European Review of Digital Administration & Law - Erdal
2026, Volume 7, Issue 1, pp. 53-60
ISSN 2724-5969 – ISBN 979-12-218-2846-7 – DOI 10.53136/9791221828467 07

*Human Responsibility and Control over Artificial Decision-making in Italy: Challenges and Problems**

Enrico Carloni
(Full Professor of Administrative Law, University of Perugia)

ABSTRACT The essay examines the principle of human responsibility and control over automated and AI-assisted decision-making in Italian public administration, building on the framework of 'algorithmic legality' developed by administrative case law and now partly codified in Article 30 of the Public Contracts Code (Legislative Decree No. 36/2023) and in the new national AI Law (Law No. 132/2025). After outlining the constitutional and European (GDPR, AI Act) roots of the non-exclusivity/human-in-the-loop principle, the essay questions whether human oversight, as currently framed, can remain effective as AI systems grow more complex and less interpretable. It identifies several structural risks – merely formal or 'rubber-stamp' supervision, automation bias, fragmented responsibility among multiple actors, and insufficient administrative skills – that may turn human-in-the-loop into a 'phantom guarantee'. Drawing a comparison with the deregulatory turn taken by the United States, the essay argues that Italian and European law nonetheless preserve an irreducible space for the human official, rooted in the personalist and impartiality principles of the Constitution. It concludes by proposing the model of the 'artificial co-pilot' as a viable path to reconcile the efficiency gains of AI with the protection of individual rights, provided it is accompanied by adequate investment in administrative capacity and governance.

KEYWORDS: Algorithmic legality - Human-in-the-loop - Artificial intelligence - Public administration - Automation bias - Reserve of humanity

TABLE OF CONTENTS: 1. Introduction: Technological Change and the Maintenance of Legal Principles. – 2. Human Responsibility and Oversight in the New Artificial Legality. – 3. Human-in-the-Loop and the Risk of Official Capture. – 4. The Irreducible Role of the Individual and the Artificial Co-pilot.

1. *Introduction: Technological Change and the Maintenance of Legal Principles*

The introduction of artificial intelligence technologies at the heart of the exercise of administrative power raises profound questions about the relationship between automation, fundamental rights, and individual guarantees, as well as challenging the traditional dynamics of the democratic State in general.

If transparency is the prerequisite for understanding and monitoring the exercise of AI power,[1] the next question concerns how automation can be integrated into decision-making processes without disempowering the human component and without producing discriminatory effects. What, more generally, emerges at the intersection of these principles is the need to construct a comprehensive artificial legality, meaning the necessary adaptation and critical review of the principles of algorithmic legality, but perhaps also a broader rethinking of the coordinates developed thus far to protect citizens in their relationship with new computational powers.[2]

Two aspects emerge as central and complement the transparency pillar (despite its limitations): on the one hand, the prevention of the risks of algorithmic discrimination, arising from the use of systems that learn from historical data riddled with biases and inequalities, developing new ones; on the other, the need to preserve a reserve of humanity in administrative decisions, which translates into the prevention of the so-called 'human in the loop'; on the other, an issue that also has a significant organizational dimension.

Both of these principles (which, as mentioned, together with transparency, constitute the fundamental core of the legality

*Article submitted to double-blind peer review.
This paper focuses on and resumes a broader reflection, for which reference is made to E. Carloni, *Critica dell'amministrazione artificiale*, Bologna, Il Mulino, 2025.

[1] On the role of transparency in the democratic state, and on some challenges related to algorithmic transparency, see E. Carloni, *Il paradigma trasparenza*, Bologna, Il Mulino, 2022; more recently, and specifically on the challenges of algorithmic transparency, E. Carloni, *Administrative transparency and AI. Attempting to catch a difficult horse*, in A. Cerrillo Martinez and J. Ponce Solé (eds.), *The EU Artificial Intelligence Act and the Public Sector*, Athens, Eplo European Public Law Series, 2025, 201 ff.

