ISSN 2723-9195

CERIDAP

Rivista Interdisciplinare sul Diritto delle Amministrazioni Pubbliche

Estratto

Fascicolo
4 / 2025

OTTOBRE - DICEMBRE

Artificial Intelligence, Technical Discretion, and the Limits of Judicial Review in Italian Public Procurement

Tommaso Cesareo

DOI: 10.13130/2723-9195/2025-4-59

Il contributo analizza le sfide poste dall'integrazione dell'intelligenza artificiale (IA) nelle gare d'appalto pubbliche in Italia, prendendo spunto dalla sentenza del TAR Lazio n. 4546 del 3 marzo 2025, successivamente confermata dal Consiglio di Stato con sentenza n. 8092 del 20 ottobre 2025, come caso di studio chiave. Questa pronuncia, la prima in Italia ad affrontare l'uso di IA generativa (ChatGPT-4) in un'offerta tecnica, funge da lente per esaminare l'adeguatezza dei quadri giuridici tradizionali. L'analisi critica la decisione del TAR di respingere il ricorso basandosi sui principi consolidati dell'ampia discrezionalità tecnica della commissione giudicatrice e sulla limitata sindacabilità delle valutazioni effettuate con il metodo AHP (Analytic Hierarchy Process). L'autore sostiene che questa impostazione crei un "doppio scudo" che rende quasi impossibile un sindacato giurisdizionale effettivo sulla sostanza tecnica delle soluzioni di IA, generando un "paradosso deferenziale".

This paper analyses the challenges arising from the integration of Artificial Intelligence (AI) in Italian public procurement, using judgement n. 4546 from the Lazio Regional Administrative Court on 3 March 2025, as subsequently affirmed by the Italian Supreme Administrative Court (Consiglio di Stato) with judgement n. 8092 of 20 October 2025, as a key case study. This ruling, the first in Italy to address the use of generative AI (ChatGPT-4) in a technical bid, serves as a lens to examine the adequacy of traditional legal frameworks. The analysis critiques the Court's decision to dismiss the appeal by relying on the established principles of the evaluation committee's broad technical discretion and the limited judicial review of assessments made using the Analytic Hierarchy Process (AHP) method. The author argues that this approach creates a "double shield", making effective judicial review of the technical substance of AI solutions nearly impossible and leading to a "deference

paradox".

Summary: 1. Introduction.- 2. The Facts of the Case and the Applicant's Objections Regarding AI.- 3. The Lazio Regional Administrative Court's Reasoning: Balancing Technical Discretion and Artificial Intelligence.- 4. Critical Issues and Points for Reflections.- 4.1. The Adequacy of Traditional Judicial Review Canons for AI.- 4.2. The Scrutiny of AI-Based Offers by the Contracting Authority.- 4.3. A Proactive Solution: Imposing the AI Act's High-Risk Regime via Lex Specialis to Overcome Technical Discretion.- 4.4. Evidentiary Challenges Revisited: the Burden of Proof and the Principle of Availability in the Algorithmic "Black Box".- 4.5. The Risks of Premature Normalization of Complex Technologies.- 5. Conclusions.

1. Introduction

The landscape of public procurement is undergoing a full-scale revolution, a wave sweeping away old shores. The cause? Artificial intelligence. Private companies, in fact, have begun to include AI-based solutions in their bids to secure contracts with public administrations. This can only translate into an earthquake shaking the foundations of evaluation processes, introducing new and complex challenges. Consequently, it also severely tests the judiciary, whose judicial review is still in an embryonic phase when confronting these implications[1].

The point is clear: the speed with which AI, especially generative AI, evolves and infiltrates the economic and social fabric, seems to far outpace the adaptability of our consolidated administrative practices and our legal framework. The result is a gap that judicial pronouncements, like the one we will analyze, seek to bridge, often, however, having to resort to interpretative tools and legal categories born in another era. Legal scholarship, for its part, has begun to explore this *mare magnum*, developing an important distinction: AI can be an auxiliary tool for procurement procedures, or it can be the object of the public procurement itself[2].

In this dynamic context, judicial review over evaluation methods and the use of AI in public procurement is still in what we might call an embryonic phase.

The Lazio Regional Administrative Court (TAR Lazio), with its judgment of March 3, 2025, n. 4546, marked a turning point. Issued in the dispute between Romeo Gestioni S.p.A. and Consip S.p.A. (with Dussmann Service S.r.l. as counter-interested party), this indeed represents the first instance where an Italian administrative judge was called upon to directly address the challenge of a technical offer that declared the use of advanced generative AI tools, namely ChatGPT-4. This ruling was appealed and subsequently affirmed in its entirety by the Italian Supreme Administrative Court (*Consiglio di Stato*) with judgment n. 8092 of October 20, 2025[3]. The ruling, which has already generated considerable discussion among practitioners, is innovative because, in principle, it recognized that AI can be used in the execution of public contracts[4].
But what does this mean for the law? And what questions does it raise?
Our objective is to critically analyze the Lazio TAR's approach, as consolidated by the *Consiglio di Stato*, with particular attention to the grounds of appeal concerning the successful bidder's use of AI. We will question what evaluation standards an examining board should apply. And, above all, what are the limits and nature of the judicial review that the administrative judge can exercise in the face of such innovative and technologically complex elements.
The fundamental question is this: are our traditional legal instruments, born in an analog era, still sufficient to guarantee a just balance? Can they promote technological innovation without sacrificing sacred principles such as transparency, competition, and the fairness of administrative action?

2. The Facts of the Case and the Applicant's Objections Regarding AI

The dispute under examination originates from a tender procedure initiated by Consip S.p.A. for the stipulation of a framework agreement, pursuant to Article 59, paragraph 4, letter a), of Legislative Decree no. 36/2023, concerning cleaning and sanitization services for National Health Service Entities. Lot no. 2, the subject of the dispute, specifically concerned properties located in the Umbria Region. Romeo Gestioni S.p.A., having ranked third in the classification for the said lot, challenged the award granted in favor of Dussmann Service S.r.l. (first ranked) and the temporary grouping of undertakings led by Samsic Italia S.p.A.

(second ranked)[5].

Among the numerous grounds of appeal, those of specific interest for this contribution were articulated through additional grounds and focused on the technical offer submitted by Dussmann Service S.r.l. In particular, Romeo Gestioni challenged the attribution of qualitative scores deemed excessively high (for a total of 16 points) in relation to Dussmann's declared intention to utilize artificial intelligence tools, specifically OpenAI's Chat GPT-4. Such use was planned for the execution of activities attributable to various technical offer evaluation criteria, namely: B1 «*Organizational Model*», C1 «*Technical-Operational Methodologies for Cleaning and Sanitization Services*», C2 «*Logics and Modalities for Service Personalization*», C3 «*Implementation and Management of the Information System and Contact Center*», and D1 «*Improvement Proposals for Service Quality Control*», as detailed in Table 10 of the tender specifications[6].

The main arguments put forward by the applicant Romeo Gestioni in support of these objections are threefold.

The first is that the description of AI's use in Dussmann's offer was characterized by excessive vagueness and abstractness. The use of «*linguaggio estremamente tecnico, talvolta perfino criptico*» would have, according to the applicant, masked the presentation of «*modelli astratti, la cui funzionalità in concreto è tutta da dimostrare*»[7]. The problem would therefore relate to the transparency and concrete evaluability of the proposed AI solutions, issues widely debated in legal scholarship with reference to the sometimes opaque nature of such systems[8].

The second argument is that the Evaluation Committee «*avrebbe accolto positivamente, senza alcun approfondimento istruttorio, l'utilizzabilità dell'IA nell'ambito del servizio di cui si discute*», thus emphasizing the obligations of due diligence and the depth of investigation that contracting authorities are required to conduct when faced with offers incorporating innovative and complex technologies such as AI[9].

Finally, Romeo Gestioni undertook a peculiar initiative, submitting in court the results of its own «*queries*» performed on the ChatGPT tool. The intent was to demonstrate, through the obtained responses, a presumed incompatibility between the actual capabilities of the language model and the specific use Dussmann intended for it within the scope of the tender. An original attempt,

certainly. But one that, at the same time, highlights the significant evidentiary difficulties encountered when challenging claims regarding the capabilities and adequacy of an AI system in litigation. Let's reflect: how can one disprove or confirm the performance of such a complex technology, which varies in its outputs[10], without direct access to the AI system, its training data, or objective and shared benchmark parameters? The judge's reaction to this initiative, as will be seen, is a clear signal of the current limitations of our system.

3. The Lazio Regional Administrative Court's Reasoning: Balancing Technical Discretion and Artificial Intelligence

In response to these objections, the Lazio Regional Administrative Court adopted a clear stance: it upheld the contracting authority's actions.

Specifically, the Lazio Regional Administrative Court rejected the additional grounds of appeal, including the objection concerning the use of AI, deeming them manifestly unfounded[11]. But on what basis was this decision grounded? The Court structured its reasoning on two fundamental pillars of administrative jurisprudence governing the evaluation of technical offers.

The first pillar is the reference to the consolidated principle of technical discretion recognized to the examining board. The Lazio Regional Administrative Court emphatically reiterated that «*l'attribuzione dei punteggi rientra nell'ampia discrezionalità tecnica riconosciuta alla commissione giudicatrice, organo tecnico competente*»[12]. Such discretion, despite having undergone an evolution in its review by the administrative judge – moving from a merely extrinsic control over the logical consistency of the process to a more penetrating verification of the reliability of technical evaluations[13] – remains a bulwark against a substitutive review. Indeed, the administrative judge can only challenge technical evaluations in the presence of macroscopic defects such as the abnormality of the technical choice made, manifest illogicality, unreasonableness, arbitrariness, or clear factual error, being precluded from independently verifying the adequacy of the offer or substituting its own judgment for that of the technical body[14]. Objections that «*impingono nel merito di valutazione per loro natura opinabili*» were, therefore, declared inadmissible[15].

