

CERIDAP

RIVISTA INTERDISCIPLINARE SUL
DIRITTO DELLE
AMMINISTRAZIONI PUBBLICHE

Estratto

FASCICOLO

3 / 2023

LUGLIO - SETTEMBRE

Large Language Models for Transparent and Intelligible AI-Assisted Public Decision-Making

Gherardo Carullo

DOI: 10.13130/2723-9195/2023-3-100

Questo articolo esamina l'implementazione dell'intelligenza artificiale nei processi decisionali all'interno della pubblica amministrazione, con un focus su come affrontare le sfide legate alla trasparenza, alla responsabilità e alla comprensibilità delle decisioni generate dall'IA. L'articolo discute in particolare l'importanza dell'imputabilità nelle decisioni prese con algoritmi di deep learning. Sottolinea che concedendo alle pubbliche amministrazioni il pieno controllo sul set di dati per l'addestramento, sul codice sorgente e sulla base di conoscenza, si può garantire l'imputabilità della decisione. Questo controllo permette alle amministrazioni di validare la pertinenza e l'accuratezza dei dati di addestramento dell'algoritmo, affrontare eventuali bias e rispettare i requisiti legali ed etici. Il documento propone quindi l'utilizzo dei Large Language Models (LLM) come soluzione per migliorare la trasparenza e la motivazione dietro le decisioni assistite dall'IA. Mette in evidenza che gli LLM possono generare output testuali articolati e comprensibili che assomigliano da vicino alle decisioni generate dall'uomo, permettendo una comprensione più profonda del processo decisionale. Inoltre, l'articolo enfatizza l'importanza di fornire l'accesso al set di dati per l'addestramento, al codice sorgente e ai precedenti amministrativi individuali per aumentare la trasparenza e la responsabilità. Sostiene che offrendo questi componenti, gli stakeholder possono valutare la validità e l'affidabilità delle decisioni assistite dall'IA, promuovendo la fiducia nel processo decisionale.

This paper examines the implementation of artificial intelligence in decision-making processes within public administration, with a focus on addressing the challenges of transparency, accountability, and the intelligibility of AI-generated decisions. The paper discusses the importance of imputability in decisions made with deep learning algorithms. It emphasises that by granting public administrations full control over the training dataset, source code, and knowledge base, the imputability of the decision

can be ensured. This control enables administrations to validate the relevancy and accuracy of the algorithm's training data, address potential biases, and comply with legal and ethical requirements. The paper then proposes the use of Large Language Models (LLM) as a solution to enhance the transparency and motivation behind AI-assisted decisions. It highlights that LLMs can generate articulate and comprehensible textual outputs that closely resemble human-generated decisions, allowing for a deeper understanding of the decision-making process. Furthermore, the paper emphasises the significance of providing access to the training dataset, source code, and individual administrative precedents to enhance transparency and accountability. It argues that by offering these components, stakeholders can evaluate the validity and reliability of AI-assisted decisions, fostering trust in the decision-making process.

Summary: 1. Introduction.- 2. Levels of AI and types of administrative powers.- 3. Briefly Assessing GDPR's Limits on Automated Decision-Making: The Human-in-the-Loop Approach.- 4. The imputability of decisions automated with deep learning.- 5. The motivation of decisions automated with Large Language Models.- 6. Conclusions.

1. Introduction

Technological evolution has led to techniques of artificial intelligence (AI) capable of interpreting a given context and acting accordingly to achieve one or more objectives^[1]. While traditionally public decisions have been taken so far mostly by human beings^[2], thanks to this capacity, these algorithms can adopt a decision similar, under certain conditions and within certain limits, to that which a human would reach. In some cases, this ability can even exceed the cognitive possibilities of the individual^[3], inferring information from the data that would be difficult for a human to identify^[4].

We can therefore assume that the implementation of AI in decision-making processes of public administrations (PA), in aid or in substitution of human intellectual work, is today technically possible^[5]. Incorporating AI in PA might also be desirable as it could empower decision-makers to arrive at more informed decisions by leveraging data-driven insights and advanced analytical capabilities^[6].

This holds true at least for repetitive and predetermined tasks that can be easily automated, and possibly even for more complex decision-making cases^[7].

Indeed, the use of AI has already been tested in various fields^[8], such as healthcare^[9] and economics^[10]. In the context of public administration, the potential benefits of AI implementation in decision-making processes are numerous^[11]. AI systems can analyze large amounts of data quickly and accurately, leading to more informed decisions^[12]. They can also reduce the risk of bias and error in decision-making, ensuring fair and consistent outcomes for all parties involved. Moreover, AI-based decision-making can save time and resources, allowing public administrators to better allocate available resources^[13].

However, the use of AI in decision-making is not without its challenges. One major concern is the lack of transparency in the decision-making process^[14]. It is important for administrators and the population to be able to understand how AI systems make decisions and ensure that these decisions are based on relevant and accurate information. Additionally, there are ethical considerations to consider, such as the potential for AI to perpetuate or amplify existing biases^[15].

To address these challenges, it is crucial to develop and implement appropriate frameworks and safeguards for the use of AI in public administration^[16]. This includes establishing guidelines for the development and deployment of AI systems to be used by the public sector, ensuring transparency and accountability in decision-making, and addressing ethical concerns. Furthermore, it is essential to involve stakeholders from diverse backgrounds and perspectives in the development and implementation of these frameworks to ensure that the use of AI in public administration is fair and equitable.