[2] Numerous requests in this direction, most recently, e.g., in P. Hacker *et al.* (eds.), *The Oxford Handbook of Foundations and Regulation of Generative AI*, Oxford, Oxford University Press, 2026.

of automated processes, first in their algorithmic form and now in that linked to the use of AI), are intertwined with transparency, which is a precondition, but which is fully realized only when the individual can assume a substantial role, at least in understanding the decision-making dynamics entrusted to machines. The very issue of human responsibility in the 'artificial' decision-making process stands as a central issue, which deserves to be repeatedly and continually questioned.

The challenge of artificial administration reveals, upon closer inspection, the complexity of the exercise of the public function, as a dynamic that runs through the organization and public action and is projected onto the protection mechanisms: in particular, if the procedure is no longer able to be the center of gravity around which to build the priority places for the protection of the interests involved in the public decision, if the procedure is therefore not the form of the artificial administrative function,[3] this does not mean that these venues are lacking. Auby explains this well concerning the 'reserve of humanity': the moments of human choice are multiple and occur before and after the decision is made, in the overall unfolding of the function (the choice to automate a procedure, the definition of the criteria that will guide it, also transferring these priorities into the definition of the clauses for the purchase of the technological solution - or in the instructions for the offices that will follow the programming -, and so on).[4] Likewise, Gallone highlights that, in practice, the algorithm (and therefore AI) exists within inescapably human contexts: the will to resort to automation; the final act, which can only be performed by an administrative body, even when produced by a machine; and the possible exercise of self-regulation powers.[5]

2. *Human Responsibility and Oversight in the New Artificial Legality*

Although the recent Italian law on AI (Law No. 132 of 2025[6]) tends to see human officials as the unavoidable protagonists even in the new scenario of artificial administration, affirming the strictly instrumental nature of machines and therefore full 'human' responsibility for the use of AI solutions, the issue is more complex, and perhaps less reassuring.

Even in the era, still current, of 'algorithmic legality', but still according to European legislation and the provisions of the Italian Code of public procurement (Article 30 of Legislative Decree No. 36 of 2023), the balance between the necessary openness to the potential and advantages of AI and the persistent need for protection and risk control is based, in no small way, on the effectiveness of the principle of 'non-exclusivity'. However, the Italian legislator's choice in Law No. 132 perhaps seems too demanding, as it states, in particularly clear terms, that 'the use of artificial intelligence is instrumental and to support the procedural activity, respecting the autonomy and decision-making power of the individual, who remains solely responsible for the provisions and procedures in which artificial intelligence has been used'.[7]

This is a principle that is worthy of support, especially since it does not exclude the use of AI even for procedural decisions (meaning those that directly impact specific legal situations), but at the same time requires full 'human control' in these cases. However, the regulation poses problems that deserve mention. The text, as formulated, perhaps fails to adequately address certain issues: among these, the resulting controversial relationship with European legislation, which expressly permits the use of AI models if they comply with the parameters and rules defined at the European level by the AI Act, even by public

[3] Paraphrasing F. Benvenuti, *L'attività amministrativa e la sua disciplina generale*, in *Atti del convivium regionale di studi giuridici*, Trento, 28 June 1957, 1958, 49 ff., then in *Scritti giuridici*, Milan, Vita e pensiero, 2006, vol. 2, 1509 ff.

[4] It is important to realize that human intervention can, and usually does, occur at several levels: the decision to use an algorithm, the development of the algorithm in dialogue with experts, the adoption of the algorithm, and finally, the final decision. The problem of the 'automatic' decision arises in the last phase, which does not mean that the negligence and errors committed by the administration in the three previous phases are irrelevant: see J.-B. Auby, *Preface*, in G. Gallone, *Riserva di umanità e funzioni amministrative*, Padua, Cedam, 2023, XII.

[5] See G. Gallone, *Riserva di umanità e funzioni amministrative*, cit., espec. 143 ff.

[6] Law No. 132 of 23 September 2025, regulates the use of Artificial Intelligence in Italy, establishing rules on intellectual property, professional use, and sanctions. It imposes human primacy, the obligation to inform customers about the use of AI, protects copyright, and increases penalties for technological crimes.