The second pillar upon which the TAR's decision rests is the Analytic Hierarchy

Process (AHP) evaluation method. This method, stipulated by the *lex specialis* for the attribution of discretionary scores, was developed by Thomas L. Saaty in the 1970s and is a systematic methodology for making complex decisions, combining qualitative and quantitative techniques: in practice, it decomposes a decision problem into a hierarchical structure of interrelated elements, typically organized as objectives, criteria, sub-criteria, and alternatives[16]. The Court highlighted how, according to consolidated jurisprudence and particularly the teachings of the *Adunanza Plenaria* of the *Consiglio di Stato*, the adoption of this evaluation method further restricts the scope of judicial review[17]. In such a system, the justification for the scores awarded is considered inherent «*nelle stesse preferenze attribuite ai singoli elementi di valutazione considerati nei raffronti con gli stessi elementi delle altre offerte*»[18]. What does this mean for the judge? It means that their review is admissible only if there is a «*uso distorto, logicamente incongruo, irrazionale del metodo in parola*», the burden of proof for which lies entirely with the applicant, who cannot merely challenge the non-shareability of the comparative judgment expressed by the Committee but must demonstrate that it is unreliable[19].

And the evidence produced by Romeo Gestioni, those «*queries*» to ChatGPT? The Lazio Regional Administrative Court dismissed them without hesitation, defining them as «*valutazioni unilaterali del tutto opinabili*», resulting from a «*lettura fuorviante, errata e parziale sia dell'offerta tecnica di Dussmann, sia dei criteri di valutazione previsti dalla lex specialis*»[20]. This stance is a strong signal. It denotes marked judicial skepticism towards attempts by litigating parties to informally "test" AI tools. It is somewhat as if the judge were saying: «*You cannot bring the results of your conversations with artificial intelligence to court and claim they are definitive proof of unreliability*». This approach, ultimately, led the Court to consider Dussmann's AI proposal as a «*mirato e specifico*» use of artificial intelligence, «*diverso dall'utilizzo del modello generale descritto nell'atto di motivi aggiunti*». In the TAR's view, AI was proposed by Dussmann as an «*ulteriore strumento di supporto matematico/statistico e di elaborazione di dati, migliorando l'efficienza e la qualità dei servizi offerti*»[21], thus making the proposal neither vague nor abstract, but sufficiently defined and plausible.

Finally, the judges, albeit in *obiter dictum*, made an observation destined to spark debate, stating that the AI tool proposed (with implicit reference to large

language models like ChatGPT) would be «*ormai di comune e diffuso utilizzo*».[22] While this may seem an obvious observation in the digital age, this statement opens the door to much deeper reflections.

4. Critical Issues and Points for Reflections

The Lazio Regional Administrative Court's approach, while anchored in consolidated jurisprudential principles, raises a series of crucial questions regarding the adequacy of our traditional legal instruments in the face of the advent of artificial intelligence in public procurement. In my opinion, at least five critical issues emerge from this judgment, which require our careful attention: the inherent inadequacy of traditional judicial review canons; the depth of scrutiny of AI-based offers required by contracting authorities; a proactive strategy to overcome the limits of technical discretions; the procedural challenges related to the burden of proof; and the risks associated with the premature normalization of complex technologies like AI.

4.1. The Adequacy of Traditional Judicial Review Canons for AI

The first critical issue concerns the adequacy of traditional judicial review canons for AI. The Lazio TAR's judgment, as we have seen, rigorously applies consolidated doctrines on technical discretion and the limits of judicial review, especially in the presence of the AHP method. But here a question naturally arises: are these canons, forged in a technologically different era, truly fully adequate to govern new issues such as the complexity, opacity – the "black box" – and the potential for «*algorithmic bias*»[23] that characterize some artificial intelligence solutions?

Let us imagine a "black box": it is an AI system whose internal mechanisms leading to a specific result we cannot fully comprehend. It is like a clock without visible gears, which tells the time but does not allow us to understand how. This opaque nature makes it extremely difficult for the judge to identify the traditional symptomatic defects of abuse of power. The concrete risk is that technical discretion transforms into a kind of screen, almost a smokescreen, behind which the opacity of AI becomes substantially unchallengeable. And this,

in my opinion, could constitute a true "deference paradox": the more complex and opaque the technology, the less effective, in fact, judicial review becomes[24]. It is as if, faced with an increasingly sophisticated machine, the engineer who should control it has fewer and fewer tools to do so.

The deferential stance was not only the cornerstone of the first-instance decision but was subsequently embraced by the *Consiglio di Stato* in its appellate ruling, which explicitly found the challenge unfounded precisely because it «*sottintende una non consentita sostituzione del convincimento opinabile e personale dell'appellante rispetto all'attribuzione dei punteggi effettuata dalla Commissione giudicatrice, nell'esercizio del proprio potere tecnico-discrezionale*»[25].

This intrinsic opacity of many AI systems directly clashes with core principles of Italian administrative law: administrative transparency and the obligation to provide reasons for administrative acts, enshrined in Law 241/1990[26]. If the administrative judge were to show excessive deference in these circumstances, it would risk compromising citizens' trust and the effectiveness of public accountability. This is particularly true for high-risk AI systems, for which the European[27] and national[28] legislator has provided reinforced safeguards. The almost unconditional reliance on the technical discretion of the examining board could, ultimately, translate into a merely formal scrutiny. And the situation becomes even more complicated if the examining boards lack adequate specialized AI competencies[29].

4.2. The Scrutiny of AI-Based Offers by the Contracting Authority

The second critical issue brings us straight to the heart of the problem: the role of the contracting authority in scrutinizing AI-based offers. The Lazio Regional Administrative Court's judgment, by valuing Dussmann's targeted and specific presentation of its AI solution[30], seems to suggest a path. A path where a well-argued and plausible AI proposal can pass the contracting authority's review and, consequently, withstand challenges in judicial proceedings.

Consiglio di Stato's reasoning adds a new layer of concern. Rather than engaging with the technical merits of AI, it dismantled the appeal on a preliminary logical ground, affirming that the score awarded to Dussmann was based on multiple

factors, not solely on its use of AI. It stated that «*la Commissione ha attribuito i punteggi prendendo in considerazione plurimi elementi e non solo l'utilizzo dell'intelligenza artificiale*»[31]. This finding, in the court's view, would have «*smentito* per tabulas *il fondamento logico su cui poggiano le censure dell'appellante*»[32].

This line of reasoning reinforces what we might term the "double shield". By focusing on the overall scoring logic and finding a procedural flaw in the appellant's argument (i.e., that AI was not the sole factor), the court relieved itself of the duty to conduct a substantive technical analysis. The first shield (technical discretion) is thus protected by a second shield (procedural formalism), creating a nearly impenetrable defense against challenges to AI-based bids.

But this should make us reflect. How in-depth must the preliminary investigation conducted by contracting authorities be? Is a mere "plausibility" check of the AI solution sufficient, or is a true "technical deep dive" necessary? We must ask whether it is incumbent to analyze the system's architecture, its transparency, its ability to explain its outputs (the so-called "Explainable AI" or "XAI"[33]), its accuracy, robustness, cybersecurity[34], proper data management (both data used for training and data processed during the AI system's operation), and so on.

This question becomes even more pressing when we talk about evaluation methodologies like the Analytic Hierarchy Process (AHP). AHP, which focuses on determining relative preferences among different options based on predefined criteria, might not be intrinsically equipped to evaluate the absolute adequacy, intrinsic technical validity, or ethical implications of a particularly innovative and complex AI solution. AHP, in this sense, assumes that evaluators have a sufficiently homogeneous understanding of the elements being compared. But this assumption, given the specificity and novelty of AI systems, might falter. If the comparison criteria have not been meticulously designed to probe these very particular aspects, the risk is high. The scientific literature on AHP and its evolution (such as the Analytic Network Process or ANP[35]) highlights its usefulness for complex problems, but also the essential need for a correct structuring of the decision model and criteria[36]. Ultimately, the danger is that, in the absence of specific criteria for AI, AHP might end up masking a true lack of substantial evaluation of the technological component.

The TAR's decision (as confirmed by *Consiglio di Stato*), in my opinion, risks, perhaps unintentionally, encouraging a dual hazard. On the one hand, it could prompt bidders to present AI solutions in an apparently «*mirato e specifico*» but actually superficial manner, merely to pass a plausibility check: metaphorically, it is like describing a product with grandiloquent terms without delving into the details of its actual functioning. On the other hand, the judgment might legitimize the use of a tool like AHP which, without adequate engineering of evaluation criteria, risks "armoring" the commission's discretion, without, however, guaranteeing genuine technical substance: metaphorically, it is like having a seal of approval on a product, but without the content having been effectively verified.

It therefore becomes fundamental to reflect on how tender documents (invitations to tender, technical specifications, and disciplinary rules) should be drafted. They must be written in such a way as to allow for a serious, transparent, and comparative evaluation of AI components, while simultaneously ensuring equal treatment among competitors and the absence of discrimination. In this regard, the integration of standard contractual clauses, such as those proposed at the European level for AI procurement ("MCC-AI"), could prove to be a valuable tool in the bid evaluation phase. These clauses would, in fact, provide an objective evaluation framework. They would allow administrations to verify the conformity of suppliers' offers with crucial requirements: transparency, robustness, accuracy, cybersecurity, risk management, data governance, and human oversight. In this way, the evaluation would go far beyond the mere plausibility of the proposal. But that's not all. The MCC-AI would also be useful after the contract award, becoming an integral part of the contract. They would oblige the supplier to keep technical documentation and the risk management system updated, to provide all necessary information to demonstrate the system's compliance, to undertake immediate corrective actions in case of non-compliance, and to guarantee the traceability of system operations through automatic event logging ("log")[37].

To conclude on the first two critical issues, we can affirm that the combined effect of two factors – a deferential approach to technical discretion and the application of the AHP method – creates what we called a "double shield," which makes it incredibly arduous for applicants to successfully challenge the

technical evaluations of the commission, especially when dealing with complex and innovative technologies like AI. Technical discretion, by itself, confines judicial review to only the most evident defects (so-called "macroscopic defects"), preventing the judge from substituting their own evaluation for that, eminently technical, of the commission. The AHP method, furthermore, as interpreted by Italian administrative jurisprudence, further radicalizes this limitation. Why? Because the justification for scores, in this system, is considered intrinsic to the comparative judgments expressed: a detailed explanation for each single point is not required; the score itself is the justification[38]. In this context, the burden of proof for anyone intending to challenge an AI-based evaluation becomes extremely heavy. It is not sufficient to demonstrate that the evaluation of the AI component is debatable or even manifestly illogical. On the contrary, it is necessary to prove that such illogicality derives from a patently distorted application of the AHP method. This "double shield" risks precluding an in-depth examination of the technical substance of the AI proposal, especially if the commission does not clearly explain how the individual aspects of the AI influenced the comparative judgments within the AHP matrix. A situation which, in my opinion, could incentivize contracting authorities to resort to methods like AHP to further "shield" their technical evaluations, raising serious questions about the actual transparency and true consistency of the evaluation process in the face of cutting-edge technological innovations.