It is important to clarify that the focus of this paper is on the implementation and potential benefits of AI in the context of public administration decision-making processes, intentionally addressing the topic as an autonomous and different issue to the one of regulation AI in the private sector^[17]. This paper explicitly refrains from discussing the need for AI regulation in the private sector, as it remains an open question whether such regulation is necessary, and if so, to what extent. In the private sector, there are ongoing debates and discussions surrounding the appropriate level of oversight for AI technologies, but this paper does not engage with those debates. Instead, the paper concentrates on the unique opportunities and challenges that AI presents for public administration^[18],

providing insights and recommendations specific to this domain. By narrowing the scope to public administration, the paper aims to offer a focused examination of AI implementation and its potential impact on decision-making processes within this context.

2. Levels of AI and types of administrative powers

There are numerous and diverse AI techniques available^[19], making it challenging to provide a unifying definition. In its broadest sense, AI can be understood as the ability of a machine to make good decisions, plans, or inferences by adhering to the principles of statistical and economic rationality^[20]. This shared characteristic allows for a common ground to appreciate the various approaches within the expansive field of artificial intelligence.

A first group of algorithms, often referred to as “conditional algorithms,” is characterized by AI systems that base their decisions on the satisfaction or non-satisfaction of specific conditions, employing a structure such as “if...else...”. The decision-making, planning, or inferencing capabilities of conditional algorithms stem from the application of predefined rules that are hardcoded into the algorithm itself^[21]. These rules are typically encoded into the program by one or more human programmers, who translate their understanding of the problem domain into a series of logical steps that the algorithm follows.

In recent times, there have been emerging forms of automation that facilitate the writing of such rules, reducing the need for direct human intervention in the coding process^[22]. Regardless of whether a human being or an automated software generates the conditional algorithm, the essential computer rules that delineate the procedural model of the algorithm are established programmatically. As a result, these rules are accessible and consultable, allowing for transparency in the decision-making process and the ability to trace the logic behind the AI system’s actions.

Conditional algorithms, as a foundational AI concept, enable the development of AI systems by providing a structured and rule-based approach to problem-solving. By clearly defining the conditions and associated actions, conditional algorithms offer a level of predictability and reliability in decision-making processes that can be invaluable in various applications, particularly when dealing

with scenarios where the potential outcomes and factors involved are well understood.

The logical process that underpins these algorithms bears some resemblance to legal reasoning, which suggests they may hold potential for applications in the field of administrative law. In the context of legal norms, a rule is established, and when certain events or conditions arise, specific legal consequences are produced as a result of applying that rule.

The incorporation of conditional algorithms into administrative action does not inherently disrupt the decision-making process of administrative bodies, provided that the applicable legal rule can be effectively translated into one or more machine-interpretable and executable rules. This translation process involves breaking down complex legal principles into simpler, conditional statements that a machine can process, understand, and apply consistently and accurately.

By aligning the algorithmic logic with legal reasoning, conditional algorithms can be used to support or automate administrative decision-making processes in a manner that is compatible with existing legal frameworks^[23]. When properly designed and implemented, these algorithms can enhance the efficiency, consistency, and transparency of administrative actions while still adhering to the underlying legal principles.

However, it is essential to consider potential challenges and limitations when incorporating conditional algorithms into administrative decision-making^[24]. For instance, some legal rules may be too complex or ambiguous to be easily translated into clear-cut conditional statements^[25]. In such cases, a careful analysis of the legal rule and the context in which it is applied may be necessary to determine whether a conditional algorithm is appropriate for the task at hand^[26].

This AI technique, namely conditional algorithms, becomes therefore impractical when the number of conditions to be evaluated is either non-quantifiable or extremely high. In the realm of administrative activities, this eventuality typically arises in situations involving discretionary powers. In such cases, the choice is not strictly predetermined by the law; instead, the administration is granted the flexibility to select the option that seems most appropriate for the specific case at hand.

Discretionary powers require a nuanced understanding of context, the balancing

of competing interests, and the consideration of multiple factors that may not be easily quantifiable or reducible to simple rules^[27]. As a result, the use of conditional algorithms is impractical, if not impossible, for situations involving some degree of discretion, as the inherent complexity and subjectivity of these decisions can make it difficult to capture and encode the decision-making process into a series of clear-cut, machine-interpretable rules^[28].

In such instances, alternative AI techniques might be more appropriate for supporting or informing the decision-making process. For example, machine learning algorithms can be employed to analyze historical data and identify patterns or trends that may inform decision-makers as they exercise their discretionary powers^[29]. Additionally, natural language processing techniques can be used to mine and analyze relevant documents or case law, providing insights and context that may guide administrators in their decision-making process^[30].

Machine learning and deep learning techniques were developed specifically to address the limitations of conditional algorithms^[31]. In deep learning systems, the essential information required for making a particular decision is extracted by the machine itself through the analysis of vast amounts of data^[32]. This data is processed to identify patterns and parameters that enable the system to subsequently make informed decisions^[33]. It is worth emphasizing that the output or “decision” generated by a deep learning-based algorithm can take various forms, such as a boolean value (true or false), a discursive text, or any other data presentation format depending on the specific requirements of the case.

Deep learning systems comprise two key components. The first component is the source code, which is a text written in a specific programming language that can be understood by a human with the requisite technical skills. The second crucial component of a deep learning system is the “model”, a collection of parameters that empowers the algorithm to make informed decisions^[34]. It is important to note that this model is typically not directly intelligible to humans^[35].

The model is developed during the training phase, wherein the algorithm iteratively attempts to represent a given set of big data accurately. Through these repeated attempts, the algorithm evaluates how well the model represents the input data, refining the model in the process^[36]. The reliability of the model is expressed using various measures, such as the percentage of approximation, mean absolute error, or other indicators, depending on the type of operation being

performed^[37].

Lastly, it's noteworthy to mention that a novel technique has recently emerged, which enables the enhancement of a model's knowledge without requiring its re-training. During the execution phase, the model's knowledge base is expanded by introducing new contextual data while it performs a specific task^[38]. In this way the model can adapt to new information and incorporate it into its knowledge base in real time.