[7] Law no. 132 of 2025, art. 14, paragraph 2.

administrations (one need only scan, in this regard, the list of 'high-risk' activities, which include entire sectors of typically public intervention). Furthermore, it should be noted that European legislation, in setting conditions and limits on the use of fully automated solutions, effectively considers them fully legitimate.[8]

Above all, however, if fully understood, the Italian legislation would introduce a principle that goes far beyond the human-in-the-loop principle, effectively making significant use of AI by public administrations in the context of decisions addressed to specific recipients impossible to implement, given that human control over procedures calls into question some of the characteristics and advantages inherent in the use of AI (including the capacity for massive information processing and rapid processing speed).

The scope of this principle, therefore, can only be scaled back, precisely because it is embedded in a dense network of European regulations that propose a precautionary yet open strategy, which does not appear to allow for closures to the AI phenomenon at the level of individual Member States: while there is room to guide the use of AI in the public sector, this cannot be exercised at the national level to excessively preclude the use of artificial intelligence solutions, not even by public deployers.[9]

The situation is different if we read this statement in conjunction with the indisputable principle (though, as we will see, not necessarily fully applicable) of non-exclusivity or 'reserve of humanity', which clearly emerged already during the jurisprudential construction of the principles of algorithmic legality. This principle, as we have seen, was first embodied in administrative case law (drawing on the formulation contained in the GDPR) and then by the legislator of the Contracts Code itself.

The need for a human role in the decision-making process – not as central as required by Italian law, but an unavoidable one – is clearly rooted in the legal system. The decision can be made by a virtual agent, but there must exist, and endure, a space for supervision/review: a 'reserve of humanity' for the benefit of the official/natural person, to use the expression proposed by Ponce.[10] This assertion, as we will see, while perhaps inadequate from the perspective of more recent Italian legislation, is in fact both salutary and problematic in operational terms.

Human oversight of AI is in fact confronted with a series of different problems. It should in any case be considered that, given the potential associated with the introduction of AI solutions, preconceived, excessive, or even merely disproportionate closure strategies paradoxically produce risks of human capture as well, which develop, however, according to 'shadow AI' dynamics - all the more dangerous because they are even more opaque and devoid of organizational and precautionary mechanisms.[11]

On the other hand, even when faced with the transparent use of AI solutions, a series of questions arises that are useful to break down.

AI transparency itself, as we've seen, is a complex system, but not always fully satisfactory because it is currently unable to ensure adequate explainability, except through human intervention. However, human intervention is effective only if the official is able to understand the machine's actions (which must therefore be adequately transparent to them): this risks producing a paradox (what we might call opaque AI transparency).

The principle of non-exclusivity, in turn,

[8] The declared 'provisional' nature of the decision, which is left to the full responsibility of the human individual, can be used to delimit the perimeter of the national provision.

[9] It can be argued that the relationship between European and national legislation is such that while there is room for internal regulation that raises the level of guarantee (on the model of essential levels, as noted by G. Gardini, *L'uso preparatorio dell'AI come limite agli eccessi regolativi. Per una valutazione "in concreto" dell'intelligenza artificiale applicata all'agire pubblico,* in *Le istituzioni del federalismo,* 2, 2025, 261-299), individual Member State regulation encounters the limitation of not being able to introduce conditions that hinder the circulation of AI systems (along this line, see O. Mir, *The AI Act from the Perspective of Administrative Law*, 6, although specifically with reference to non-high-risk systems).

[10] J. Ponce, *Inteligencia artificial, Derecho administrativo y reserva de humanidad*, in *Revista General de Derecho Administrativo*, no. 50, 2019, 26.

[11] In particular, as a result of the unauthorized use of AI systems by employees, 'who autonomously and spontaneously, often to improve the efficiency of their work, use this technology as a cognitive and decision-making aid without the approval of their organization': A. Innocenti, *Shadow AI: cos'è e come gestire l'AI Ombra, l'Intelligenza Artificiale non ufficiale nelle aziende*, in *Rivista AI*, October 2024, (in www.rivista.ai). The topic is carefully treated in G. Gardini, *L'uso preparatorio dell'AI come limite agli eccessi regolativi*, cit.

presupposes a minimum level of transparency, such that the second principle is a prerequisite for the first (and for the ability to recognize hypotheses of discriminatory decisions, except here for the possibility of verification based on comparison with similar decisions and therefore on the results, even without knowing the procedural rationale). The problem, then, is that for the human-in-the-loop to be consistent, the principle of transparency must be effective, as must an adequate ability to understand the decision-making logic that guides the choice.[12]

Technological evolution is leading us into a decision-making environment that becomes increasingly less interpretable and comprehensible based on 'human' criteria and logic, the further we move away from algorithmic decisions and AI-driven decisions based on predetermined parameters, using databases selected and trained on these data.