4.3. A Proactive Solution: Imposing the AI Act's High-Risk Regime via *Lex Specialis* to Overcome Technical Discretion

As to the third critical issue, we can proactively address the limits of technical discretion, showing that the "deference paradox" is not an insurmountable obstacle. In particular, a legally sound solution is already available to contracting authorities, allowing them to preemptively resolve the conflict between technological complexity and judicial review. The solution is to strategically use their administrative power to define the *lex specialis* for the tender procedure.
Italian public procurement law grants contracting authorities broad discretion in establishing the special requirements for participation in a tender. This power, now codified in Article 100 of the Italian Legislative Decree 36/2023 (*codice dei*

contratti pubblici), allows them to introduce technical and professional requirements that are more stringent and rigorous than the minimums established by the law, provided that they remain proportionate, reasonable, and pertinent to the specific object of the contract[39].

In the context of emerging and potentially high-impact technologies like AI, this faculty can be construed as a duty stemming from the overarching principles that govern public procurement. The new Italian Public Contracts Code elevates principles such as the "principle of the result" (Article 1) and the "principle of trust" (Article 2) to primary criteria for exercising discretionary power[40]. Pursuing the best possible outcome in terms of quality and efficiency, while safeguarding the public interest, requires a prudent and forward-looking approach. In a sector as sensitive as AI, imposing higher standards of reliability and transparency is a direct implementation of these principles, ensuring the administration acquires solutions that are not only innovative but also robust and accountable.

The proposed strategy consists of contracting authorities explicitly incorporating, within the tender documents (which constitute the *lex specialis*), the obligation for any proposed AI systems to comply with the requirements laid down for high-risk systems in Chapter III, Section 2 of Regulation (EU) 2024/1689 (the AI Act). This imposition would apply regardless of whether the specific AI application falls within the typified high-risk categories listed in Annex III of the Regulation. Specifically, the *lex specialis* would mandate compliance with the core technical and ethical safeguards of the AI Act, transforming them into binding contractual requirements. These requirements would include, for instance, Article 9, which mandates a continuous and iterative risk management system throughout the AI's lifecycle to identify, analyze, and mitigate foreseeable risks to health, safety, or fundamental rights[41]. The framework is complemented by Article 10, which requires that the training, validation, and testing datasets be of high quality (relevant, representative, and as free of errors and biases as possible) and subject to appropriate data governance practices[42]. Furthermore, to ensure accountability and verifiability, Article 11 obliges the provider to create and maintain comprehensive technical documentation demonstrating compliance with all requirements[43], while Article 12 requires the system to have the technical capability to automatically and

traceably records events ("logs") throughout its operational life[44]. Moreover, Article 13 demands that the AI systems be designed for sufficient transparency, enabling users to interpret its output and use it appropriately, and necessitates the provision of concise, complete, and clear instructions for use detailing the system's capabilities, limitations, expected accuracy, and foreseeable risks[45]. The mandate would also cover Article 14 (human oversight), ensuring that the system is designed to allow for effective oversight by natural persons, who must be able to understand the system's capacities, monitor its operation, remain aware of automation bias, and possess the ability to disregard, override, or interrupt the system's output through a "stop" button or similar procedure[46]. Additionally, compliance with Article 15 would be obligatory, requiring the system to be designed to achieve an appropriate and declared level of accuracy, to be resilient against errors, and to be protected against attempts by unauthorized third parties to alter its use or performance by exploiting vulnerabilities[47].

By referencing these normative standards, the contracting authority would provide an objective and verifiable framework for the evaluation of AI solutions, moving far beyond a generic assessment of plausibility. This strategic incorporation of legal standards into the tender documents triggers a fundamental shift in the nature of the evaluation and, consequently, in the nature of any potential defect. The assessment conducted by the evaluation committee would no longer be a purely technical-discretionary judgement on the merit of a technological solution. Instead, it would become a verification of compliance with predetermined legal requirements set forth in the *lex specialis*.

This transformation would also have a profound impact on the scope and intensity of judicial review. A challenge against the evaluation of the AI component would no longer be limited to the narrow grounds of abuse of power (*eccesso di potere*). On the contrary, the challenge would be grounded in a violation of law (*violazione di legge*), specifically the violation of the *lex specialis*, which serves as the binding law for that particular procurement procedure. The administrative court's role would thereby be fundamentally altered: it would no longer be required to defer to the committee's technical expertise, a deference that creates what we called the "paradox" when faced with opaque systems. Instead, the judge would be empowered to conduct a full, strong, and intrinsic review of legal requirements. The central question for the court would shift from

the subjective ("Was the committee's technical judgement reasonable?") to the objective ("Did the proposed AI system demonstrably meet the legal requirements for transparency, human oversight, accuracy, robustness, and cybersecurity as mandated by the tender documents?").
In my opinion, this approach would effectively pierce what we previously referred to as the "double shield" of technical discretion and the AHP method. The technical judgment would be restrained by objective legal parameters, making the committee's assessment fully and effectively justiciable. In other words, it provides an *ex ante* solution that empowers public bodies to manage AI risks, enhances accountability, and aligns procurement practices with the broader EU policy goal of promoting trustworthy AI, without waiting for judges to develop new (and potentially inconsistent) standards of review *ex post*.

4.4. Evidentiary Challenges Revisited: the Burden of Proof and the Principle of Availability in the Algorithmic "Black Box"

The fourth critical issue concerns the need to "test" artificial intelligence from an evidentiary perspective. The Lazio Regional Administrative Court's rejection of Romeo Gestioni's attempt to use its own ChatGPT queries as evidence[48] highlights a fundamental question: how can a plaintiff effectively challenge in court the declared capabilities, or the alleged inadequacy, of an AI system proposed by a competitor in a public tender? The rejection of Romeo's evidence, without the Court indicating an alternative path for verifying such technical claims, leaves an enormous void on how such challenges can be validly brought to the judge's attention and properly investigated. Also, *Consiglio di Stato*'s ruling has affirmed the lower court's position by dismissing the appellant's expert opinion, reasoning that once the logical foundation of the appeal was found to be flawed (as AI was not the sole scoring factor), the expert's technical observations became irrelevant[49]. The judiciary's position is now very clear: not only are informal tests by a plaintiff inadmissible, but even formal expert opinions can be rendered irrelevant on procedural grounds before their technical substance is ever considered.
The traditional principle of burden of proof, codified in Article 2697 of the

Italian Civil Code (*onus probandi incumbit ei qui dicit*), places the burden squarely on the party making an assertion. In the context of administrative litigation, this principle is nuanced by the "*principio dispositivo con metodo acquisitivo*", which allows the judge certain official powers to acquire evidence[50]. However, this tempting is insufficient in the face of the algorithmic black-box. The key lies in a particular interpretation of Article 64 of the Code of Administrative Court Procedure, whose Paragraph 1 states that parties have the burden of providing evidence that is in their "*disponibilità*" (availability). This term cannot be interpreted statically, because, in the context of a proprietary AI system, the crucial evidence (the source code, the architecture, the training and validation datasets, the logic of the algorithm, the results of internal testing, etc.) is, by definition, not in the plaintiff's availability. Thus, it is exclusively and entirely available to the successful bidder who developed or proposed the system.
To address such structural imbalances, Italian jurisprudence, across civil and administrative courts, has developed the principle of *vicinanza della prova* (proximity of evidence). This principle, rooted in the constitutional right to an effective defense (Article 24 of the Constitution) and the principle of a fair trial (Article 111), acts as a necessary corrective to the formal application of Article 2697 of the Italian Civil Code. It dictates that the burden of proof should be allocated to the party who is closer to the source of the evidence and for whom providing proof is materially easier[51].
The principle of *vicinanza della prova* provides the doctrinal foundation for interpreting the concept of *disponibilità* in Article 64 of the Italian Code of Administrative Court Procedure in the context of AI litigation. In a dispute over an AI system's capabilities, the successful bidder is unequivocally the party closest to the evidence. To place the full burden of proving a negative (i.e., that the system cannot do what is claimed) on the plaintiff, who has no access to the system's inner workings, would be to impose a *probatio diabolica* and effectively nullify their right to challenge the award[52].
Therefore, a correct application of these principles requires a dynamic rebalancing of the evidentiary burden. The plaintiff should be required to meet a threshold of providing a *principio di prova*, that is a serious evidentiary starting point that raises credible doubts about the AI system's functionality or compliance. This could be achieved by presenting expert opinions questioning

the technical plausibility of the bidder's claims, highlighting manifest contradictions within the technical offer, or providing documented evidence of failures in similar systems.

Once the plaintiff has met this initial burden, the principle of *vicinanza della prova* should trigger a reversal, shifting the onus onto the defendant (the successful bidder, and by extension the contracting authority that approved the bid) to provide positive, concrete, and verifiable proof that the AI system functions as declared and meets all the requirements of the *lex specialis*. This procedural rebalancing, in my opinion, is the only way to overcome the structural information asymmetry. It also creates a powerful *ex post* incentive for bidders to be truthful and transparent in their offers, knowing they may be compelled to substantiate their claims with hard evidence in court.

In this scenario, a procedural tool emerges strongly, which, though not new to administrative justice, becomes the most appropriate, if not essential: the court-appointed technical expert (*consulente tecnico d'ufficio* - CTU)[53]. But caution is advised: the effective use of this tool in high-tech disputes requires deep reflection on the expert's profile and mandate. It is no longer enough to appoint a generic IT expert. The specificities of modern AI systems, and particularly Large Language Models (LLMs), necessitate the identification of a new type of CTU, which we might define as the "expert of the new millennium". Consequently, the formulation of the questions by the judge becomes the key procedural act, as it must guide the court-appointed technical expert (CTU) to verify the precise correspondence between the performance promised in the technical offer and the actual capabilities of the AI system. For example, the question could ask to ascertain, through specific tests and analyses, whether the AI solution is effectively capable of performing the described tasks, with what margin of error, on what data basis it was trained, and whether the declared risk mitigation methodologies have been concretely implemented. In this way, the CTU would act as a conceptual bridge between the technical-IT world and the legal-administrative one, translating factual findings into elements evaluable by the judge according to the traditional categories of abuse of power, such as manifest illogicality, unreasonableness, or misrepresentation of factual premises. The CTU's report, therefore, would not substitute the commission's evaluation. Rather, it would provide the judge with the indispensable cognitive tools to

verify whether such evaluation falls within one of those symptomatic defects that legitimize its annulment, even when the technology involved is complex and opaque in nature. However, the *Consiglio di Stato*'s ruling makes it clear that without a proactive judicial embrace of these tools, the courthouse doors will remain closed to meaningful challenges against AI systems.

