Generative Artificial Intelligence in Public Administration: An Integrated Framework for Regulatory Compliance, Ethical Governance, and Digital Transformation

Gaetano Riccio[1,*]

[1] School of Medicine and Surgery, University of Naples Federico II, Via Sergio Pansini, 5, 80131 Napoli NA, Italy

Abstract

This article presents and explains many of the common variations, as well as many of the formatting elements an author may use in the preparation of the documentation of their work. The integration of Generative Artificial Intelligence (GIA) in Public Administration (PA) is emerging as a powerful tool to optimize decisionmaking processes, improve service delivery and strengthen interaction with citizens. However, the regulatory complexity and ethical risks associated with the use of these technologies require a structured approach to ensure their responsible adoption. This study presents an integrated framework that harmonizes the requirements of the AI Act, GDPR and international standards (ISO 42001, NIST AI RMF), adapting them to the specifics of the public sector. The analysis highlights how multilevel governance, risk assessment, and human oversight are key elements to mitigate algorithmic bias, ensure transparency, and protect fundamental rights. Through a comparative study of use cases in different European administrations, the research shows that institutions that adopt an ethical and technically robust governance model more effectively address challenges related to regulatory compliance, cybersecurity and public trust. The results suggest that a strategic and regulated implementation of IAG in PA not only improves operational efficiency but also contributes to the creation of sustainable and inclusive public value.

Keywords

Generative Artificial Intelligence, Public Administration, Framework

1. Introduction

Intelligence (AI) systems in Public Administrations represent a technological frontier that promises to revolutionize public services, optimize internal processes and improve the relationship with citizens. However, this transformative process requires a structured approach that balances technological innovation, regulatory compliance and protection of fundamental rights. The analysis of the AI Framework for Public Administration presented in this article demonstrates that the effective adoption of AI requires a multilevel governance model that integrates requirements of the AI Act, the GDPR and international standards, adapting them to the specificities of the public sector. The research highlights how public administrations need not only technical tools, but also transversal skills, adequate organizational structures and continuous risk assessment processes to ensure ethical and responsible implementations of AI. The results suggest that administrations that adopt an integrated framework address more effectively the challenges of regulatory compliance, bias mitigation and data protection, generating greater trust in citizens and sustainable public value.

Ital-IA 2025: 5th National Conference on Artificial Intelligence, organized by CINI, June 23-24, 2025, Trieste, Italy
*Corresponding author.
gaetano.riccio2@unina.it

2. Theoretical Framework and Reference Literature

Generative Artificial Intelligence represents a subset of AI characterised by the ability to create original content rather than merely analysing existing data. As defined by Radford et al. (2019), IAG comprises "computational models designed to generate novel content that mimics statistical features of training data". These systems, primarily based on transformer and deep learning architectures, have experienced exponential acceleration in recent years, culminating in models such as GPT (Generative Pretrained Transformer), DALLE and Midjourney, which have democratised access to powerful generative tools. Their adoption in PA introduces transformative potential, but also raises unique issues compared to traditional AI systems, particularly in terms of reliability, content control and legal accountability.

The academic literature on the implementation of AI in the public sector has started to expand significantly but still has significant gaps when focusing specifically on generative systems. Wirtz and Müller (2023) conducted a systematic review identifying 135 academic articles on AI in PA published between 2000 and 2022 and found that only 12% addressed governance issues and 7% ethical aspects. Even more meagre is the specific literature on generative AI in PA, with pioneering studies such as that of Criado and Gil Garcia (2019) exploring the potential of advanced chatbots in public service delivery but highlighting the lack of comprehensive implementation frameworks.

The theoretical framework for analysing IAG in PA can be articulated through three complementary perspectives. The first is public value theory (Moore, 1995; Twizeyimana & Andersson, 2019), which assesses the impact of IAG in terms of service improvement, operational efficiency and institutional trust. The second perspective is digital ethics (Floridi & Cowls, 2019), which examines the moral implications of decision automation in the public sphere. The third is technology acceptance theory in the public sector (Venkatesh et al., 2016), which analyses the organisational and cultural factors that influence the adoption of advanced technologies in public bureaucracies.