All these conditions are no longer present in the case of Gen-AI: we are therefore entering a field where the affirmation of the principle of non-exclusivity risks losing its operational capacity and effectiveness, fading into a manifesto principle useful for enabling automated decision-making modules that, in practice, forgo any human capacity to challenge and refute them (except, however, the possibility of asking AI to explain and justify its conclusions).[13] While this allows for an explanation in human-understandable language, it leaves the machine alone free to self-falsify, even based on human stimuli and prompts.

Even assuming the overall transparency of AI systems and their functioning in individual decisions is sufficient, critical questions remain. We can attempt to briefly outline a series of open questions.

A first problem is linked to the merely formal role (in law or in fact) of the individual: in this case, we speak of superficial human supervision, meaning the requirement that the individual's contribution be significant and effective for the requirement of human intervention to be considered satisfied. In many contexts, even formally supervised by humans, the individual lacks the time, expertise, or authority to intervene to control or modify the decision made by the machine.[14]

In particular, in high-volume services (e.g., mass provision of social benefits), requiring consistent and in-depth human review of every case can be impractical: the human supervisor receives complex information flows but has little time to evaluate alternatives, and human oversight degrades to superficial or formal oversight.

A second problem, which appears destined to become increasingly pressing (also due to the 'qualification' of AI required by the AI Act, but more generally to the growth in AI capabilities), is the presence of automation bias and overreliance on AI systems. Essentially, when AI is deemed reliable, and its output is typically adequate (think, for example, of a translation), operators tend to rely on the system (automation bias), losing critical capacity and ignoring contrary signals. In the absence of system support, if the system fails to suggest areas of uncertainty or risk, and therefore fails to facilitate human interpretation of deficiencies and malfunctions, the human-in-the-loop risks failing.[15]

A third problem concerns the fragmentation of responsibilities, resulting in a loss of effective control. When multiple stakeholders (manufacturer, service provider, integrator, administrator, or official) exist, responsibility becomes fragmented, and one individual may not have all the necessary resources to exercise effective control, or may defer to human control exercised by other stakeholders in the chain without effective human verification.[16]

A fourth problem concerns the skills required for effective supervision: these skills must be possessed by the individual and the department as a whole. Human supervision requires resources, training, clear roles and responsibilities, continuous monitoring procedures, and impact assessments. Without

[12] From a non-legal perspective, see P. Benanti, *Human in the loop. Decisioni umane e intelligenze artificiali*, Milan, Mondadori Università, 2022.

[13] In this sense, see the proposal of G. Carullo, *Large Language Models for Transparent and Intelligible AI-Assisted Public Decision-Making*, in *CERIDAP*, no. 3, 2023, 1 ff.

[14] Both the GDPR (see: Article 22, and the WP29/EDPB Guidelines) and the ICO Guidelines require that the human reviewer understands the model, can disagree and change the decision, and that the process is not 'based solely on automation'.

[15] The European Union AI Act (art. 14) requires, in any case, that high-risk systems are 'effectively monitorable' and that technical measures facilitate human interpretation and corrective intervention.

[16] See e.g. U.S. Government Accountability Office, *'Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities* (GAO-21-519SP)', GAO, 30 June 2021, which emphasizes traceability, auditability, and 'governability' as conditions for effective HITL in the public sector.

these conditions, the human-in-the-loop approach is ineffective.

3. *Human-in-the-Loop and the Risk of Official Capture*

The different expressions used in legal literature to frame this discussion are significant: the principle of non-exclusivity is sometimes defined as a 'reserve of humanity', protecting an 'irreducible human', and sometimes as 'human oversight'.