4.5. The Risks of Premature Normalization of Complex Technologies

The fifth and final critical issue, deserving careful examination, concerns an incidental statement by the TAR, almost an "aside" in the judgment, but of considerable weight. The Court maintained that AI (or at least the type of tool proposed, comparable to ChatGPT) is «*ormai di comune e diffuso utilizzo*»[54]. Now, while it is undeniable that Large Language Models (LLMs) have experienced rapid proliferation and become accessible to many[55], it is equally true that their mature and responsible integration into the public sector, especially in critical or high-risk areas, is still in an initial phase. And it presents non-negligible complexities[56]. Qualifying such powerful tools as "common" could, unintentionally, diminish their novelty and potential risks, as it reduces the perceived need for particularly in-depth analysis, both by tender committees and, subsequently, by the administrative judges. Prematurely normalizing AI, especially complex systems like LLMs, risks making us overlook their intrinsic limitations. And here the catalog is long: consider the so-called "hallucinations"[57], meaning the ability of LLMs to generate plausible but factually incorrect or nonsensical information. Imagine an AI system that, in a public procurement contract for critical infrastructure management, "hallucinates" crucial data: the consequences could be disastrous. Or consider the "biases" present in training data, which can lead to discrimination and perpetuate social inequalities[58]. And further, the lack of a real understanding of context, which means that LLMs, while excellent at mimicking language, do not possess true intelligence[59]. The risk, in summary, is a premature reduction of vigilance against potential negative consequences. *Consiglio di Stato*'s decision, by endorsing its analyzed approach, exacerbates this risk. If contracting authorities and judges were to treat these tools as ordinary technologies, without implementing the necessary safeguards

required for their responsible use, we could find ourselves facing a pitfall far greater than innovation itself.

The regulatory framework, both at national and supranational levels, is undergoing rapid evolution. At the national level, we have Italian Legislative Decree 36/2023[60] and the recent Law No. 132 of 23 September 2025[61]. At the European level, the EU Artificial Intelligence Regulation, the so-called "AI Act"[62], stands out. It will be crucial to monitor how these new regulations, once fully operational and interpreted by jurisprudence and practice, will provide clearer and more specific guidance for the integration of AI in public procurement, seeking to fill existing gaps and address current uncertainties. In particular, Article 30 of Italian Legislative Decree 36/2023, which explicitly allows for the use of automated procedures – including AI – in bid evaluation, requires that such procedures ensure the knowability (transparency) and comprehensibility of the decision-making process, as well as the non-exclusivity of the algorithmic decision (by providing for human oversight). The interpretation and concrete application of these requirements will constitute fertile ground for future doctrinal and jurisprudential debate.

5. Conclusions

The Lazio Regional Administrative Court's judgment n. 4546/2025, as confirmed by the ruling of the *Consiglio di Stato* n. 8092/2025, is, undoubtedly, an inescapable benchmark. However, as often happens in law when confronting innovation, it is an ambivalent judgment. On the one hand, it offers the stability of a secure approach, anchored to the consolidated principles of technical discretion. On the other hand, however, it reveals the fragility of such an approach in the face of the rising tide of technological innovation: it is like an ancient dam facing a tsunami.

The analysis we have conducted has sought to demonstrate a crucial point: the application of what we have termed the "double shield" – the combination of a deferential judicial review and structured evaluation methodologies like AHP – risks transforming technical discretion into an area of substantial unreviewability. And this occurs precisely when the opacity and complexity of AI would, on the contrary, require a much more penetrating scrutiny.

The Romeo Gestioni case, ultimately, is a true litmus test. It has highlighted not only the limits of what currently exists, but also the pressing need to chart new paths. This analysis has proposed a two-part strategy to address the challenges identified.

First, an *ex ante*, substantive solution is available to contracting authorities. By strategically using their power to define the *lex specialis*, public bodies can proactively incorporate the requirements for high-risk systems from the EU AI Act as binding criteria for all AI solutions. This act of "legalization" transforms the nature of the evaluation from a review of the technical discretion to a review of the legal requirements, thereby empowering the administrative judge to conduct a strong and intrinsic assessment of compliance and fostering what we might call an "accountability by design".

Second, an *ex post*, procedural solution is available to the judiciary. Through a sophisticated application of Article 64 of the Italian Code of Administrative Court Procedure, informed by the principle of *vicinanza della prova*, judges can rebalance the evidentiary burden in litigation. This approach overcomes the profound information asymmetry inherent in the algorithmic black box, compelling the party closest to the evidence to demonstrate its claims and enabling a meaningful evidentiary inquiry, supported by specialized court-appointed experts.

These two strategies should provide a robust legal framework to govern AI in public procurement, equipping all stakeholders in the system.

First actor: the legislator, who is called upon to translate the general principles of the AI Act into operational rules for procurement, giving flesh and blood to abstract concepts.

Second actor: contracting authorities, who can no longer be mere passive recipients of technological solutions, but must acquire new competencies, a true «*AI procurement literacy*»[63].

And finally, the most challenging actor: administrative judges. They faces the most arduous challenge: to innovate its own instruments, to refine legal categories and investigative means, in order to ensure that judicial review does not stop at the threshold of the AI "black box" (and here we think, again, of the crucial role of a specialized court-appointed technical expert).

The balance between promoting innovation and safeguarding the core principles

of administrative action will represent the fundamental challenge for the future. The true pitfall, this ruling suggests, is not so much the use of AI itself – which is a given and an opportunity – but rather the risk of confronting it with legal categories that are formally correct but substantially inadequate to govern its disruptive novelty. If we fall into this trap, we will betray the ultimate mission of administrative justice: to ensure full and effective protection of public interest and fundamental rights in the digital era.

1. In the USA, for example, a part of legal scholarship has put forward an interesting proposal for the establishment of a «*Tech Court*» specialized in cases involving emerging technologies within the federal judiciary under Article III of the U.S. Constitution. Such a court, according to the proponent, would increase uniformity and predictability, promote judicial efficiency, and prevent «*forum shopping*». For this proposal, see M.B. Neitz, *Ready or Not: How Congressional Dysfunction and Loper Bright Enterprises v. Raimondo will Shift U.S. Regulation of Emerging Technologies to the Federal Bench*, in *SMU Law Review*, 78(1), 2024, pp. 119-156, esp. 144-153.
2. On this distinction, see G.F. Licata, *Intelligenza artificiale e contratti pubblici: problemi e prospettive*, in *CERIDAP*, 2, 2024, pp. 36-37.
3. Cons. St. (sec. III), judgment of 20 October 2025, n. 8092.
4. R. Damonte, M. Bersi, *L'intelligenza artificiale e gli appalti pubblici*, in *italiAppalti*, 30 April 2025, p. 4; L. Di Giacomo, *L'intelligenza artificiale negli appalti pubblici: la sentenza TAR Lazio n. 4546/2025,* in *Diritto.it*, *2* May 2025; G. Di Stefano, A.A. Sessa, *ChatGPT nelle gare pubbliche: il TAR Lazio legittima l'impiego dell'IA generativa [sentenza n. 4546/2025]*, in *Diritto di Internet*, 15 May 2025; M. Briccarello, *Chat GPT vince la gara e il giudizio*, in *italiAppalti*, 20 May 2025.
5. T.A.R. Lazio-Roma (sec. II), judgment of 3 March 2025, n. 4546, § 1.
6. *Ibid*., §§ 4, 14.
7. *Ibid*., § 4.
8. On the topics of algorithmic opacity, transparency, and the "black box" in public administration, see, *inter alia*, A. Tronci, *Le nuove frontiere della trasparenza nell'amministrazione algoritmica*, in *Federalismi.it*, 9, 2025, pp. 141-167; L. Martucci, *Il fragile equilibrio tra l'intelligenza artificiale e la trasparenza nelle decisioni amministrative*, in *Amministrativamente*, 1, 2025, pp. 402-412; E. Tagliasacchi, *Il sindacato del giudice sull'attività amministrativa algoritmica tra etica benthamiana e "black box": il futuro ha un cuore antico?*, in *Contratto e impresa*, 1, 2025, pp. 26-33; M. Ramajoli, *La convivenza tra trasparenza e riservatezza*, in *Diritto amministrativo*, 2, 2024, pp. 471-495; S. Foà, *Intelligenza artificiale e cultura della trasparenza amministrativa. Dalle "scatole nere" alla "casa di vetro"?*, in *Diritto amministrativo*, 3, 2023, pp. 515-548; A. Corrado, *La trasparenza necessaria per infondere fiducia in una amministrazione*

algoritmica e antropocentrica, in *Federalismi.it*, 5, 2023, pp. 175-215; C. Di Lello, *La trasparenza algoritmica del procedimento amministrativo "algoritmico"*, in *Iustitia*, 4, 2023, pp. 48-58; G. Lo Sapio, *La black box: l'esplicabilità delle scelte algoritmiche quale garanzia di buona amministrazione*, in *Federalismi.it*, 16, 2021, pp. 114-127; G. Pesce, *Il giudice amministrativo e la decisione robotizzata. Quando l'algoritmo è opaco*, in *Judicium*, 15 June 2020; A. Mascolo, *Gli algoritmi amministrativi: la sfida della comprensibilità*, in *Giornale di diritto amministrativo*, 3, 2020, pp. 366-375; S. Sassi, *Gli algoritmi nelle decisioni pubbliche tra trasparenza e responsabilità*, in *Analisi Giuridica dell'Economia*, 1, 2019, pp. 109-128;