3. Briefly Assessing GDPR's Limits on Automated Decision-Making: The Human-in-the-Loop Approach

Articles 13 and 14 of the GDPR mandate that the processor of personal data informs the data subject about the existence of automated decision-making, including profiling, as outlined in Article 22(1) and (4). In such cases, the data subject must be provided with meaningful information about the logic behind the automated decision-making process, as well as the significance and anticipated consequences of this processing.

In the context of public administration, this obligation requires that the administration notify all individuals whose personal data are processed within the decision-making procedure about the presence of an automated decision-making process. Ideally, this notification should include *«simple ways to tell the data subject about the rationale behind, or the criteria relied on in reaching the decision»*^[39].

The purpose of these provisions is to ensure transparency and accountability in the processing of personal data^[40], particularly when automated decision-making is employed. By providing data subjects with information about the logic and consequences of automated decision-making, the GDPR seeks to empower individuals with the knowledge necessary to understand the potential impact of such systems on their lives and to exercise their rights accordingly^[41].

Subject to prevailing national legislation^[42], these European provisions do not inherently preclude the adoption of automated decision making within public administration. Instead, these provisions aim to establish a framework that ensures that data subjects are in a position to *«understand the logic underlying artificial intelligence decisions»*^[43].

Art. 22 of the GDPR is another crucial aspect to consider when discussing the use of AI and deep learning in public administration decision-making processes. The first paragraph of this article states that data subjects have *«the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her»*.

The adverb *«solely»* plays a critical role in the context of using deep learning techniques by public administrations. Administrative action is typically proceduralized, and, as of now, key human figures remain integral to the formation of any decision. Consequently, it can be reasonably argued that, as long as human intervention remains genuine and meaningful^[44], i.e., *«by a human who is competent and professionally authorised to do so»^[45]*, a decision made by an authority in the exercise of its functions would not be adopted *«solely»* in an automated manner^[46].

This concept of maintaining a human-in-the-loop ensures that, while AI and deep learning systems contribute to the decision-making process, human operators retain ultimate control and responsibility. By keeping a human presence at critical stages, public administrations can maintain compliance with the GDPR and address potential concerns related to accountability, transparency, and fairness.

The human-in-the-loop approach allows public administrations to harness the benefits of AI and deep learning techniques in their decision-making processes while adhering to the GDPR's provisions. This approach strikes a balance between leveraging advanced technology and ensuring that the human element remains central to the administrative process, safeguarding the rights of data subjects and maintaining the integrity of the decision-making process.

While certain national legislations may be in the process of sanctioning fully automated decisions^[47], we maintain that it remains crucial that human participation is retained within the administrative procedure^[48]. This ensures adherence to the GDPR and other pertinent legal structures, such as the right to a good administration, while upholding accountability in decision-making processes. Consequently, for the purposes of this paper and given the existing legal landscape, we may conclude that further exploration of the GDPR's provisions concerning fully automated decision-making is unnecessary, provided

that the human-in-the-loop approach is maintained.

A distinct issue pertains to the utilization of personal data during both the training and execution phases of AI, as exemplified by the case involving OpenAI, which was instigated by the Italian data protection authority^[49]. However, this issue goes beyond the scope of this paper, as it primarily pertains to data protection, as opposed to our specific focus on AI in public administration decision-making processes. Consequently, we will not delve into this aspect further in this context.

4. The imputability of decisions automated with deep learning

During the training phase, deep learning generates the “rules” to be applied in the subsequent execution phase^[50]. This scenario is fundamentally different from the one involving conditional algorithms. As outlined above, in the case of conditional algorithms, an agent writes the rules based on the legal provisions applicable to a specific case. In contrast, deep learning does not involve a definite codification of the rules that will guide the machine’s decisions.

In deep learning, the model is derived from the analysis of the dataset on which the algorithm has been trained. Essentially, deep learning generates the parameters that will govern the execution phase by attempting to create an action model based on pre-existing data^[51]. This implies that deep learning must be trained using a vast collection of administrative decisions pertaining to specific cases^[52]. Additionally, if new information is incorporated during the execution phase through a knowledge-base approach^[53], this additional information must also be pertinent to the case under consideration.

The administrative decision made with the support of deep learning can thus be contextualized within the framework of administrative precedent^[54]. The model provides a representation of administrative cases upon which future evaluations of similar cases can be based. Consequently, the imputability of decisions automated with deep learning is grounded in the historical and practical knowledge derived from previously resolved cases. This knowledge, when combined with the algorithm’s ability to identify patterns and relationships within the data, can lead to more informed and effective decision-making in

public administration^[55]. However, it is essential to acknowledge and address the potential biases and limitations that may arise from non-objective data, or the risk that data «*can easily be manipulated and may be biased to reflect cultural, gender, national or other prejudices and preferences*»^[56].

It is thus of utmost importance that the administration maintains complete control over the dataset used during the training phase^[57]. Ensuring that the training dataset and knowledge base is attributable to the administration is a crucial factor in establishing the imputability of the model generated from such dataset.

By exercising control over the training data, the administration can be made accountable to confirm that the information used to train the deep learning algorithm is relevant, accurate, and representative of the cases that the AI system will encounter^[58]. This, in turn, can help guarantee that the model's predictions and recommendations are grounded in the reality of the administration's decision-making context.

Additionally, having full control over the training dataset and knowledge base allows the administration to minimize potential biases and inaccuracies that may be present in the data. This is particularly important in ensuring that AI-based decisions are fair, transparent, and adhere to ethical standards, including for compliance with the GDPR^[59].

One could argue that there is a potential risk that the human official responsible for overseeing the decision-making process might simply accept the decision proposed by the AI system without any genuine review. As a matter of fact, assuming the algorithm functions properly, the extent of the human's constructive contribution to the automated decision-making process may not be of great significance. As previously mentioned, the algorithm is built upon administrative precedents and can be instructed to apply them to new cases by formulating the text of the decision.