The principle is linked to the concept of human-in-the-loop (HITL), which originates in complex systems engineering and indicates the presence of human intervention in automated processes. In the context of artificial intelligence applied to public decision-making, human-in-the-loop has become a fundamental principle: it reflects the idea that decisions cannot be entrusted to algorithmic systems and AI, but must involve human oversight or intervention to ensure their legitimacy, accountability, and protection of fundamental rights.[17] The principle is distinct (and must be distinguished) from that of human on the loop - HOTL, which refers to the idea of human supervision of systems rather than of the individual output, and therefore without substantial intervention in the individual decision.[18]

It is a principle that has been confronted by the doctrine for a long time, which has analyzed the problem from several angles, called to ensure human responsibility (both in terms of imputation, responsibility in the strict sense, and accountability): with the advent of Gen-AI the principle poses new problems and in these terms questions the law, in search of AI models capable of ensuring an effective HITL (an outcome that is not technically a foregone conclusion).[19]

As mentioned, HITL is a necessary condition for the effective protection of fundamental rights and the guarantee of individual dignity. This is primarily because only human intervention can ensure that individuals are not reduced to the 'object' of automated decisions, thus ensuring the right balance between efficiency and the protection of human rights.[20]

On the other hand, not all human oversight is truly meaningful. In this sense, it's important to note the risk of mere formal approval by an official, who can only make the principle effective if competent to understand, challenge, or modify the decision. In short, the mere presence of an official formally approving the output isn't enough: the oversight must be 'meaningful', that is, exercised by a competent individual truly capable of understanding, challenging, or modifying the algorithmic decision, without being subject to AI. For this to happen, a series of preconditions is necessary.

The administration's ability to avoid being captured by machines and to provide individual responses to automated decisions resulting from AI depends in no small measure on the quality of the administration and the breadth of expertise it possesses. On the other hand, not all human oversight is truly meaningful. In this sense, it's important to note the risk of mere formal approval by an official, who can only make the principle effective if competent to understand, challenge, or modify the decision. In short, the mere presence of an official formally approving the output isn't enough: the oversight must be 'meaningful', that is, exercised by a competent individual truly capable of understanding, challenging, or modifying the algorithmic decision, without being subject to AI. For this to happen, a series of preconditions is necessary.[21]

The new artificial administration has an extraordinary need for high-level and interdisciplinary skills, which must focus on optimizing deployment strategies, algorithm

[17] The issue is of particular importance, evidently, not only in the field of AI applied to administrative action: the case of the use of unsupervised AI systems in the war field is exemplary (on hard cases in this regard, see D. Amoroso, *Sistemi di supporto alle decisioni basati sull'IA e crimini di guerra: alcune riflessioni alla luce di una recente inchiesta giornalistica*, in *Diritti umani e diritto internazionale*, no. 2, 2024, 347 ff.).

[18] These are concepts that clearly exclude, in any case, the phenomenon of human out of the loop, that is, the hypothesis of full artificial exclusivity.

[19] See e.g. J. Laux and H. Ruschemeier, *Automation bias in the AI Act: On the legal implications of attempting to de-bias human oversight of AI*, in *arXiv* (arXiv.2502.10036), 2025; M. Chiodo *et al., Formalising human-in-the-loop: Computational reductions, failure modes, and legal-moral responsibility*, in *arXiv* (arXiv.2505.10426), 2025, analyzes the different HITL configurations, classifying them based on legal responsibility and the enforceability of explainability.

[20] See, for example, in the framework of the European GDPR, L. Bygrave, *Article 22 Automated individual decision-making, including profiling*, in C. Kuner, L.P. Aust and V.J.S. Greenwood (eds.), *The EU General Data Protection Regulation (GDPR): A commentary*, Oxford, Oxford University Press, 2022, 248 ff.

[21] See again the now outdated (but not for this reason not worthy of attention as a regulatory model) Executive order 14110, Biden, 'On Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence'.

control, and building public AI, or AI designed consistently with the constitutional characteristics and missions of administrations, new algorithms capable of revealing themselves and their computational power. From this perspective, the government's AI bill's decision to base the new policy on the principle of spending invariance cannot fail to have a negative impact.[22]

In the absence of adequate skills and (moreover) an adequate critical stance, HITL risks being rendered meaningless. This is not only due to the incomprehensibility of the algorithmic process, but also due to automation bias, that is, the tendency of humans to rely on AI.