9. T.A.R. Lazio-Roma, cit., § 4.
10. The variability and unreliability in AI systems stem from multiple sources that fundamentally differentiate them from traditional software. At their core, AI systems – especially those built on machine learning – operate on statistical and probabilistic principles rather than deterministic ones. This characteristic means that «*software 'defects' are not well defined*» in AI systems, requiring a reevaluation of traditional quality management techniques (P. Santhanam, *Quality Management of Machine Learning Systems*, in O. Shehory, E. Farchi, G. Barash (eds.), *Engineering Dependable and Secure Machine Learning Systems*, Springer, Cham, 2020, pp. 1-13). This variability can be categorized into two distinct types of uncertainty. «*Aleatoric uncertainty*» arises from random noise and data variability and is considered non-reducible, while «*epistemic uncertainty*» stems from model limitations and training data gaps, which can potentially be reduced through improved modeling strategies or additional data collection (H.B. Braiek, F. Khomh, *Machine Learning Robustness: A Primer*, in M. Lorenzi, M.A. Zuluaga (eds.), *Trustworthy AI in Medical imagining*, Elsevier, London, 2024, available at arXiv, pp. 4-5). Large Language Models (LLMs) exhibit a particularly challenging form of variability known as the "Model Variability Problem" (MVP), characterized by inconsistent sentiment classification, polarization, and uncertainty that arises from stochastic inference mechanisms, prompt sensivity, and biases in training data (D. Herrera-Poyatos, C. Peláez-González, C. Zuheros, A. Herrera-Poyatos, V. Tejedor, F. Herrera, R. Montes, *An overview of model uncertainty and variability in LLM-based sentiment analysis. Challenges, mitigation strategies and the role of explainability*, 6 April 2025, available at arXiv, pp. 1-25). Even simple changes, such as rearranging the word order in prompts, can lead to different outputs, highlighting the models' inherent inconsistency (M. Mock, S. Schmidt, F. Müller, R. Görge, A. Schmitz, E. Haedecke, A. Voss, D. Hecker, M. Poretschkin, *Developing trustworthy AI applications with foundation models*, 8 May 2024, available at Fraunhofer Publica, p. 20). In function-calling scenarios, studies have quantified this reliability issue by identifying «*fluctuating queries*» – inputs that yield different results across multiple attempts – demonstrating how non-determinism affects consistency in practical applications (N. Bhan, S. Gupta, S. Manaswini, R. Baba, N. Yadav, H. Desai, Y. Choudhary, A. Pawar, S. Shrivastava, S. Biswas, *Benchmarking Floworks against OpenAI & Anthropic: A Novel Framework for Enhanced LLM Function*

Calling, 23 October 2024, available at arXiv, p. 8).

11. T.A.R. Lazio-Roma, cit., § 13.
12. *Ibid*., § 14.
13. On the topic of technical discretion and judicial review by administrative judges in Italy, cfr. F. Francario, M.A. Sandulli (a cura di), *Sindacato sulla discrezionalità e ambito del giudizio di cognizione*, Editoriale Scientifica, Napoli, 2023, *passim*, e R. Labriola, *Discrezionalità`tecnica e poteri del giudice amministrativo*, Giuffrè, Milano, 2020, *passim*. See also C. Andreini, *La discrezionalità tecnica: evoluzione storica e avvento della sindacabilità giurisdizionale alla luce del passaggio dallo Stato autoritario allo Stato di diritto*, in *Il diritto amministrativo*, 12 March 2024; F. Liguori, *Dalla discrezionalità tecnica insindacabile alle valutazioni tecniche sostituibili. Indagine sul sindacato giurisdizionale sugli atti delle Autorità Indipendenti*, in *Munus*, 2, 2023, pp. 349-386; P. Carpentieri, *La discrezionalità tecnica e il suo sindacato, da un punto di vista logico*, in *Diritto e società*, 3, 2022, pp. 489-515; F. Orabona, *La discrezionalità tecnica ed il sindacato giurisdizionale sulle valutazioni espresse dalla commissione* esaminatrice, in *Cammino Diritto*, 5, 2018, pp. 48-73; S. Cognetti, *Il controllo giurisdizionale sulla discrezionalità tecnica: indeterminatezza della norma e opinabilità dell'apprezzamento del fatto da sussumere*, in *Diritto e processo amministrativo*, 2-3, 2013, pp. 349-379; M. Allena, *Il sindacato del giudice amministrativo sulle valutazioni tecniche complesse: orientamenti tradizionali versus obblighi internazionali*, in *Diritto processuale amministrativo*, 2012, pp. 1602-1635.
14. *Ibid*., § 10.1, 14. In this sense, the consolidated orientation of the Italian Supreme Administrative Court (*Consiglio di Stato*) is applicable: see, *inter alia*, Cons. St. (sec. VII), judgment of 20 June 2025, no. 5392; Cons. St. (sec. IV), judgment of 20 April 2023, n. 4019; Cons. St. (sec. IV), judgment of 1 March 2022, n. 1445; Cons. St. (sec. IV), judgment of 11 January 2021, n. 359; Cons. St. (sec. III), judgment of 9 December 2020, n. 7758; Cons. St. (sec. III), judgment of 9 June 2020, n. 3694; Cons. St. (sec. IV), judgment of 7 October 2019, n. 6753; Cons. St. (sec. III), judgment of 2 September 2019, n. 6058; Cons. St. (sec. V), judgment of 8 January 2019, n. 173; Cons. St. (sec. III), judgment of 21 November 2018, n. 6572; Cons. St. (sec. III), judgment of 7 March 2014, n. 1072; Cons. St. (Ad. Plen.), judgment of 3 February 2014, n. 8.
15. *Ibid*., § 14.
16. T.L. Saaty, *A Scaling Method for Priorities in Hierarchical Structures*, in *Journal Mathematical Psychology*, 15(3), 1977, pp. 234-281; Id., *The Analytic Hierarchy Process (AHP)*, in *The Journal of the Operational Research Society*, 41, 1980, pp. 1073-1076; Id., *Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World*, Wadsworth Publishing, Belmont (California), 1982, *passim*; Id., *What Is the Analytic Hierarchy Process?*, in G. Mitra (ed.), *Mathematical Models for Decision Support*, Springer, Berlin-Heidelberg, 1988, pp. 109-121; Id., *Decision Making with the Analytic Hierarchy Process*, in *Scientia Iranica*, 9(3), 2002, pp. 215-229). This structured approach simplifies complex decisions by breaking them down into more manageable components

that can be analyzed independently (B.I. Belinda, A.A. Emmanuel, N. Solomon, A.B. Kayode, *Evaluating Software Quality Attributes using Analytic Hierarchy Process, in International Journal of Advanced Computer Science and Applications*, 12(3), 2021, p. 168). The AHP methodology involves four fundamental steps: 1) problem decomposition to form a hierarchy of decision elements; 2) pairwise comparison of elements using a numerical scale (typically 1-9); 3) prioritization to calculate relative weights of elements; 4) synthesis to aggregate these priorities and determine rankings of alternatives (P. Mimović, J. Stanković, V.J. Milić, *Decision-making under uncertainty – the integrated approach of the AHP and Bayesian analysis*, in *Economic Research-Ekonomoska Istraživanja*, 28(1), 2015, pp. 869-870).

17. The reference is to the *Adunanza Plenaria* of the *Consiglio di Stato*, judgment of 14 December 2022, no. 16, which clarified that the AHP method «*[...] non mira ad una ponderazione atomistica di ogni singola offerta rispetto a standard ideali, ma tende ad una graduazione comparativa delle varie proposte dei concorrenti mediante l'attribuzione di coefficienti numerici nell'ambito di ripetuti "confronti a due", con la conseguenza che il sindacato giurisdizionale incontra forti limitazioni e non può sovrapporsi a valutazioni di merito spettanti all'amministrazione, salvi i casi di un uso distorto, logicamente incongruo, irrazionale del metodo in parola, che è, però, preciso onere dell'interessato allegare e dimostrare, evidenziando non già la mera (e fisiologica) non condivisibilità del giudizio comparativo, bensì la sua radicale ed intrinseca inattendibilità tecnica o la sua palese insostenibilità logica*» (§ 30.2.).
18. T.A.R. Lazio-Roma, cit., § 14. In the same vein, see Cons. St. (sec. V), judgment of 29 March 2022, n. 2320; Cons. St. (sec. III), judgment of 29 May 2020, n. 3401; Cons. St. (sec. V), judgment of 27 December 2018, n. 7250; Cons. St. (sec. VI), judgment of 19 June 2017, n. 2969; Cons. St. (sezione III), judgment of 25 June 2016, n. 4364; Cons. St. (sec. V), judgment of 24 October 2016, n. 4415; Cons. St. (sec. III), judgment of 24 April 2015, n. 2050.
19. T.A.R. Lazio-Roma, cit., § 14.
20. T.A.R. Lazio-Roma, cit., § 14.
21. *Ibid*.
22. T.A.R. Lazio-Roma, cit., § 14.
23. On legal issues related to algorithmic bias and discrimination, see, *inter alia*, A. Poama, E. Fosch-Villaronga, *AI bias*, Oxford, Oxford University Press, 2025, *passim*; J. Dailey, *Algorithmic Bias: AI and the Challenge of Modern Employment Practices*, in *UC Law Business Journal*, 21(2), 2025, pp. 215-240; G.F. Lendvai, G. Gosztonyi, *Algorithmic Bias as a Core Legal Dilemma in the Age of Artificial Intelligence: Conceptual Basis and the Current State of Regulation*, in *Laws*, 14(3), 2025, pp. 1-15; J. Ruohonen, *On Algorithmic Fairness and the EU Regulations*, available at arXiv, 13 November 2024; L. Deck, J.L. Müller, C. Braun, D. Zipperling, N. Kühl, *Implications of the AI Act for Non-Discrimination Law and Algorithmic Fairness*, available at arXiv, 26 June 2024, pp. 1-9; H. Weertz, R. Xenidis, F. Tarissan, H.P. Olsen, M. Pechenizkiy, *Algorithmic Unfairness*

through the Lens of EU Non-Discrimination Law: Or Why the Law is not a Decision Tree, in *2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*, June 12-15, 2023, Chicago, IL, USA, pp. 805-816; J. Adams-Prassl, R. Binns, A. Kelly-Lyth, *Directly Discriminatory Algorithms*, in *The Modern Law Review*, 86(1), 2023, pp. 144-175; European Union Agency for Fundamental Rights, *Bias in algorithms – Artificial intelligence and discrimination*, available at FRA website, 8 December 2022, pp. 1-101; A. Xiang, *Reconciling Legal and Technical Approaches to Algorithmic Bias*, in *Tennessee Law Review*, 88(3), 2021, pp. 649-724; T.B. Nachbar, *Algorithmic Fairness, Algorithmic Discrimination*, in *Florida State University Law Review*, 48, 2021, pp. 509-558. For a focus on this problem in the public sector, see S. Alon-Barkat, M. Busuioc, *Human-AI interactions in public sector decision making: "Automation bias" and "selective adherence" to algorithmic advice*, in *Journal of Public Administration Research and Theory*, 33, 2023, pp. 153-169; J. Tomlinson, *Proving algorithmic discrimination in government decision-making*, in *Oxford University Commonwealth Law Journal*, 20(2), 2020, pp. 352-360.