While such a situation might pose issues under the GDPR^[60], regarding the imputability of the decision, if the competent authority endorses the decision and affixes their signature, the content of the administrative act must be ascribed to the administration to which they belong. This ensures that the principles of accountability and responsibility are maintained in the decision-making process. Consequently, judicial review remains possible under the same conditions as if

the administrative act had been carried out entirely without the assistance of any algorithms.

The primary concern here is ensuring that human officials do not become overly reliant on AI-generated decisions and abdicate their responsibility to carefully review and assess the recommendations provided by the AI system. This can be addressed through proper training and guidelines for officials working alongside AI tools, emphasizing the importance of critical thinking and active engagement in the decision-making process^[61].

As a result, the decision made with the support of deep learning algorithms can be attributed to the administration, provided that it has full control, choice, and access to the dataset used in the algorithm's training phase, as well as the source code of the algorithm itself, and any other knowledge base used during the execution phase^[62]. By ensuring these conditions are met, the administration is able to maintain its influence and authority over the model utilized in the decision-making process.

This approach emphasizes the importance of the administration's active involvement in every stage of the AI implementation, from selecting the data and overseeing the algorithm's development to monitoring its performance and outcomes. By retaining control over these aspects, the administration can ensure that the AI system operates in line with its intended goals and values, while also taking responsibility for the resulting decisions. This approach not only maintains the accountability of the administration but also strengthens the legitimacy and transparency of AI-assisted decision-making processes.

5. The motivation of decisions automated with Large Language Models

A critical aspect of addressing decisions made with the assistance of digital tools is the motivation behind the decision. As previously mentioned, deep learning algorithms do not rely on hard-coded rules that can be easily verified by humans. Instead, the decision is based on an evaluation of the issue through a model generated during the training phase.

Consequently, a common contention is that AI methodologies, particularly deep learning, should be avoided in the public sector due to their inherent "opacity"^[63].

Particularly, concerns about the transparency and intelligibility of the decision-making process frequently arise. Given that the algorithm's rationale is not explicitly articulated in a format understandable by humans^[64], it is contended that comprehending the motivations behind the decisions, and ensuring their adherence to legal and ethical standards, presents a substantial challenge.

We, however, posit that this purported opacity is more perceived than real, thanks to the most recent advancements in AI techniques. It is crucial to underscore that a deep learning system can be designed to generate a textual output that mirrors the structure and consistency of its training data, but with a different content^[65]. This ability to generate articulate decisions that closely resemble human-generated decisions is a powerful feature of deep learning algorithms. Presently, this class of AI systems is known as Large Language Models (LLM)^[66].

This means that if deep learning is trained using discretionary administrative acts as a basis, the algorithm can be configured to learn how to formulate texts that resemble those found in these acts. Consequently, once the training is complete, during the execution phase, the algorithm can produce a well-structured decision that employs the same terms as those observed during the training phase, but with new content that is coherent with the new input data^[67].

Given the necessity to comprehend the rationale behind an automated decision^[68], we can thus now rely on AI systems that produce intelligible decisions articulated in textual format. As such, these decisions are comprehensible to humans. This may also be supported by additional output information, to provide clearer and more comprehensible explanations of the algorithm's decision-making process.

A deep learning system intended for use by public bodies should therefore be designed to generate textual output that closely resembles what a human would produce when addressing the same issue. Rather than providing a simple binary outcome, such as accepted or rejected, the algorithm must be configured to create an argumentative text that effectively communicates the reasoning behind the decision. Of course, this assumes that the algorithm functions correctly, meaning it has the ability to produce coherent and meaningful text.

In the contemporary advancements witnessed in the field, the probability of a LLM generating outputs that are incoherent or incomprehensible is considerably low. Nevertheless, in the rare occurrence that such discrepancies do manifest, a

meticulous case-by-case evaluation would be necessitated to ascertain the validity of the output. Any resulting text found to lack coherence or relevance shall be classified as unusable.

Moreover, it should be noted that it is more likely for LLMs, including prominent examples like chatGPT^[69], to sometimes generate outputs that are not entirely aligned with the true complexity of the issue at hand, occasionally resulting in “hallucinations” or inaccurate representations of the input data. This risk is particularly heightened when LLMs interface with complex situations comprising many facts, where they might create seemingly sound legal reasoning based on the given context, even though they lack a genuine comprehension of the underlying motivations or intricacies of these results.

Given these potential pitfalls, it becomes essential that public officials and operators rigorously review and validate the textual suggestions offered by LLMs. In cases where the AI system might not fully understand the context or nuances of a particular prompt or question, it’s incumbent upon human oversight to discern the veracity and applicability of AI-generated content. This requires the development of robust frameworks for evaluating the outputs of these systems, thereby ensuring that the utilization of LLMs complements rather than compromises the integrity of public decision-making processes^[70]. Thus, the challenge lies not just in crafting algorithms capable of generating human-like textual outputs, but in seamlessly merging the capacities of these AI systems with expert human judgment to forge a pathway to more transparent and intelligible public decision-making.

Consider a deep learning system that has been trained on millions of environmental assessment procedures to evaluate the compatibility of various interventions on a given territory. When applied to a new case, the output generated by this algorithm should resemble the format and structure of previous assessments while addressing the unique aspects of the specific case at hand. Instead of producing a mere boolean result indicating compatibility or incompatibility, the algorithm should generate a comprehensive compatibility analysis, drawing upon the parameters established during the training phase.

In this scenario, the deep learning system would effectively analyze the new case by considering the various factors and criteria that have been learned from the vast array of prior environmental assessments. The resulting output would

provide a detailed and reasoned compatibility analysis, enabling stakeholders to better understand the rationale behind the decision and its implications on the proposed intervention.