The concept of automation bias is interesting: it reveals the tendency of humans to overly trust recommendations generated by automated systems, even when they are incorrect or questionable. Applied to law and public administration, it means that even if there is formally 'human control', this can be reduced to an uncritical ratification of the algorithm's decisions.[23]

Problems such as the so-called 'uncritical ratification' (or 'rubber stamping': the official approves the output out of deference to technology), the asymmetry of skills (officials do not have sufficient technical knowledge to question the result), make the ultimate responsibility of man only apparent, which makes that of the human in the loop a 'phantom guarantee', as suggested by some authors. Whether we frame the problem with the notion of algorithmic bias or algorithmic over-reliance, the literature and official documents recognize the phenomenon as a critical point in AI governance.[24]

The problem is not simply a matter of faulty human-machine interaction, but takes on a systemic nature that impacts the legality, reliability, and even the transparency of administrative action, requiring regulatory, organizational, and training interventions to ensure that the use of AI remains compatible with constitutional principles and the substantive protection of fundamental human rights.

The Italian administrative judge's fear, when faced with an administrative decision, of 'legitimizing its devolution to a completely impersonal computerized or mathematical mechanism lacking the ability to assess individual concrete cases, which is typical of the traditional, procedural investigation that should inform administrative activity, especially when it results in measures that impact the subjective legal positions of private individuals', therefore remains relevant and cannot be overcome by cosmetic solutions.

4. *The Irreducible Role of the Individual and the Artificial Co-pilot*

The recognition of the role of the official is not, as mentioned, merely technical: it is not simply a tool for avoiding errors, nor merely a mechanism for substantial transparency. If this were the case, we could indeed envision a prospect, linked to technological evolution, of full automation: machines could supervise other machines, correcting their errors; XAI could provide us with the answer to the dilemma of explaining the decision made by AI.[25]

The question then becomes whether the 'value' of the official is inalienable, and whether this must necessarily be a human official. The possibility, utility, and legitimacy of using virtual agents instead of officials in the performance of tasks (here, and relevantly, public ones) is a long-standing issue. Without delving into more complex issues and thus

[22] A recurring issue, but no less questionable for that, in Italian reforms affecting public administration.

[23] See R. Calo, *Robotics and the Lessons of Cyberlaw*, in *California Law Review*, vol. 103, no. 3, 2015, 513 ff.; D.K. Citron and F. Pasquale, *The Scored Society: Due Process for Automated Predictions*, in *Washington Law Review,* no. 89, 2014; see also K. Yeung, *Algorithmic Regulation: A Critical Interrogation*, in *Regulation & Governance,* vol. 12, no. 4, 2018, 505.

[24] See L. Edwards and M. Veale, *Slave to the Algorithm? Why a "Right to an Explanation" is Probably Not the Remedy You Are Looking For*, in *Duke Law & Technology Review*, no. 16, 1, 2019, 8 ff.; S. Wachter, B. Mittelstadt and L. Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the GDPR*, in *International Data Privacy Law*, vol. 7, issue 2, 2017, 76 ff.; see also Alon-Barkat and Busuioc, that in studying automation bias and selective adherence, which have empirically highlighted how, even in 'human-in-the-loop' contexts, excessive deference to algorithmic output can undermine the fairness and legitimacy of public action. S. Alon-Barkat and M. Busuioc, *Human–AI interaction in public sector decision making: "Automation bias" and "Selective adherence" to algorithmic advice*, in *Journal of Public Administration Research and Theory*, vol. 31, issue 2, 296–312.