24. In this regard, it has been observed that the «*opaque nature of AI can be potentially dangerous to the right to a reasoned decision*», as the «*impossibility of understanding and validating the decision process of the system can both lead its users to doubt the reliability of the decision that is proved, and violate the right to a reasoned decision*» (M. Dymitruk, *The Right to a Fair Trial in Automated Civil Proceedings*, in *Masaryk University Journal of Law and Technology*, 13(1), 2019, p. 40). This is issue is compounded by what Forti terms the «*black box barrier*», which «*indicates the impossibility for external human observers to identify, understand, and replicate the reasoning patterns chosen by machine learning algorithms*» (M. Forti, *Addressing Algorithmic Errors in Data-Driven Border Control Procedures*, in *German Law Journal*, 25/2024, p. 641). The distinction between "black-box" and "white-box" models helps clarify this issue. "White-box" models allow users to inspect internal mechanism and understand the rationale behind predictions, while "black-box" models operate with opaque internal processes that are not easily interpretable (J. Kim, H. Maathuis, D. Sent, *Human-centered evaluation of explainable AI applications: a systematic review*, in *Frontiers in Artificial Intelligence*, 17 October 2024, p. 2), Foundations models, which represent some of the most advanced AI systems today, exemplify this black-box nature, as they receive inputs and generate outputs with no human-interpretable explanation of their internal decision-making processes (M. O'Neill, M. Connor, *Amplifying Limitations, Harms and Risks of Large Language Models*, 6 July 2023, available at arXiv, pp. 12-13).
25. Cons. St. (sec. III), judgment of 20 October 2025, n. 8092, § 13.
26. On the obligation to provide reasons for algorithmic administrative acts, see L. Martucci, *Il fragile equilibrio tra l'intelligenza artificiale e la trasparenza nelle decisioni amministrative*, in *Amministrativamente*, 1, 2025, pp. 405-406; O. Mir, *The AI Act from the Perspective of Administrative Law: Much Ado About Nothing?*, in *European Journal of Risk Regulation*, 30 September 2024, pp. 11-12; H.C.H. Hofmann, *Automated Decision-*

Making (ADM) in EU Public Law, in H.C.H. Hofmann, F. Pflücke (eds.), *Governance of Automated Decision-Making and EU Law*, Oxford University Press, Oxford, 2024, p. 20; S. Foà, *op. cit.*, pp. 530-547; A. Cerrillo i Martinez, *¿Son fiables las decisiones de las Administraciones públicas adoptadas por algoritmos?*, in *European Review of Digital Administration & Law*, 1(1-2), 2020, pp. 31-33; C. Napoli, *Algoritmi, intelligenza artificiale e formazione della volontà pubblica: la decisione amministrativa e quella giudiziaria*, in *Rivista AIC*, 3, 2020, p. 342; E. Carloni, *I principi della legalità algoritmica. Le decisioni automatizzate di fronte al giudice amministrativo*, in *Diritto amministrativo*, 2, 2020, p. 293; J. Cobbe, *Administrative law and the machines of government: judicial review of automated public-sector decision-making*, in *Legal Studies*, 39, 2019, p. 648.

27. According to Article 6, point 2, of Regulation (EU) 2024/1689 (AI Act), AI systems of the types listed in Annex III are always considered high-risk, unless they don't pose a significant risk to people's health, safety, or rights. Annex III lists some important activities carried out by public administration, such as critical infrastructure, access to and enjoyment of essential public services and benefits, migration, asylum and border control management. In these cases, public administration bodies shall ensure that artificial intelligence systems they deploy are compliant with the requirements set forth in Chapter III, Section II of the Regulation. Pursuant to the transparency obligations under Article 13, high-risk AI systems must be so designed as to enable the user to interpret the system's output and use it appropriately, and they shall be accompanied by comprehensive and clear instructions for use detailing their characteristics, capabilities, performance limitations, and expected levels of accuracy, thereby providing the administration with all necessary information for the proper and lawful use thereof. Furthermore, Article 14 establishes the requirement of human oversight, mandating that high-risk AI systems be supervised by one or more duly qualified human operators throughout their period of use. Such oversight must be substantive, affording the operator the full capacity to understand the system's processes and the authority to decide not to use the AI system, to disregard or overturn its output, and to intervene in or halt its operation at any given time. As a further requirement under Article 15, high-risk AI systems must be designed to achieve an appropriate level of accuracy, robustness, and cybersecurity throughout their entire lifecycle.
28. Article 14 of the bill approved by the Italian Senate on 20 March 2025, and currently under parliamentary discussion, entitled «*Use of artificial intelligence in public administration*», provides that «*1. Le pubbliche amministrazioni utilizzano l'intelligenza artificiale allo scopo di incrementare l'efficienza della propria attività, di ridurre i tempi di definizione dei procedimenti e di aumentare la qualità e la quantità dei servizi erogati ai cittadini e alle imprese, assicurando agli interessati la conoscibilità del suo funzionamento e la tracciabilità del suo utilizzo. 2. L'utilizzo dell'intelligenza artificiale avviene in funzione strumentale e di supporto all'attività provvedimentale, nel rispetto dell'autonomia e del potere decisionale della persona che resta l'unica responsabile dei provvedimenti e dei*

procedimenti in cui sia stata utilizzata l'intelligenza artificiale. 3. Le pubbliche amministrazioni adottano misure tecniche, organizzative e formative finalizzate a garantire un utilizzo responsabile dell'intelligenza artificiale e a sviluppare le capacità trasversali degli utilizzatori. 4. Le pubbliche amministrazioni provvedono agli adempimenti previsti dal presente articolo con le risorse umane, strumentali e finanziarie disponibili a legislazione vigente».

29. See L. Shen, Z. Hu, *An Immune Evolutionary Algorithm with Punishment Mechanism for Public Procurement Expert Selection*, in *Mathematical Problems in Engineering*, 9 February 2021, p. 2, who correctly observes that «*as a public procurer has less discretion to select any other bidder than the one awarded the highest score, the evaluation committee is critical because the quality and knowledge of the experts can significantly affect the final evaluation results*». On the lack of specialized AI competences in the public sector, see the recent study by the European Commission's Joint Research Centre: R. Medaglia, P. Mikalef, L. Tangi, *Competences and governance practices for artificial intelligence in the public sector*, Publications Office of the European Union, Luxembourg, 2024. The report highlights that the «*skill gap*» is one of the main barriers to AI adoption and underscores the difficulty that public administrations face in attracting and retaining technical talent due to slow hiring procedures and wage gaps with the private sector (p. 51). Furthermore, the study breaks down the necessary competences, emphasizing the need not only for generic IT skills but for specific literacy and know-how, such as understanding machine learning models (p. 32), prompt engineering (p. 35), and, above all, a specific «*AI procurement literacy*» (p. 46).
30. T.A.R. Lazio-Roma, cit., § 14.
31. Cons. St. (sec. III), judgment of 20 October 2025, n. 8092, § 13.
32. *Ibid*.
33. Explainable AI (XAI) represents «*a set of techniques and methods to convert the so-called blac-box AI algorithms to white-box algorithms, where the results achieved by these algorithms and the variables, parameters, and steps taken by the algorithm to reach the obtained results, are transparent and explainable*» (F. Hussain, R. Hussain, E. Hossain, *Explainable Artificial Intelligence (XAI): An Engineering Perspective*, available at arXiv, 10 January 2021, p. 1). A critical distinction in XAI is between «*interpretability*» and «*explainability*». Interpretability is an inherent property of a prediction algorithm that makes it directly understandable to users, while explainability is an acquired property of the decision process, usually implemented through external means. In other words, «*interpretability reveals the internal structure of machine learning models [...]. Unlike interpretability, explainability is primarily focuses on external users of an AI systems*» (S. Chalyi, V. Leshchynskyi, *Possible evaluation of the correctness of explanations to the end user in an artificial intelligence system*, in *Advanced Information Systems*, 7(4), 2023, p. 75).
34. Article 15 of the AI Act is dedicated to the requirements of accuracy, robustness, and cybersecurity for high-risk AI systems.
35. The Analytic Network Process (ANP) represents an evolution of AHP, extending the

methodology to address situations where interdependencies and feedback relationships exist between decision elements (T.L. Saaty, *Decision Making with Dependence and Feedback: The Analytic Network Process*, RWS Publications, Pennsylvania, 1996, *passim*). While AHP assumes a linear top-to-bottom hierarchical structure with independent relationships between elements at different levels, ANP extends this concept by incorporating interdependencies and feedback relationships between criteria and alternatives (M.C. Lee, *The Analytic Hierarchy and the Network Process in Multicriteria Decision Making: Performance Evaluation and Selecting Key Performance Indicators Based on ANP Model*, 2010, in G. Lee, D. Howard, D. Ślęzak (eds.), *Convergence and Hybrid Information Technologies*, Springer, Cham, 2011, p. 125).

