal of Public Sector Management*, 32(5), 438-450.
- Desouza, K. (2018). *Delivering Artificial Intelligence in Government*. IBM for Business of Government.
- Diakopoulos, N., & Koliska, M. (2016). Algorithmic transparency in the news media. *Digital Journalism*, 5(7), 809-828.
- Eisenhardt, K.M. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532-550.
- European Commission. (2019). A definition of Artificial Intelligence. Main capabilities disciplines. High-Level Expert Group on Artificial Intelligence. [Access date October, 1<sup>st</sup>, 2020: <https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>].
- European Parliament. (2019). Artificial Intelligence: Challenges for EU Citizens and Consumers. Policy Department for Economic, Scientific and Quality of Life Policies. Directorate-General for Internal Policies.
- Evans, T. (2010). *Professional Discretion in Welfare Services: Beyond Street-Level Bureaucracy*. London: Ashgate.
- Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, 28(1), 62-70.
- Fink, K. (2018). Opening the government's black boxes: freedom of information and algorithmic accountability. *Information, Communication & Society*, 21(10), 1453-1471.
- Geiger, R.S. (2017). Beyond opening up the black box: investigating the role of algorithmic systems in wikipedian organizational culture. *Big Data & Society*, 4(2), 1-14.
- Giménez, P., Gavara A., Jerez, A.M., Pérez, C., Martínez, J., & Pellicer, L. (2018). El manejo del Sistema de Alertas por parte de la Inspección General de Servicios en la Generalitat Valenciana de España. XXIII Congreso Internacional del CLAD sobre la Reforma del Estado y de la Administración Pública, Guadalajara, México, 6-9 November 2018.
- Grimmelikhuijsen, S. (2019). Deciding by algorithm: testing the effects of algorithmic (non-)transparency on citizen trust. EGPA International Conference, Belfast, 10-13 September 2019.
- Grimmelikhuijsen, S.G., & Meijer, A.J. (2014). Effects of transparency on the perceived trustworthiness of a government organization: evidence from an online experiment. *Journal of Public Administration Research and Theory*, 24(1), 137-157.
- Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18(1), 59-82.
- Harris, J.G., & Davenport, T.H. (2005). Automated decision making comes of age. *MIT Sloan Management Review*, 46(4), 2-10.
- Introna, L.D. (2015). Algorithms, governance, and governmentality: on governing academic writing. *Science, Technology & Human Values*, 41(1), 17-49.
- Janssen, M., & Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly*, 33(3), 371-377.
- Lee, K.F. (2018). *AI Super-powers. China, Silicon Valley, and the New World Order*. New York: Houghton Mifflin Harcourt.
- Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology*, 31(4), 611-627.
- Lipsky, M. (1971). Street-level bureaucracy and the analysis of urban reform. *Urban Affairs Quarterly*, 6(4), 391-409.
- Loi, M., Ferrario, A., & Viganò, E. (2019). Transparency As Design Publicity: Explaining and Justifying Inscrutable Algorithms. *Social Science Research Network*.
- Margetts, H., & Dorobantu, C. (2019). Rethink government with AI. *Nature*, 568(April), 163-165.
- Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics*, 160(4), 835-850.
- Martínez, R. (2019). Inteligencia artificial desde el diseño. Retos y estrategias para el cumplimiento normativo. *Revista Catalana de Dret Públic*, 58, 64-81.
- Martínez, F., Gómez, J.A., & Ferri, C. (2018). La estructura informática del Sistema de Alertas Tempranas. XXIII Congreso Internacional del CLAD sobre la Reforma del Estado y de la Administración Pública, Guadalajara, México, 6-9 November 2018.
- Meijer, A., & Wessels, M. (2019). Predictive policing: review of benefits and drawbacks. *International Journal of Public Administration*, 42(12), 1031-1039.
- Orlikowski, W.J., & Scott, S.V. (2014). What happens when evaluation goes online? Exploring apparatuses of valuation in the travel sector. *Organization Science*, 25(3), 868-891.
- Parrado, S., Dahlström, C., & Lapuente, V. (2018). Mayors and corruption in Spain: same rules, different outcomes. *South European Society and Politics*, 23(3), 303-322.
- Pencheva, I., Esteve, M., & Mikhaylov, S. (2020). Big data and AI – a transformational shift for government: so, what next for research? *Public Policy and Administration*, 35(1), 24-44.
- Ponce, J. (2018). La prevención de riesgos de mala administración y corrupción, la inteligencia artificial y el derecho a una buena administración. *Revista Internacional Transparencia e Integridad*, 6.
- Ponce, J. (2019). Inteligencia artificial, Derecho administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico. *Revista General de Derecho Administrativo*, 50.
- Puncel, A. (2019). Inteligencia artificial para la transparencia pública. *Boletín Económico de ICE*, 3116.
- Ruvalcaba-Gomez, E.A., Criado, J.I., & Gil-Garcia, J.R. (2017). Public Managers' Perceptions About Open Government: A