[25] For the state of reflection in the field of digital technology sciences, see: M. Chen *et al., Evaluating Large Language Models Trained on Code*, in *arXiv* (arxiv.2107.03374), 2021; B. Peng *et al.*, *Check Your Facts and Try Again: Improving Large Language Models with External Knowledge and Automated Feedback*, in *arXiv* (arxiv.2302.12813), 2023. See G. Carullo, *Large Language Models for Transparent and Intelligible AI-Assisted Public Decision-Making*, in *Ceridap*, no. 3, 2023.

having to grapple with the extensive literature on the limits of the 'nonhuman',[26] e might ask what space is reserved for the human (and, specifically, the human official) in the journey from normative command to its concrete execution, from the abstract need to respond to needs to their necessary satisfaction and effective care, and whether this space could therefore be occupied by a machine.[27]

This is a problem that is now evolving with new potential accelerations in the comparative scenario, first of all due to the transformation that the new American administration intends to bring to the federal apparatus (and, more generally, to AI).[28] Donald Trump, with the Executive Order that marked the overcoming of the precautionary policies of the Biden Presidency, seems to have clear ideas on this point: in a strategy aimed at 'Removing barriers to American leadership in Artificial Intelligence', the precautions regarding the introduction of AI solutions are decidedly reduced, without this raising fears.[29] On the other hand, all the most clearly optimistic rhetoric aimed at promoting development (free or uncontrolled, as you wish) accompanies recent US measures on artificial management and the regulation (or non-regulation) of AI as a whole.

In a context where technology advances at an exponential pace, it's now clear that, without preconceived notions (often unhelpful in practice), potential risks cannot be ignored, and they must be managed. These risks increase as AI becomes more autonomous and acquires a central role in public decision-making.

In the Italian scenario, a 'place' reserved for human officials is, in any case, unavoidable. For several reasons, which stem from the personalistic dimension inscribed in the first articles of our Fundamental Charter and permeate its entire structure. The responsibility of the official cannot be subsumed by that of the machine, and impartiality itself (in its 'inclusive' sense) requires adequate consideration of all significant elements relating to each position. This prevents the administration's field of vision from being limited to the 'forms' imagined by a deterministic algorithm, but also to that of intelligent systems operating based on classifications and correlations that force individual positions into simplifying boxes.

The Italian Constitution, on the other hand, is clear in placing the official at the center of the administrative system, responsible for and guarantor of the proper exercise of public functions. The exercise of entrusted functions 'with discipline and honor',[30] even taking an oath where the law requires it, clearly calls into question the individual's personal commitment and refers to a series of 'qualities' inherent to the human person.[31]

The transformations underway, however, require avoiding rigid interpretations and perhaps beginning to carefully reread the text of the Constitution. From the Charter's deepest threads, a finer line can be drawn. This line (in its relationship with the law, which is responsible for implementing the effective conditions of impartiality and proper functioning) may perhaps allow us to better define the space of the irreducible human. This will be increasingly necessary as machines succeed in assimilating the qualities of the individual,[32] and as the concrete fulfillment of people's rights and the effective removal of the obstacles they encounter in everyday life require seizing the opportunities inherent in their use.

The risks associated with the full automation of decision-making processes can obviously be mitigated by granting machines a lesser, more subservient, and investigative role in procedures that remain the responsibility of a human. And this is the path we are most often faced with.[33]

It's conceivable that, beyond the purely

[26] See. e.g. T. Chamorro-Premuzic, *Io, Umano. Ai, automazione e il tentativo di recuperare quello che ci rende unici*, Milan, Apogeo, 2023.

[27] On the need for human supervision when decisions affect individuals, see e.g., I.M. Wróbel, *Artificial intelligence systems and the right to good administration*, in *Review of European and Comparative Law*, no. 49 (1), 2022, 219 ff.

[28] The reference is to the temporary structure called DOGE (Department of Government Efficiency), foreseen by White House, Executive order, 20 January 2025, 'Establishing and implementing the President's "Department of Government Efficiency"'.

[29] White House, Executive order, 23 January 2025.

[30] Art. 54 of the Italian Constitution.

[31] For a comprehensive rereading of Article 54 of the Constitution, see, for example, G. Sirianni, *I doveri dei politici e degli altri decisori pubblici in Italia*, Naples, ESI, 2019. A clearly 'human' dimension to the exercise of public functions is also derived from Article 28 of the Constitution, as well as from a declaration of the personalist principle enshrined in Articles 2 and 3 of the Constitution.

[32] If not by imitating human reasoning, then by the ability to achieve a result that is in line with that required by the values defined by law and the Constitution.