36. J.J. Huang, C.Y. Chen, *Resource Allocation of Cooperative Alternatives Using the Analytic Hierarchy Process and Analytic Network Process with Shapley Values*, in *Algorithms*, 17, 152, 2024, p. 3.
37. The Model Contractual Clauses for AI (MCC-AI) are a standardized set of non-binding contractual terms drafted by the European Commission for use by public contracting authorities. Their primary purpose is to ensure that the public procurement of an AI system complies with the legal obligations stipulated in the EU's Artificial Intelligence Act. Two distinct versions of the clauses were elaborated: 1) MCC-AI-High-Risk, intended for AI systems identified as «*high-risk*» under the AI Act; 2) MCC-AI-Light, designed for AI systems that are not classified as «*high-risk*» but still warrant contractual safeguards, especially regarding transparency and explainability. Plus, an official commentary was elaborated, providing guidance on how to use, customize and apply the clauses in practice. The MCC-AI are not a standalone agreement but are designed to be incorporated as an annex into a broader procurement contract, addressing key areas such as risk management, data governance, transparency, technical documentation, and human oversight. MCC-AI and commentary are available on the European Commission website.
38. Indeed, «*tanto più è dettagliata l'articolazione dei criteri e sub-criteri di valutazione, tanto più risulta esaustiva l'espressione del punteggio in forma numerica*» (Cons. St. (sec. V), judgment of 14 June 2023, n. 5854; in the same vein, see, *inter alia*, Cons. St. (sec. V), judgment of 3 February 2025, n. 839; Cons. St. (sec. V), judgment of 18 December 2024, n. 10195; Cons. St. (sec. III), judgment of 24 July 2024, n. 6677; Cons. St. (sec. V), judgment of 17 May 2024, n. 4440; Cons. St. (sec. III), judgment of 12 October 2023, n. 8893; Cons. St. (sec. III), judgment of 12 March 2021, n. 2118; Cons. St. (sec. V), judgment of 20 September 2016, n. 3911).
39. On the special requirements for participation in a tender, see E. Dalli Cardillo, *I requisiti di ordine speciale (Art. 100)*, in G.F. Cartei, D. Iaria (eds.), *Commentario al Codice dei Contratti pubblici. Dopo il correttivo*, Editoriale Scientifica, Napoli, pp. 1026-1039; V. Pampanin, *Commento all'Art. 100. Requisiti di ordine speciale*, in R. Villata, M. Ramajoli (eds.), *Commentario al Codice dei contratti*, Pacini Editore, Pisa, 2024, pp. 570-579; G. Caputi, *Art. 100. Requisiti di ordine speciale*, in A. Botto, S. Castrovinci Zenna (eds.), *Commentario alla normativa sui contratti pubblici*, Giappichelli, Torino, 2024, pp.

913-921.

40. On the principles of the new Public Contracts Code as criteria for exercising discretionary power, see M.R. Spasiano, *La codificazione dei principi del Codice dei contratti pubblici e, in particolare, del risultato, alla prova del correttivo*, in *Federalismi.it*, 10, 2025, pp. 191-234; L.R. Perfetti, *Discrezionalità amministrativa e principio del risultato*, in *P.A. Persona e Amministrazione*, 14(1), 2024, pp. 69-96; E. Quadri, *Il principio della fiducia alla luce del nuovo Codice dei contratti pubblici e delle prime applicazioni della giurisprudenza*, in *Rivista Trimestrale di Scienza dell'Amministrazione*, 4, 2024, pp. 1-15; E. Carloni, *Verso il paradigma fiduciario? Il principio della fiducia nel nuovo codice dei contratti e le sue implicazioni*, in *Diritto Pubblico*, 1, 2024, pp. 131-162; P. Marzaro, *I principi generali del nuovo codice dei contratti pubblici: la funzione nomopoietica del consiglio di stato per un 'nuovo paradigma' di amministrazione*, in *Il diritto dell'economia*, 3, 2024, pp. 441-460; M.R. Spasiano, *Principi e discrezionalità nel nuovo codice dei contratti pubblici: primi tentativi di parametrazione del sindacato*, in *Federalismi.it*, 24, 2023, pp. 222-239.
41. On the Article 9 of AI Act, see *ex multis* M. Álvarez Fernández, *Risk Management System (Article 9)*, in A. Huergo Lora (ed.), *The EU Regulation on Artificial Intelligence: A Commentary*, Wolter Kluwers, Milano, 2025, pp. 139-181; P.S. Castellano, *Risk Management Systems as a Specific Obligation for High-Risk Artificial Intelligence Systems in Article 9 of the Regulation*, in L. Cotino Hueso, D.U. Galetta (eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, Editoriale Scientifica, Napoli, 2025, pp. 567-594; J. Schuett, *Risk Management in the Artificial Intelligence Act*, in *European Journal of Risk Regulation*, 15, 2024, pp. 367-385.
42. On the Article 10 of AI Act, see *ex multis* A. Palma Ortigosa, *Data and Data Governance (Article 10)*, in A. Huergo Lora (ed.), cit., pp. 183-205; M. Loza Corera, *Data and Data Governance and Connections to Data Protection Principles in Article 10 of the Artificial Intelligence Act*, in L. Cotino Hueso, D.U. Galetta (eds.), cit., pp. 595-627.
43. On the Article 11 of AI Act, see *ex multis* G. Gallone, *Technical Documentation (Article 11)*, in A. Huergo Lora (ed.), cit., pp. 207-211; F. Ramón Fernández, *Quality Management Systems, Technical Documentation and Documentation Keeping in the Regulation*, in L. Cotino Hueso, D.U. Galetta (eds.), cit., pp. 629-647.
44. On the Article 12 of AI Act, see *ex multis* A.G. Orofino, *Record-Keeping (Article 12)*, in A. Huergo Lora (ed.), cit., pp. 213-216; W. Arellano Toledo, A. Merchán Murillo, *The Obligation to Keep Records of High-Risk Systems in the Artificial Intelligence Act*, in L. Cotino Hueso, D.U. Galetta (eds.), cit., pp. 649-663.
45. On the Article 13 of AI Act, see *ex multis* A.P. Ortigosa, *Transparency and Provision of Information to Deployers (Article 13)*, in A. Huergo Lora (ed.), cit., pp. 217-242; M. Estrella Gutiérrez David, *Transparency and Provision of Information to Deployers in Article 13 of the Artificial Intelligence Act*, in L. Cotino Hueso, D.U. Galetta (eds.), cit., pp. 217-242; pp. 665-716; K. Söderlund, *Hight-risk AI transparency? On qualified transparency mandates for oversight bodies under the EU AI Act*, in *Technology and*

Regulation, 2025, pp. 96-113; T. Gils, F. Heymans, W. Ooms, *From Policy to Practice: Prototyping The EU AI Act's Transparency Requirements,* 31 January 2024, available at SSRN: https://dx.doi.org/10.2139/ssrn.4714345.

46. On the Article 14 of AI Act, see *ex multis* G. Lazcoz Moratinos, *Human Oversight (Article 14)*, in A. Huergo Lora, cit., pp. 252-274; G. Lazcoz Moratinos, *Human Oversight or Monitoring in Article 14 of the Artificial Intelligence Act: A Mere Mandatory Requirement for High-Risk Systems?*, in L. Cotino Hueso, D.U. Galetta (eds.), cit., pp. 717-737; A.M. Corrêa, S. Garsia, A. Elbi, *Better together? Human oversight as means to achieve fairness in the European AI Act governance*, in *Cambridge Forum on AI: Law and Governance*, 1, 2025, pp. 1-17; M. Fink, Human Oversight under Article 14 ot the EU AI Act, 14 February 2025, Available at SSRN: https://dx.doi.org/10.2139/ssrn.5147196; L. Enqvist, *'Human oversight' in the EU artificial intelligence act: what, when and by whom?*, in *Law, Innovation and Technology*, 2 (15), 2023, pp. 508-535.
47. On the Article 15 of AI Act, see *ex multis* O. Fernández Fernández, *Accuracy, Robustness and Cybersecurity (Article 15)*, in A. Huergo Lora, cit., pp. 276-296; A. Aba Catoira, *Accuracy and Robustness of High-Risk Artificial Intelligence Systems in Article 15 of the Artificial Intelligence Act*, in L. Cotino Hueso, D.U. Galetta (eds.), cit., pp. 739-760; M.E. Sánchez Acevedo, C*ybersecurity in High-Risk Artificial Intelligence Systems in Article 15 of the Artificial Intelligence Act*, in L. Cotino Hueso, D.U. Galetta (eds.), cit., pp. 761-784.
48. T.A.R. Lazio-Roma, cit., § 14.
49. Cons. St. (sec. III), judgment of 20 October 2025, n. 8092, § 13: «*Risulta pertanto smentito* per tabulas *(...) che l'utilizzo dell'intelligenza artificiale abbia avutoun peso determinante ai fini dell'attribuzione di un punteggio tecnico elevato infavore di Dussmann, il che priva di rilevanza anche tutte le rimanentiargomentazioni dell'appellante, volte a censurare l'omessa valorizzazione delleosservazioni del consulente di parte, contenute nella relazione allegata agli attidel giudizio di primo grado*».
50. On the *principio dispositivo con metodo acquisitivo*, see *ex multis* F. Saitta, *La distribuzione dell'onere della prova nel processo amministrativo, tra principi e regole*, in *Il Processo*, 1, 2025, pp. 105-132; L.R. Perfetti, G. Pesce, *L'istruzione nel processo amministrativo*, Giuffrè, Milano, 2025, *passim*.
51. On the principle of *vicinanza della prova*, see *ex multis* G.M. Sacchetto, *L'onere della "vicinanza" della prova*, in *Diritto Processuale Civile e ADR*, 8-9, 2023, pp. 1852-1857; R. Rordorf, *Onere della prova e vicinanza della prova*, in *Jus Civile*, 1, 2023, pp. 11-18; R. Mazzariol, *Fatti costitutivi e vicinanza della prova*, in *Actualidad Juridica Iberoamericana*, 18, 2023, pp. 578-613;
52. In this regard, see F. Saitta, *Vicinanza della prova e codice del processo amministrativo: l'esperienza del primo lustro*, in *Judicium*, 15 June 2015, pp. 5-6, where the Authors observes that the concept of *disponibilità* set out in the Article 64 of the Administrative Procedure Code essentially coincides with the criterion of *vicinanza della prova*, whereby «*ha l'onere di provare chi più vicino alla prova o, per dirla con le parole del codice, chi ha la disponibilità della prova*». This provision, therefore, should not be confused with the

mere rule of allegation in the Civil Code (Article 2697), but constitutes the express legislative translation of the principle of proximity in administrative litigation, aimed at rebalancing the asymmetry between the administration and private parties.