A review of the literature highlights three main research gaps. First, there is a dearth of empirical studies on IAG implementation in PAs, with few analyses of success factors and organisational barriers. Van der Voort et al. (2019) point out that most studies focus on technical aspects rather than implementation and organisational dimensions. Secondly, there is a lack of integrated frameworks that harmonise regulatory requirements, ethical considerations and operational needs. Misuraca et al. (2020) highlight the fragmentation of approaches, with models tending to focus on single aspects (regulatory compliance, technical performance, organisational acceptance) without a holistic view. Finally, there is a dearth of research on the specificities of IAG versus traditional AI in the public context, with little attention paid to the particular challenges posed by generative systems in terms of content control, legal accountability and integration into public decision-making processes.

This article aims to contribute to bridging these gaps by analysing an integrated framework for implementing IAG in PA, which balances regulatory requirements, ethical considerations and operational needs in a holistic approach. As pointed out by Veale and Borgesius (2021), 'effective AI governance in the public sector requires the integration of multiple regulatory frameworks into a coherent and operationally sustainable approach', a challenge that this article aims to address through the analysis of a specific implementation model for IAG.

3. The Integrated Framework for Generative AI in the PA

The Artificial Intelligence Framework for Public Administration developed and analysed in this study (attached) represents a structured approach to the implementation of AI systems, including generative ones, in the public sector context. Its distinctive feature lies in the integration of regulatory requirements, ethical considerations and operational needs into a coherent and applicable

model. The framework consists of four main dimensions: strategic governance, risk assessment, cybersecurity and training, each adapted to the specifics of IAG and the peculiarities of the public context.

Strategic governance constitutes the fundamental pillar of the framework, laying the foundations for an implementation aligned with both institutional objectives and regulatory requirements. The model proposes a multilevel governance structure that includes an interdisciplinary Ethics Technology Committee, a Centralised AI Systems Registry and a Three-year Adoption Plan. As pointed out by Janssen and van der Voort (2020), effective AI governance requires "institutional structures that balance technical, legal and ethical expertise in an adaptive approach", a principle that the framework implements through the multidisciplinary composition of the Committee. Particularly relevant for generative systems is the provision of an AI Governance Board (AGB) that includes not only managers and technicians, but also legal experts and ethical representatives from NGOs and universities, thus ensuring a multistakeholder approach to governance.

Alignment with international standards is another distinctive feature of the framework, which incorporates principles and methodologies from the NIST AI RMF, ISO 42001 and the MIT AI Risk Repository. This integration responds to Lewis et al.“s (2022) recommendation to 'harmonise technical and regulatory standards into a coherent approach to AI governance”. Particularly significant for IAG is the adoption of NIST's Generative AI Profile, which introduces specific robustness parameters for systems such as chatbots and virtual assistants, thus addressing the unique quality control challenges posed by generative systems.

Risk assessment is a crucial component of the framework, particularly relevant for IAG given its potentially unpredictable nature. The model proposes an innovative Integrated Assessment that combines the Fundamental Rights Impact Assessment (FRIA) required by the AI Act with the Data Protection Impact Assessment (DPIA) required by the GDPR. This integration addresses the need, highlighted by Kaminski and Malgieri (2021), to “overcome the fragmentation of impact assessments in a holistic approach to risk management”. The framework provides specific methodologies for the identification of bias in datasets, using tools such as AI Fairness 360, and protocols for robustness testing against adversarial attacks, particularly relevant for generative systems exposed to manipulation attempts.

A distinctive element of the framework is the emphasis on human-in-the-loop mechanisms to ensure human oversight and explainability of automated decisions. As pointed out by Veale and Borgesius (2021), “generative systems require adaptive supervisory models that combine prior control and continuous monitoring”. The framework implements this principle through real-time monitoring dashboards that flag statistical anomalies, uncertainty thresholds for the activation of human channels, and “gradient” explainability mechanisms that provide different levels of explanatory detail depending on the criticality of the decision.

Cybersecurity is another pillar of the framework, with a focus on risks specific to generative systems. The model provides for a “security by design” approach with special emphasis on protection against targeted attacks such as data poisoning, which are particularly dangerous for LLM-based systems. Significant is the introduction of homomorphic encryption techniques that allow sensitive data to be processed without exposing them in the clear, thus responding to Floridi et al.“s (2020) recommendation to 'implement technical protections that guarantee privacy by design in public AI systems”. The framework also provides specific protocols for incident management that integrate GDPR requirements with those of the AI Act, ensuring timely notifications and structured postmortem analysis.