The first aspect of the motivation to consider in a decision made by a deep learning system is the output text generated by the machine. On top of this, in a decision adopted by a LLM, additional factors that would not be available in a decision made by a human can be taken into account. These may include the deviation of the algorithm from previous cases, as well as the relevance of those cases in informing the decision at hand.

In evaluating the output text, it is crucial to acknowledge that this text should mirror what a human would have produced. Therefore, such text should only be incorporated into the final decision of the PA if its contents are valid and legally sound, which can be determined by the competent public official responsible for the case in question.

If the text clears this initial hurdle, further reviews of the output are feasible. As previously stated, the dataset utilized for instructing the system must include decisions already rendered by the administration on cases analogous to those the algorithm will be ruling on. Therefore, the system could, for example, highlight the key precedents that played a decisive role in shaping the decision. Incorporating such a feature into a deep learning system would require the algorithm to be programmed to recognize relevant precedents during the execution phase, with techniques such as vectorization and similarity comparison^[74]. By incorporating an automated mechanism capable of comparing the output with pertinent precedents, it can be ensured that new decisions are in alignment with those previously made by the same administration for similar cases under the same regulatory framework, ultimately enhancing the transparency and accountability of the decision-making process.

To make these circumstances verifiable by interested parties, it is necessary to grant access to both the source code of the algorithm and the dataset used for training^[72]. The dataset is crucial for understanding how the algorithm operates – particularly in evaluating whether the algorithm has been trained on data that genuinely corresponds to the case being assessed – and if any bias may exist^[73]. For this reason, the entire dataset used by the AI should be made freely available to anyone. Moreover, since any modification of the data can result in variations of

the final output, it is essential that the data is made available exactly as it was during the training process. This transparency ensures that all parties involved can confidently assess the validity and reliability of decisions made with the assistance of deep learning algorithms.

In order to address the evident concerns surrounding personal data protection, it is crucial to ensure that any dataset used by a public administration for instructing a deep learning system is fully anonymized before it is ever utilized. By taking this precautionary measure, the privacy of individuals can be protected, while still allowing for the effective training of deep learning algorithms.

To ascertain the relevance of the dataset fed to the system, it is essential to examine whether the cases used to instruct the model genuinely align with the situation being decided. Before exploring how this might be accomplished, it is important to emphasize that all interested parties must have access to the complete dataset. The responsibility of determining which portions of the dataset are relevant for validating the deep learning system's model should not rest solely on the administration.

By offering access to both the final text of the decision, and these additional components that are normally not available when the decision is adopted solely by a human being, a deeper understanding of the decision-making process can be achieved, ensuring a more robust and transparent justification. In doing so, stakeholders can have greater confidence in the decisions made using deep learning systems, as the motivation behind the decisions is well-documented and can be scrutinized by interested parties. This comprehensive approach to justification not only fosters trust in automated decision-making processes but also promotes accountability and transparency within public administration.

6. Conclusions

To ensure the reliability and effectiveness of LLM in public administration, it is crucial to establish rigorous quality control measures and testing procedures. These measures should verify the algorithm's ability to produce accurate, coherent, and meaningful argumentative text that adheres to the standards and expectations of human decision-makers. Additionally, ongoing monitoring and evaluation of the system's performance will help identify any issues or anomalies

that may arise, ensuring that the AI-assisted decision-making process remains transparent, understandable, and accountable.

In the context of automated decisions made using LLM, we argue that the duty to state the reasons of the decision^[74] can be met even more comprehensively and thoughtfully than in cases where decisions are not supported by such systems. Beyond the content of the administrative act as generated by the computer system, the rationale can be further substantiated by incorporating the source code, the training dataset, and the individual administrative precedents that best matched the case at issue.

The utilization of LLMs as outlined in the preceding sections presents a viable solution to the “black box” problem often associated with algorithmic decision-making processes^[75]. The transparent and intelligible nature of decisions generated by these models facilitates a level of scrutiny that parallels, if not exceeds, that of traditional administrative decisions. Decision recipients, as well as overseeing bodies, are in fact granted the ability to review not only the resulting text but also the underlying arguments and considerations that influenced the outcome, fostering a deeper layer of transparency and accountability.

The integration of LLMs essentially creates a dual-layer review mechanism where decisions can be evaluated based on the generated textual output, in a manner synonymous with conventional administrative scrutiny, coupled with an additional layer of analysis focusing on the supporting arguments that the system utilized during the decision-making process. This facilitates a more in depth and comprehensive review, allowing stakeholders to better dissect and understand administrative decisions.

Consequently, by designing deep learning systems that generate argumentative text outputs comparable to those of human decision-makers, public administrations can harness the power of AI while setting a new benchmark in administrative accountability and scrutiny. At the same time, by enhancing the transparency and understandability of AI-assisted decisions, public administrations can build trust in these systems and ensure that their use is welcomed by the recipients of the decisions.

AI represents a remarkable resource for public authorities, with the potential to substantially enhance administrative action by automating various aspects of public bodies’ operations, thus improving the effectiveness and efficiency of their

activities. Specifically, the implementation of AI in the public sector can automate a wide range of repetitive tasks, enabling officials to redirect their efforts towards more critical functions that directly impact citizens' lives.

As AI systems become more advanced and sophisticated, they can not only process information more rapidly than human beings but also analyze vast amounts of data to identify patterns and trends, thereby enabling more informed and data-driven decision-making. The use of AI in the public sector has the potential to revolutionize the way public services are delivered, leading to more personalized and accessible services for citizens while also reducing the time and cost associated with manual processes.