53. In this regard, see C. Cavallaro, G. Smorto, *Decisione pubblica e responsabilità dell'amministrazione nella società dell'algoritmo*, in *Federalismi.it*, 16, 2019, pp. 17-18, who, in order to ensure transparency in administrative decision-making, advocate for the intervention of «*algoritmisti*», a sort of court-appointed expert with IT skills who has the task of certifying the correctness of a certain algorithm. Similarly, see M.B. Neitz, *op. cit*., p. 155, who proposed providing judges with full-time in-house «*Tech Advisors*» who would serve as neutral experts to help judges and their staff understand complex technological issues, similar to «*Chief AI Officers*» in executive agencies or «*Tech Fellows*» in Congress. Nonetheless, practical challenges remain even when an expert consultant is appointed: in an interesting case dealt by the *Consiglio di Stato* (sec. VI), judgment of 13 January 2023, n. 448) where an expert consultant was asked to acquire and analyze the operating system of a smartphone, the request was deemed «*impractical*» due to the immense volume of documentation and source code (hundreds of thousands, if not millions, of pages) and the difficulty of verifying if the executable code matched the source code. This challenge is magnified in AI models, because they exist as trained neural networks with billions of numerical weights rather than interpretable source code, and the relationship between training code and final model parameters is non-deterministic and cannot be verified through traditional code analysis methods.
54. T.A.R. Lazio-Roma, cit., § 14.
55. A recent Stanford University study (W. Liang, Y. Zhang, M. Codreanu, J. Wang, H. Cao, J. Zou, *The Widespread Adoption of Large Language Model-Assisted Writing Across Society*, available at arXiv, 17 February 2025, pp. 1-23) provides systematic evidence of rapid LLM adoption across multiple domains, analyzing over 687,241 consumer complaints, 537,413 corporate press releases, 304.3 million job postings, and 15,919 UN press releases. In particular: 1) LLM usage surged following ChatGPT's release in November 2022, with adoption occurring 3-4 months after launch; 2) by late 2024, 18% of financial consumer complaint text appears to be LLM-assisted; 3) up to 24% of corporate press releases are attributable to LLMs; 4) nearly 14% of United Nations press releases show LLM assistance; 5) 10-15% of job postings from small firms use LLM-assisted writing. Also McKinsey's 2024 Global Survey on AI findings demonstrated that: 1) 78% of organizations now use AI in at least one business function, up from 72% in early 2024 and 55% in 2023; 2) 71% of organizations regularly use generative AI, up from 65% in early 2024; 3) for the first time, most organizations are using AI in more than one business function, with companies averaging three business functions using AI (A. Singla, A. Sukharevsky, L. Yee, M. Chui, B. Hall, *The state of AI. How organizations are rewiring to capture value*, available at McKinsey website, 12 March 2025).
56. Despite their impressive capabilities, LLMs encounter substantial technical limitations such as the lack of transparency and interpretability in how these models generate their

outputs. As LLMs become more sophisticated, their decision-making processes grow increasingly opaque, creating significant obstacles for applications where understanding the reasoning behind recommendations is essential (S. Ahmed and others, *BELL: Benchmarking the Explainability of Large Language Models*, available at arXiv, 22 April 2025, pp. 1-2). With the advent of Large Language Models (LLMs), even the field of Explainable AI (XAI) has faced new challenges and undergone significant transformation. The complexity and advanced capabilities of LLMs have made XAI methods insufficient, necessitating new approaches. The very ability to «*open the black box*» has become increasingly limited with LLMs (U. Ehsan, M.O. Riedl, *Explainable AI Reloaded: Challenging the XAI Status Quo in the Era of Large Language Models*, available at arXiv, 13 August 2024, p. 1), shifting XAI's focus from merely exposing internal mechanisms to enhancing the productivity and applicability of these models in real-world settings (X. Wu and others, *Usable XAI: 10 Strategies Towards Exploiting Explainability in the LLM Era*, available at arXiv, 18 May 2025, pp. 1-43). However, research on explainability and transparency for LLMs remains in early stages, indicating substantial unexplored potential and challenges ahead (T. Shen and others, *Large Language Model Alignment: A Survey*, available at arXiv, 26 September 2023, p. 49).

57. The most widely documented and concerning limitation of LLMs is their propensity for «*hallucinations*» – the generation of plausible but factually incorrect or nonsensical information –, as LLMs generate responses based on statistical patterns rather than retrieving facts from reliable sources (J. Shao, J. Tong, Q. Wu, W. Guo, Z. Li, Z. Lin, J. Zhang, *WirelessLLM: Empowering Large Language Models Towards Wireless Intelligence*, available at arXiv, 26 September 2023, p. 3). The hallucination problem is particularly acute in specialized domains, with studies showing that ChatGPT generates hallucinated content in approximately 19.5% of responses (J. Li, X. Cheng, W.X. Zhao, J.U. Nie, J.R. Wen, *HaluEval: A Large-Scale Hallucination Evaluation Benchmark for Large Language Models*, available at arXiv, 23 October 2023, pp. 1-6) and can produce fabricated references in research contexts (S.A. Athaluri, S.V. Manthena, V.S.R.K.M. Kesapragada, V. Yarlagadda, T. Dave, R.T.S. Duddumpudi, *Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References*, in *Cureus*, 15(4), 2023, pp. 1-5).
58. LLMs also suffer from significant bias and fairness issues, as they inherit and can amplify biases present in their training data (M.A. Bouchiha, Q. Telnoff, S. Bakkali, R. Champagnat, M. Rabah, M. Coustaty, Y. Ghamri-Doudane, *LLMChain: Blockchain-based Reputation System for Sharing and Evaluating Large Language Models*, in *2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC)*, 2 July 2024 to 4 July 2024, Osaka (Japan), p. 439). These biases can relate to gender, race, socioeconomic status, and other demographic factors, potentially leading to discriminatory outputs and perpetuating existing social inequalities (N. Hicham, H. Nassera, S. Karim, *Strategic Framework for Leveraging Artificial Intelligence in Future Marketing Decision-Making*, in *Journal of Intelligent Management Decision*, 2(3), 2023,

p. 145). The presence of biased training data means that LLMs may produce outputs that reflect societal prejudices while appearing objective and authoritative (G. Tolomei – C. Campagnano, F. Silvestri, G. Trappolini, *Prompt-to-OS (P2OS): Revolutionizing Operating Systems and Human-Computer Interaction with Integrated AI Generative Models*, in 2023 IEEE 5th International Conference on Cognitive Machine Intelligence (CogMI), pp. 131-132).

59. In this regard, see E.M. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell, *On the Dangers of Stochastic Parrots: Can Language Models Bee Too Big?*, in *FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency,* March 3-10, 2021, Canada, pp. 614-623, where the Authors introduced the term «*stochastic parrots*» to describe LLMs, suggesting they are adept at mimicking language based on statistical patterns from massive training data but lack any true understanding of meaning. In particular, it has been argued that while LLMs show impressive performance on benchmarks, they are not performing true natural language understanding. They are trained only on linguistic form, not its corresponding meaning.
60. On the use of artificial intelligence in Italian public procurement, see G.F. Licata, *Intelligenza artificiale e contratti pubblici: problemi e prospettive*, in *CERIDAP*, 2, 2024, pp. 30-63; R. Esposito, *Il principio (di conservazione) dell'equilibrio contrattuale tra risultato e intelligenza artificiale*, in *www.giustizia-amministrativa.it*, 2024, pp. 1-12; A. Corrado, *I nuovi contratti pubblici, intelligenza artificiale e blockchain: le sfide del prossimo futuro*, in *Federalismi.it*, 19, 2023, pp. 128-154; D. Diaco, *Fidarsi è bene, non fidarsi è...? L'Intelligenza Artificiale entra nel nuovo Codice dei contratti pubblici*, in *Federalismi.it*, 17, 2023, pp. 119-152; M. Ippolito, *L'implementazione delle tecnologie emergenti nel processo di automazione del ciclo di vita dei contratti pubblici. Alcune riflessioni a margine del D.L.gs n. 36/2023*, in *PA Persona e Amministrazione*, 14(1), 2024, pp. 579-606. For an international perspective, see K.S. Aboelazm, *A new era of public procurement: critical issues of procuring artificial intelligence systems to produce public services*, in *International Journal of Law and Management*, 14 March 2025, pp. 1-24; K. McBride, C. van Noordt, G. Misuraca, G. Hammerschmid, *Towards a systematic understanding on the challenges of public procurement of artificial intelligence in the public sector*, in Y. Charalabidis, R. Medaglia, C. van Noordt (eds.), *Research Handbook on Public Management and Artificial Intelligence*, Edward Elgar Publishing, Cheltenham, 2024, pp. 62-78; M. Hickok, *Public procurement of artificial intelligence systems: new risks and future proofing*, in *AI & Society*, 39, 2024 pp. 1213-1227; T.C. Oliveira, A.L. Monteiro da Rocha, M. Scatolino de Rezende, *Alice:Desafios, resultados e perspectivas da ferramenta de auditoria contínua de compras públicas governamentais com uso de inteligência artificial*, 26, 2022, pp. 296-308.
61. Law No. 132 of 23 September 2025, containing «*Disposizioni e deleghe al Governo in materia di intelligenza artificiale*». For an initial comment on the provisions of the bill of this Law related to use of AI the public sector, cfr. C. Polidori, *Procedimento amministrativo e algoritmi*, in *www.giustizia-amministrativa.it*, 2025, pp. 4-22; A. Di

Filippo, *L'intelligenza artificiale nella Pubblica Amministrazione: sfide, minacce, opportunità*, in *Azienditalia*, 11, 2024, pp. 1212-1213; A. Lirosi, *L'intelligenza artificiale nel diritto amministrativo – tra riserva di umanità e necessità di garantire una maggiore efficienza amministrativa*, in *Rivista della Corte dei Conti*, 2, 2024, pp. 130-131; P. Miele, *Intelligenza artificiale e pubblica amministrazione*, in *Rivista della Corte dei Conti*, 2, 2024, pp. 157-159.

62. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence. For an initial comment on the relationship between the AI Act and the use of AI in the public sector, see V. D'Antino, *L'approccio basato sul rischio nell'AI Act: un nuovo paradigma di regolazione dell'intelligenza artificiale*, in *Federalismi.it*, 18, 2025, pp. 15-35; G. Gallone, *L'improcrastinabile esigenza di tracciare una via "italiana" per l'intelligenza artificiale nel procedimento amministrativo. Opportunità e legittimità di un intervento regolatorio nazionale a corredo dell'AI Act*, in *Giustizia Insieme*, 26 June 2025; V. Neri, *AI Act e diritto amministrativo*, in *LavoroDirittiEuropa*, 1, 2025, pp. 1-19; S. Weertz, *Generative AI in public administration in light of the regulatory awakening in the US and EU*, in *Cambridge Forum on AI: Law and Governance*, 1, 2025, pp. 1-19, esp. 10-16; S. Francario, *AI Act e pubblica amministrazione*, in *Giustizia Insieme*, 31 October 2024; O. Mir, *The AI Act from the Perspective of Administrative Law: Much Ado About Nothing?*, in *European Journal of Risk Regulation*, 30 September 2024, pp. 1-13; T. Monaco, *L'amministrazione digitale: principi e prospettive al banco di prova della regolazione italiana e comunitaria (dell'Unione Europea)*, in *Politica.eu*, 1, 2024, pp. 179-203.
63. On which see footnote 27.