Finally, the framework devotes ample space to training and cultural change, recognising that the effective implementation of IAG requires not only technical tools but also adequate organisational skills. The model provides certified courses on AI Ethics, case simulations on real scenarios and departmental toolkits with specific guidelines, e.g. on prompt engineering for generative systems. This approach responds to Mergel et al.“s (2019) recommendation to 'invest in the development of soft skills that combine technological understanding and applied ethics”.

4. Implementation Challenges and Critical Success Factors

The implementation of Generative Artificial Intelligence systems in Public Administration presents distinctive challenges that require targeted strategies. The analysis of case studies and literature has identified five critical challenge areas: organisational resistance, skills shortage, integration into existing processes, algorithmic uncertainty management and economic sustainability. Each of these dimensions presents specificities related to IAG and requires differentiated management approaches.

Organisational resistance emerges as a significant barrier, manifesting itself in various forms: from distrust of automation of traditionally human functions to fear of loss of decision-making control. As highlighted by Mergel et al. (2019), "the introduction of AI systems in PA clashes with established organisational cultures and rigid bureaucratic structures". In the case of IAG, this resistance takes on peculiar characteristics related to the perceived unpredictability of generative systems. The analysis of the Italian municipal administration case showed that resistance was particularly pronounced for functions perceived as “creative” or requiring administrative discretion. Effective mitigation strategies included codesign approaches that involved officials in defining the operational parameters of the generative system and the gradual introduction of automation, while maintaining significant scope for human control.

The skills shortage represents a particularly acute challenge for IAG, which requires not only technical knowledge but also the ability to critically evaluate the generated outputs. Wirtz and Müller (2023) point out that “the internal digital divide within PAs is a greater obstacle to the adoption of AI than technological or economic constraints”. The cases analysed highlighted the need for a threefold level of skills: technical (understanding the mechanisms of operation), operational (knowing how to interact effectively with systems) and evaluative (knowing how to critically judge outputs). The framework addresses this challenge through multilevel training programmes, communities of practice between agencies and specific operational toolkits for different administrative functions.

Integration into existing processes is a crucial challenge, particularly complex for generative systems that may produce non standardised outputs. The case analysis of the Spanish government agency revealed difficulties in integrating the advanced chatbot with existing document management systems and established administrative workflows. Janssen and van der Voort (2020) emphasise the importance of “adaptive integration approaches that balance innovation and business continuity”. Effective strategies have included the adoption of standardised interfaces, the development of dedicated APIs and the incremental redesign of processes, initially maintaining parallel paths (automated and traditional) to ensure business continuity during the transition.

The management of algorithmic uncertainty represents a distinctive challenge of IAG, whose outputs may present varying degrees of accuracy, relevance and regulatory compliance. This challenge is particularly relevant in the public context, where administrative decisions must comply with stringent requirements of fairness, impartiality and traceability. The case of the Dutch ministry highlighted how the management of algorithmic uncertainty requires an articulated quality control strategy, combining pre-emptive filters (prompt engineering, pre-emptive checks), real-time monitoring (confidence scores, anomaly alerts) and post-generation human review. As pointed out by Floridi et al. (2020), “generative systems require a continuum of human supervision calibrated to the level of criticality of the function”.

Finally, economic sustainability represents a significant challenge, considering the costs of implementing, maintaining and updating generative systems. The case analysis showed that initial technology acquisition costs represent only a fraction of the total cost of ownership, with significant investments required for training, customisation, integration and continuous monitoring. The framework proposes models for sharing resources between administrations, incremental approaches to implementation and cost benefit evaluation methodologies that consider not only operational efficiency but also broader dimensions of public value such as transparency, accessibility and institutional trust.

5. Conclusions and Future Research Directions

The implementation of Generative Artificial Intelligence (GAI) in Public Administration represents a frontier of innovation with transformative potential, yet it is marked by complex challenges that require structured and context-sensitive approaches. The analysis of the Artificial Intelligence Framework for Public Administration has highlighted that effective adoption of GAI requires the integration of multiple dimensions: strategic governance, risk assessment, cybersecurity, and training each adapted to the specificities of generative systems and the particularities of the public sector context. The framework analysed offers a holistic approach that balances technological innovation, regulatory compliance, and the protection of fundamental rights, addressing the need to harmonize the requirements of the AI Act, GDPR, and international standards within an operationally sustainable model.