The adoption of AI in public administration is becoming an inevitable evolution to ensure its sustainability, both from an economic and a functional standpoint. In other words, implementing AI is essential for equipping administrations with the necessary tools to effectively address citizens' needs and safeguard public interests, particularly in a context of continuous change and rapid scientific advancements that increasingly challenge public institutions.

This technological progress must be carried out in compliance with individuals' rights and in adherence to the procedural guarantees provided to protect them. Specifically, the obligation to provide reasons for administrative decisions must be preserved to supply citizens with the appropriate tools to assess the legitimacy of decisions anyhow adopted by public bodies.

Among these factors, data assumes a pivotal role. In a data-driven society, access to data is a crucial element for understanding public authorities' decisions made with deep learning algorithms. And since data, without the appropriate tools for analysis, cannot be effectively interpreted to extract the information they potentially convey, it is imperative that the administration, along with the data, also provide the tools necessary for accurate and effective analysis and review.

By doing so, administrations can strike a balance between harnessing the transformative power of AI technology and ensuring transparency, accountability, and respect for individuals' rights. This approach will enable public institutions to evolve alongside technological advancements while preserving the fundamental principles of fairness, justice, and due process that underpin democratic societies^[6].

1. For a brief history of the field of artificial intelligence, from a technical perspective, see F. Chollet, *Deep learning with Python*, II Ed., Manning, Shelter Island, New York, 2021, p. 13.
2. E. Szewczyk, *Artificial intelligence in administrative law and procedure*, in *Human Rights - From reality to the virtual world*, WSGE, Józefów, 2021, p. 116.
3. On the «*biological constraints on perceptual and computational capacity*» of humans decision-makers, see K. Bamberger, *Regulation as Delegation: Private Firms, Decisionmaking, and Accountability in the Administrative State*, in *Duke Law Journal*, 2006, p. 410.
4. I. Pilving, *Guidance-based Algorithms for Automated Decision-Making in Public Administration: the Estonian Perspective*, in *CERIDAP*, 1, 2023, p. 65.
5. T.J. Barth, E. Arnold, *Artificial Intelligence and Administrative Discretion*, in *ARPA*, vol. 29, 1999, p. 337.
6. M. Grygorak, I. Isaienko, T. Kuznietsova, *The role of artificial intelligence in the europeanization of logistics public services*, in *International Journal of New Economics, Public Administration And Law*, 2, 2018; K.M. Wiig, *Knowledge management in public administration*, in *Journal of Knowledge Management*, 2002.
7. T.M. Vogl, C. Seidelin, B. Ganesh, J. Bright, *Smart Technology and the Emergence of Algorithmic Bureaucracy: Artificial Intelligence in UK Local Authorities*, in *PAR*, 2020, Wiley (Blackwell Publishing), p. 950.
8. A. Sigfrids, M. Nieminen, J. Leikas, P. Pikkuaho, *How Should Public Administrations Foster the Ethical Development and Use of Artificial Intelligence? A Review of Proposals for Developing Governance of AI*, in *Frontiers in Human Dynamics*, vol. 4, 2022, Frontiers Media SA, p. 1.
9. E. Topol, *The Topol Review: Preparing the healthcare workforce to deliver the digital future*, in *HEE, Leeds*, 2019.
10. J.M. Cavanillas, E. Curry, W. Wahlster (Eds), *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe*, Springer, Cham, 2016.
11. M. Grygorak, I. Isaienko, T. Kuznietsova, *The role of artificial intelligence in the europeanization of logistics public services*, cit..
12. L. Kysh, *Institutional Support for the Digital Transformation of Public Administration*, in *Knowledge, Education, Law, Management*, 3, 2022; K.M. Wiig, *Knowledge management in public administration*, cit..
13. T.M. Vogl, C. Seidelin, B. Ganesh, J. Bright, *Smart Technology and the Emergence of Algorithmic Bureaucracy*, cit., p. 953.
14. E. Szewczyk, *Artificial intelligence in administrative law and procedure*, cit., p. 116.
15. I. Sobrino-García, *Artificial Intelligence Risks and Challenges in the Spanish Public Administration: An Exploratory Analysis through Expert Judgements*, in *Administrative Sciences*, vol. 11, 2021, p. 4.
16. D. Piana, *Legal services and digital infrastructures: a new compass for better governance*, Routledge, New York, 2021; I. Ulnicane, W. Knight, T. Leach, B.C. Stahl, W.-G.