- Factor Analysis of Concepts and Values. In *Proceedings of the 18th Annual International Conference on Digital Government Research*, pp. 566-567. ACM.
- Simon, H.A. (1973). Applying information technology to organization design. *Public Administration Review*, 33(3), 268-278.
- Simon, H.A. (1991). Bounded rationality and organizational learning. *Organization Science*, 2(1), 125-134.
- Sun, T.Q., & Medaglia, R. (2018). Mapping the challenges of artificial intelligence in the public sector: evidence from public healthcare. *Government Information Quarterly*, 36(2), 368-383.
- Tummers, L., & Bekkers, V. (2014). Policy implementation, street-level bureaucracy, and the importance of discretion. *Public Management Review*, 16(4), 527-547.
- Valero, 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*, 58, 82-96.
- Valle-Cruz, D., Criado, J.I., Sandoval-Almazán, R., & Ruvalcaba-Gomez, E.A. (2020). Assessing the public policy-cycle framework in the age of artificial intelligence: from agenda-setting to policy evaluation. *Government Information Quarterly*, 37(4).
- Van der Voort, H.G., Klievink, A.J., Arnaboldi, M., & Meijer, A.J. (2019). Rationality and politics of algorithms. Will the promise of big data survive the dynamics of public decision making? *Government Information Quarterly*, 36(1), 27-38.
- Velasco, C. (2019). Dossier sobre l'Administració a l'era digital. *Revista Catalana de Dret Públic*, 58, 208-230.
- Villoria, M., Jiménez, F., & Revuelta, A. (2014) 'Corruption Perception and Collective Action'. In Mendilow, J., & Peleg, I. (2014). *Corruption in the Contemporary World*. London: Lexington Books, pp. 197-222.
- Vogl, T.M., Seidelin, C., Ganesh, B., & Bright, J. (2020). Smart Technology and the Emergence of Algorithmic Bureaucracy: Artificial Intelligence in UK Local Authorities. *Public Administration Review*.
- Winfield, A.F.T., & Jirotko, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Phil. Trans. R. Soc.*
- Wirtz, B.W., Weyerer, J.C., & Geyer, C. (2018). Artificial intelligence and the public sector. Applications and challenges. *International Journal of Public Administration*, 13(7), 1-20.
- Wright, N. (2018). How Artificial Intelligence Will Reshape the Global Order. *Foreign Affairs*. Available at: <https://www.foreignaffairs.com/articles/world/2018-07-10/how-artificial-intelligence-will-reshape-global-order>.
- Yin, R.K. (2004). *The Case Study Anthology*. London: Sage.
- Young, M.M., Bullock, J.B., & Lecy, J.D. (2019). Artificial discretion as a tool of governance: a framework for understanding the impact of artificial intelligence on public administration. *Perspectives on Public Management and Governance*, 2(4), 301-313.

## Appendix

Interview protocol (translated from Spanish).

### Part 1. General questions

1. What is your name and position in your organization? What is your academic background? What is your work experience? Could you please give me a brief description of your job and main duties? What is your relationship with SALER system?
2. Could you explain the origin and evolution of this system? What organizational needs do you think the system intends to cover? Do you think they fit with the mission, vision and objectives of the organization? Could you please tell us if you perceived any problem in the design/implementation of the system?
3. From your point of view, to what extent this algorithm can improve the transparency of your institution? To what extent do you think it may have an impact on the fight against corruption? What kind of implications are we talking about?

### Part 2. Impact of SALER system into decisions of the Service Inspectorate unit at Generalitat Valenciana (GVA)

4. In your opinion, to what extent a tool like SALER can affect daily work dynamics of GVA Service Inspectorate? How would you say that the risk models and indicators produced by the algorithm are

impacting the daily work of Service Inspectors?

5. Some people believe that artificial intelligence systems will replace jobs. Do you agree with that statement? Why? In the case of SALER, do you know how the machine learning process occurs? How do you perceive that it can affect the functions related to Service Inspectors? Which functions does it affect the most?

6. The utilization of SALER is enabled by legal norms. In your opinion, what are the impacts of these regulations on your daily work with SALER system? Would you say that it gives you legal certainty, or that, on the contrary, it does not properly cover the use of the system?

7. From your point of view, what impacts might SALER have on the autonomy of Service Inspectors? (We understand “autonomy” as how free you feel to pursue your duties and formal tasks inside the organization).

### *Part 3. Perceptions about algorithmic transparency of SALER system*

8. Some argue that algorithm-based systems are “difficult to understand”. In your opinion, to what extent the information shown by the system is understandable by people? To what extent do you understand how the algorithm bring about a specific response? Do you think this algorithmic process would be easy for citizens to be understood? Why?

9. In your opinion, is it easy (or not) to assess the code and the responses of the algorithms by an external auditor? Do you know if external audits have been carried out to ensure it is working as expected? In your opinion, to what extent can these systems be biased by the values of companies, organizations and people in charge of its development and implementation?

10. In your opinion, how do you think the fact that results of an algorithm like SALER can be easily understood by civil servants might affect their work? How do you think that algorithm understandability affects citizens and the legitimacy of the public administration and government?

11. In your opinion, do you think it is important that the SALER code is free? Do you think that freeing the code may have any impact on the work of the Service Inspectors? And what about the legitimacy of the administration and the government?

### *Part 4. Final comments*

12. Do you have additional comments?

13. Do you have extra documents or reports that could be useful for our research?

14. In your view, who else could be interviewed for this project?