The comparative analysis between the ideal framework and actual implementation experiences revealed significant gaps, particularly in the areas of multistakeholder governance, risk assessment methodologies, "human-in-theloop" mechanisms, and cybersecurity practices. These gaps do not necessarily stem from shortcomings in the framework, but rather from the organizational, technical, and cultural complexities inherent in real-world applications. Their identification provides opportunities for refining the framework and developing more effective implementation strategies, as part of an iterative process of learning and adaptation.

The proposed policy recommendations offer a structured roadmap for public decisionmakers and implementers, articulated across three complementary levels: strategic (National Strategy, Competence Center, Regulatory Sandbox), tactical (sector-specific Implementation Toolkits, AI Evaluation Framework), and operational (Competency Certification, Collaborative Platform, Monitoring System). These recommendations do not represent definitive solutions, but rather guidelines for a transformation path that will require continuous adaptation to technological, regulatory, and organizational evolutions.

This research contributes to the literature on AI in Public Administration by partially addressing identified gaps, particularly regarding implementation frameworks for generative systems and the integration of regulatory requirements into operationally sustainable models. However, several aspects require further exploration through future research.

First, more robust methodologies need to be developed for empirically evaluating the impact of GAI in Public Administration, moving beyond anecdotal approaches to systematically measure effects on administrative efficiency, service quality, and institutional trust. As Wirtz and Müller (2023) note, "the lack of longitudinal empirical studies limits the understanding of the transformative effects of AI in the public sector"—a gap future research should seek to fill.

It is also crucial to deepen the understanding of organizational acceptance dynamics of GAI in Public Administration, identifying cultural, structural, and individual factors influencing resistance and adoption. Particularly relevant would be the analysis of interactions between generative systems and public workers in real contexts, examining how professional roles, skills, and work identities are reconfigured. Mergel et al. (2019) emphasize that "the transformation of professional roles represents one of the least explored dimensions of AI adoption in Public Administration," a gap that would benefit from ethnographic and longitudinal approaches.

A third line of research concerns governance models for GAI in Public Administration, with particular attention to mechanisms for democratic accountability and civic participation. As Veale and Borgesius (2021) highlight, "automated decision-making in the public sector raises fundamental issues of democratic accountability that go beyond purely technical or legal aspects," calling for interdisciplinary research that integrates legal, political, and sociological perspectives.

Finally, it is necessary to develop evaluation approaches that measure the impact of GAI not only in terms of efficiency but also with respect to broader public values such as equity, transparency, and legitimacy. Such research should adopt mixed methodologies, combining quantitative performance analysis with qualitative investigations into stakeholder perceptions and experiences, within a multidimensional and participatory evaluation approach.

In conclusion, GAI represents not only a technological challenge for Public Administration but also a broader opportunity for institutional rethinking, in which digital innovation and fundamental public values must be integrated into a new governance paradigm. As Janowski et al. (2022) stress, "the success of AI in the public sector will be measured not only in terms of operational efficiency, but in the ability to strengthen democratic legitimacy and institutional trust"—a perspective that should guide both practical implementation

Declaration on Generative AI

During the preparation of this work, the author used ChatGPT to translate some sentences. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

6. References

[1] Bannister, F., & Connolly, R. (2014). ICT, public values and transformative government: A framework and programme for research. Government Information Quarterly, 31(1), 119128.

[2] Chesterman, S. (2021). We, the Robots? Regulating Artificial Intelligence and the Limits of the Law. Cambridge University Press.

[3] Criado, J. I., & Gil Garcia, J. R. (2019). Creating public value through smart technologies and strategies: From digital services to artificial intelligence and beyond. International Journal of Public Sector Management, 32(5), 438450.

[4] Cugurullo, F. (2021). Urban artificial intelligence: From automation to autonomy in the smart city. Frontiers in Sustainable Cities, 3, 673033.

[5] de Bruijn, H., Janssen, M., & Warnier, M. (2020). Accountability challenges in the transnational algorithmic governance: The case of public sector algorithms. Public Administration Review.

[6] Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1).

[7] Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). How to design AI for social good: Seven essential factors. Science and Engineering Ethics, 26(3), 17711796.

[8] Janowski, T., Estevez, E., & Ojo, A. (2022). Digital government for sustainable development. Government Information Quarterly, 39(2), 101667.

[9] Janssen, M., & van der Voort, H. (2020). Adaptive governance: Towards a stable, accountable and responsive government. Government Information Quarterly, 37(1), 101411.

[10] Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. Government Information Quarterly, 37(3), 101493.