- Wanjiku, *Governance of Artificial Intelligence*, in M. Tinnirello (Ed.), *The Global Politics of Artificial Intelligence*, Taylor & Francis, 2022, p. 31.
17. R. Kennedy, *E-Regulation and the Rule of Law: Smart Government, Institutional Information Infrastructures, and Fundamental Values*, Social Science Research Network, Rochester, NY, 2015, p. 22.
 18. C.E. Jimenez-Gomez, J. Cano-Carrillo, F.F. Lanas, *Artificial Intelligence in Government*, in *Computer*, vol. 53, 10, 2020, p. 25.
 19. F. Chollet, *Deep learning*, cit., p. 13.
 20. S. Russell, T. Dietterich, E. Horvitz, B. Selman, F. Rossi, D. Hassabis, S. Legg, M. Suleyman, D. George, S. Phoenix, *Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter*, in *AI Magazine*, Winter, 2015, p. 3.
 21. F. Chollet, *Deep learning*, cit., p. 3.
 22. M. Chen, J. Tworek, H. Jun, Q. Yuan, H.P. de O. Pinto, J. Kaplan, H. Edwards, Y. Burda, N. Joseph, G. Brockman, A. Ray, R. Puri, G. Krueger, M. Petrov, H. Khlaaf, G. Sastry, P. Mishkin, B. Chan, S. Gray, N. Ryder, M. Pavlov, A. Power, L. Kaiser, M. Bavarian, C. Winter, P. Tillet, F.P. Such, D. Cummings, M. Plappert, F. Chantzis, E. Barnes, A. Herbert-Voss, W.H. Guss, A. Nichol, A. Paino, N. Tezak, J. Tang, I. Babuschkin, S. Balaji, S. Jain, W. Saunders, C. Hesse, A.N. Carr, J. Leike, J. Achiam, V. Misra, E. Morikawa, A. Radford, M. Knight, M. Brundage, M. Murati, K. Mayer, P. Welinder, B. McGrew, D. Amodei, S. McCandlish, I. Sutskever, W. Zaremba, *Evaluating Large Language Models Trained on Code*, in *arXiv*, 2107.03374, 2021.
 23. T.M. Vogl, C. Seidelin, B. Ganesh, J. Bright, *Smart Technology and the Emergence of Algorithmic Bureaucracy*, cit., p. 953. For a practical implementation of such systems, see G. Contissa, F. Lagioia, M. Lippi, H.-W. Micklitz, P. Palka, G. Sartor, P. Torroni, *Towards Consumer-Empowering Artificial Intelligence*, International Joint Conferences on Artificial Intelligence Organization, 2018, p. 5152.
 24. G. Contissa, F. Lagioia, M. Lippi, H.-W. Micklitz, P. Palka, G. Sartor, P. Torroni, *Towards Consumer-Empowering Artificial Intelligence*, cit., p. 5152.
 25. M. Adler (Ed.), *Administrative Justice in Context*, Bloomsbury Publishing, Oxford, 2010, p. 168. For an example related to risk management, see T. Mahler, *Tool-supported Legal Risk Management: A Roadmap*, in *EJLS*, 3, 2010, p. 22.
 26. L.B. Moses, *Recurring Dilemmas: The Law's Race to Keep up with Technological Change*, in *U. Ill. J.L. Tech. & Pol'y*, vol. 2007, 2, 2007, p. 283.
 27. D.E. Hall, *Administrative Law: Bureaucracy in a Democracy*, VI Ed., Pearson Prentice Hall, 2015, pp. 341–343.
 28. S. Zouridis, M. Thaens, *E-Government: Towards a Public Administration Approach*, in *Asian Journal of Public Administration*, vol. 25, 2, 2003, p. 175.
 29. E.J. Walters, *Data-Driven Law: Data Analytics and the New Legal Services*, CRC Press, 2018.
 30. C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P.J. Liu, *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*, in

- arXiv*, 1910.10683, 2020.
31. F. Chollet, *Deep learning*, cit., p. 10.
 32. M. Mohri, A. Rostamizadeh, A. Talwalkar, *Foundations of Machine Learning*, MIT Press, Cambridge, 2012
 33. *Ibid.*, p. 1.
 34. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, Cambridge, Massachusetts, 2016.
 35. J.L. Slosser, F.F.I.U.I. Rechtsforschung, *Thoughts on the Black Box: Getting to Cooperative Intelligence in Public Administration*, 2022.
 36. D.E. Rumelhart, G.E. Hinton, R.J. Williams, *Learning representations by back-propagating errors*, in *Nature*, vol. 323, 1986, p. 533.
 37. F. Chollet, *Deep learning*, cit., p. 5.
 38. B. Peng, M. Galley, P. He, H. Cheng, Y. Xie, Y. Hu, Q. Huang, L. Liden, Z. Yu, W. Chen, J. Gao, *Check Your Facts and Try Again: Improving Large Language Models with External Knowledge and Automated Feedback*, in *arXiv*, 2302.12813, 2023.
 39. Article 29 Data Protection Working Party, “Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679”, of 6 February 2018 (17/EN WP251rev.01), p. 25.
 40. On «*the core principle of transparency underpinning the GDPR*» on this topic, see the already mentioned “Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679”, by the Article 29 Data Protection Working Party, p. 16.
 41. M. Krzysztofek, *GDPR: General Data Protection Regulation (EU) 2016/679: Post-Reform Personal Data Protection in the European Union*, Kluwer Law International B.V., Alphen aan den Rijn, 2018, p. 144.
 42. National legislation, depending on the specific context and jurisdiction, may impose additional restrictions on the use of AI in the public sector. These limitations could be in the form of more stringent transparency requirements, or stronger safeguards against potential biases in AI-driven decisions. See, for example, J.-P. Schneider, F. Enderlein, *Automated Decision-Making Systems in German Administrative Law*, in *CERIDAP*, 1, 2023. For Sweden, see J. Reichel, *Regulating Automation of Swedish Public Administration*, in *CERIDAP*, 1, 2023 For an overview of the approaches adopted by Member States, see H.C.H. Hofmann, *Comparative Law of Public Automated Decision-Making. An Outline*, in *CERIDAP*, 1, 2023.
 43. A.P. Czarnowski, K. Kunda, M. Gawronski, M. Kibil, M. Sztaberek, P. Punda, *Data Subject’s Rights*, in M. Gawronski (Ed.), *Guide to the GDPR*, Kluwer Law International, Alphen aan den Rijn, 2019, p. 163
 44. M. Krzysztofek, *GDPR*, cit., p. 172.
 45. A.P. Czarnowski, K. Kunda, M. Gawronski, M. Kibil, M. Sztaberek, P. Punda, *Data Subject’s Rights*, cit., p. 179.
 46. On the need to ensure «*human supervision on outcomes affecting people*», see I.M. Wróbel,