[33] This appears, in fact, to be the model suggested for the regulatory activity by the Italian law on AI, no. 132 of 2025.

instrumental use of AI by officials (who use it, for example, to process data, translate text, or formally restructure a document, as if it were a mere 'virtual assistant'[34]), there will be increasingly frequent use of action modules featuring an artificial co-pilot working alongside the human pilot.[35] This prospect can improve administrative capacity, freeing officials from repetitive tasks and increasing public performance (this, however, is not without the risk of worsening employee conditions, as some studies warn).[36]

In these scenarios, however, the risk is that the (human) official-decision maker will be substantially subordinated to the (artificial) official-instructor: if the use of automation is justified by the need to process a considerable amount of data, or to integrate diverse and complex data, or to operate in areas beyond the experience and expertise of officials, it is clear that it will be difficult for the human agent to truly substitute their own decisions for those proposed by the machine.

The interaction between humans and AI in decision-making processes can lead, specifically, in addition to the aforementioned automation bias (excessive reliance), to forms of selective adherence (partial acceptance of output consistent with the official's stereotypes). Studies, based on experiments in the Netherlands, highlight potential negative consequences in these scenarios, especially for vulnerable citizens.[37]

Administrative efficiency can be achieved by 'empowering' the human official through integration with an environment in which artificial intelligence intervenes: the more expertise the human official has with respect to the phenomenon they are called upon to administer, and with respect to the functioning of the machines with which they interact, the more this hybridization is a factor of efficiency and poses no risk. Conversely, the more the official operates in an environment they are unable to understand and interpret (without the aid of machines), and are unaware of the functioning of the tools they use, the more the decision ends up being essentially AI-based and incapable of being contradicted (except in erratic terms) by the individual.

In this hypothesis, the formally subservient nature of AI becomes misleading: it allows us to avoid in-depth questions about the guarantees associated with the use of AI, but this translates into a substantial lack of protections: the human 'veil' prevents us from fully grasping the substantial role of the machine and from activating the albeit limited protection mechanisms established by the legal system against truly automated decisions.

Subordination to the machine is, on the other hand, the result of the evolution of the tool, which makes it increasingly sophisticated (just think, in everyone's experience, of how the translation capacity of automated systems evolves), but also of a rhetoric that accompanies its advent: as Sadin points out, the use of notions like 'artificial intelligence' (but also the use of other expressions, such as neural networks, etc.) conceals the idea of technologies whose status has changed, now carrying an aletheic function (which is responsible for showing the truth). AI risks appearing suitable for solving the dilemma of replacing humans thanks to an augmented anthropomorphism, which is typical of virtual solutions. The conclusions of these are therefore indisputable for several reasons, not least because of the symbolic prestige that accompanies their advent.[38]

Given all this, and warned of these risks, the artificial co-pilot model is an interesting solution because it allows for the combination of human resources and machine performance. However, it requires significant investment in quality administration, in offices that integrate socio-legal and technical-IT skills. And it requires governance, with administrative policies, the use of AI solutions consistent with public needs, and transparency.

[34] On possible categorizations, see A. Cerrillo Martinez, *Robots, asistentes virtuais e automatização das administrações públicas*, in *Revista Galega de Administración Pública,* no. 61, 2021, 271 ff.

[35] On the prospect of a 'virtual assistant', see already E. Schweighofer, *An e-Government Interface for the Director-General – Or: How to Support Decision Makers with an Electronic Chief Secretary?*, in R. Traunmueller (ed.), *Electronic Government: Third International Conference, EGOV 2004, Zaragoza, Spain, August 30-September 3, 2004, Proceedings*, Berlin Heidelberg, Springer-Verlag, 2004, 148.

[36] Studies, it must be said, not referring to public administrations: M. Bano *et al.*, *The Rise of AI Copilots - Literature Review and Research Agenda*, preprint in *ResearchGate*, December 2024.

[37] S. Alon-Barkat and M. Busuioc, *Human-AI Interactions in Public Sector Decision-Making: "Automation Bias" and "Selective Adherence" to Algorithmic Advice*, cit.

[38] Ideas not dissimilar from A. Vespignani, *L'algoritmo e l'oracolo*, Milan, Il Saggiatore, 2019.