- Artificial intelligence systems and the right to good administration*, in *Review of European and Comparative Law*, vol. 49, 2022, p. 219.
47. See for example Estonia I. Pilving, *Guidance-based Algorithms for Automated Decision-Making in Public Administration: the Estonian Perspective*, cit.. For a broader comparative perspective, see G. Gallone, *Riserva di umanità e funzioni amministrative*, CEDAM, Milano, 2023, p. 128.
 48. Similar conclusions are reached, in relation to the private sector, by B. Cappiello, *AI-systems and non-contractual liability*, Giappichelli, Torino, 2022, p. 215.
 49. On March 30, 2023, the Italian Data Protection Authority expressed concerns about ChatGPT's handling of user data, particularly its lack of age verification and inadequate legal basis for personal data collection and processing. The authority highlighted inaccuracies in the data provided by ChatGPT, as well as potential risks to minors under 13 due to a lack of filters. As a result, they issued a provisional restriction on ChatGPT's processing of personal data for Italian users, effective immediately. See the Decision of March 30, 2023, n. 112, available at www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9870832.
 50. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, cit..
 51. F. Chollet, *Deep learning*, cit., p. 133.
 52. On the availability, for larger organization, of pre-categorized data, suitable for training AI models, see F. Sebastiani, *Machine Learning in Automated Text Categorization*, in *ACM Comput. Surv.*, vol. 34, 1, 2002, p. 2.
 53. B. Peng, M. Galley, P. He, H. Cheng, Y. Xie, Y. Hu, Q. Huang, L. Liden, Z. Yu, W. Chen, J. Gao, *Check Your Facts and Try Again*, cit..
 54. G. Contissa, F. Lagioia, M. Lippi, H.-W. Micklitz, P. Palka, G. Sartor, P. Torroni, *Towards Consumer-Empowering Artificial Intelligence*, cit., p. 5151.
 55. Similar conclusions are reached by T.M. Vogl, C. Seidelin, B. Ganesh, J. Bright, *Smart Technology and the Emergence of Algorithmic Bureaucracy*, cit., p. 953. See also T.J. Barth, E. Arnold, *Artificial Intelligence and Administrative Discretion*, cit., p. 337, who proposed that «*there is the potential to have an AI system that approximates ideal rational man in the sense that all assumptions underlying the thinking and therefore decision-making processes can be understood. From the administrative discretion perspective, such capabilities raise the possibility of an unquestioningly loyal system open to any set of values, motives, and goals that are imposed on it*».
 56. E. Szewczyk, *Artificial intelligence in administrative law and procedure*, cit., p. 116.
 57. On the risks of obtaining datasets from external companies, see I. Sobrino-García, *Artificial Intelligence Risks and Challenges in the Spanish Public Administration*, cit., p. 4.
 58. On the correlation between training data contents and accuracy of the trained model, M. Mohri, A. Rostamizadeh, A. Talwalkar, *Foundations of Machine Learning*, cit..
 59. A.P. Czarnowski, K. Kunda, M. Gawronski, M. Kibil, M. Sztaberek, P. Punda, *Data Subject's Rights*, cit., p. 181.
 60. See what outlined in paragraph 3 above.

61. Similar conclusions are reached, in relation to the private sector, by B. Cappiello, *AI-systems and non-contractual liability*, cit., p. 215.
62. On the elements of the AI system to analyze, see C. Natali, *AI Ethics Auditing. Risk Impact Assessment of Artificial Intelligence in the Public Administration of Justice*, 2022, University of Milan, Milan, p. 119.
63. I.M. Wróbel, *Artificial intelligence systems and the right to good administration*, cit., p. 210.
64. T. Mahler, *Introduction*, cit., p. 17.
65. T.B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D.M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, *Language Models are Few-Shot Learners*, in *arXiv*, 2005.14165, 2020.
66. W.X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, Y. Du, C. Yang, Y. Chen, Z. Chen, J. Jiang, R. Ren, Y. Li, X. Tang, Z. Liu, P. Liu, J.-Y. Nie, J.-R. Wen, *A Survey of Large Language Models*, in *arXiv*, 2303.18223, 2023, data di consultazione 14 maggio 2023, in <http://arxiv.org/abs/2303.18223>.
67. In this sense, the realization of "electronic assistants" capable of supporting public decision-making processes has been proposed by E. Schweighofer, *An e-Government Interface for the Director-General – Or: How to Support Decision Makers with an Electronic Chief Secretary?*, in R. Traunmueller (Ed.), *Electronic Government: Third International Conference, EGOV 2004, Zaragoza, Spain, August 30-September 3, 2004, Proceedings*, Springer-Verlag, Berlin Heidelberg, 2004 (Lecture Notes in Computer Science), p. 148.
68. E. Szewczyk, *Artificial intelligence in administrative law and procedure*, cit., p. 116.
69. Available at <https://chat.openai.com>.
70. Similar conclusions are drawn by I.M. Delgado, *La aplicación del principio de transparencia a la actividad administrativa algorítmica*, in E. Gamero Casado (a cura di), *Inteligencia artificial y sector público*, Tirant lo Blanch, Valencia, 2023, p. 174.
71. O. Shahmirzadi, A. Lugowski, K. Younge, *Text Similarity in Vector Space Models: A Comparative Study*, in Atti del Convegno «2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)», dicembre 2019, in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*.
72. On the need to make the training dataset available, see I.M. Wróbel, *Artificial intelligence systems and the right to good administration*, cit., p. 219.
73. E. Szewczyk, *Artificial intelligence in administrative law and procedure*, cit., p. 116.
74. As defined by EU case law, see for example the judgment of the Court of Justice in case C-137/92 P, *Commission/B.A.S.F.*, paragraph 67
75. See T. Wischmeyer, *Artificial Intelligence and Transparency: Opening the Black Box*, in T. Wischmeyer; T. Rademacher (Eds.), *Regulating Artificial Intelligence*, Springer, Cham, 2020, p. 87.
76. On the requirements of public decision-making processes in democracies, see S. Rose-

CERIDAP

Ackerman, *Democracy and executive power: policymaking accountability in the US, the UK, Germany, and France*, Yale University Press, New Haven, 2021.