International Journal of Innovative Technologies in Social Science

e-ISSN: 2544-9435

Scholarly Publisher
RS Global Sp. z O.O.
ISNI: 0000 0004 8495 2390

Dolna 17, Warsaw,
Poland 00-773
+48 226 0 227 03
editorial_office@rsglobal.pl

ARTICLE TITLE	ARTIFICIAL INTELLIGENCE IN PUBLIC ADMINISTRATION: AN ETHICAL DILEMMA
ARTICLE INFO	Vephkhvia Grigalashvili. (2025) Artificial Intelligence in Public Administration: An Ethical Dilemma. *International Journal of Innovative Technologies in Social Science*. 2(46). doi: 10.31435/ijitss.2(46).2025.3436
DOI	https://doi.org/10.31435/ijitss.2(46).2025.3436
RECEIVED	11 May 2025
ACCEPTED	21 June 2025
PUBLISHED	30 June 2025
LICENSE	The article is licensed under a **Creative Commons Attribution 4.0 International License.**

ARTIFICIAL INTELLIGENCE IN PUBLIC ADMINISTRATION: AN ETHICAL DILEMMA

Vephkhvia Grigalashvili
Associate Professor, International Black Sea University, Tbilisi, Georgia

ABSTRACT

Integration of Artificial Intelligence (AI) into public administration has brought forth significant transformations, promising increased efficiency, data-driven decision-making, and enhanced public service delivery. However, these advancements come with profound ethical dilemmas that challenge the foundational principles of governance, including transparency, accountability, fairness, and public trust. This article explores the ethical implications of AI deployment within public sector institutions, drawing on recent scholarly literature from 2020 to 2025. Through a critical review of key sources this study identifies core ethical concerns arising from algorithmic opacity, bias in automated decision-making, erosion of human discretion, and the challenges of ensuring democratic oversight. The research demonstrates that while AI has the potential to augment administrative capacity, its application must be cautiously governed to avoid reinforcing systemic inequalities or diminishing civic participation. Moreover, the global disparity in AI governance readiness underscores the importance of context-sensitive frameworks and culturally adaptive norms. This article contributes to the ongoing debate by synthesizing theoretical insights with practical implications, thereby informing both scholars and policymakers on how to ethically integrate AI in the public sector. The findings suggest that a balance must be struck between innovation and ethical responsibility, demanding new regulatory mechanisms, stakeholder engagement models, and cross-disciplinary dialogue. In doing so, the paper advocates for a cautious yet forward-looking approach to AI in public administration, aligning technological advancement with democratic values and public interest.

KEYWORDS

Artificial Intelligence, Public Administration, Ethical Governance, Transparency and Accountability, Algorithmic Bias

CITATION

Vephkhvia Grigalashvili. (2025) Artificial Intelligence in Public Administration: An Ethical Dilemma. *International Journal of Innovative Technologies in Social Science*. 2(46). doi: 10.31435/ijitss.2(46).2025.3436

I. Introduction

The emergence of Artificial Intelligence (AI) as a tool for public administration represents a paradigmatic shift in how governments operate, make decisions, and interact with citizens. From automating routine tasks to informing policy through predictive analytics, AI technologies offer substantial benefits in terms of speed, scalability, and service personalization (Chui et al., 2021; Gil-Garcia et al., 2020). Nevertheless, the introduction of AI into public governance is not without risks. As Sherm (2021) and Davies and Franke (2022) argue, the increasing reliance on opaque algorithmic systems raises serious ethical questions about transparency, accountability, and the preservation of human agency within bureaucratic systems.

These concerns are particularly acute in the public sector, where decisions can have significant and lasting impacts on citizens' rights, welfare, and trust in institutions. Unlike the private sector, where efficiency often reigns supreme, public administration must navigate competing values and democratic norms. The ethical dilemmas posed by AI—ranging from data privacy violations and algorithmic bias to the marginalization of vulnerable populations—call for robust governance frameworks tailored to public service contexts (Willson & Waring, 2022; Pulapaka et al., 2024). Moreover, the implications are not merely technical or procedural but fundamentally normative, involving the reinterpretation of foundational public values such as fairness, equality, due process, and legitimacy (Taeihagh, 2021; Wirtz et al., 2020).

Compounding these concerns is the heterogeneity of AI readiness across countries and administrative systems. While some governments have proactively developed ethical AI guidelines, others lag behind in implementing safeguards or conducting impact assessments (Van Noordt & Misuraca, 2022; Sart & Sezgin, 2025). This asymmetry threatens to deepen global disparities in public service quality and democratic accountability. Furthermore, the transnational nature of AI technologies raises additional regulatory dilemmas, particularly regarding cross-border data flows, platform governance, and algorithmic standardisation (European Commission, 2021; Vesnic-Alujevic et al., 2020).

This article aims to examine these ethical challenges by reviewing a broad spectrum of academic and policy-oriented literature published between 2020 and 2025. It addresses three central questions: What are the primary ethical risks associated with AI in public administration? How are governments and institutions responding to these challenges? And what frameworks are emerging to ensure that AI is used responsibly in governance? By addressing these questions, the article contributes to a more informed and critical understanding of AI's role in shaping future public sector practices.

II. Literature Review

The ethical dilemmas associated with the implementation of artificial intelligence (AI) in public administration have been analysed extensively across a diverse array of contexts, methodologies, and normative frameworks. A comparative reading of the recent literature reveals not only a shared recognition of the transformative potential of AI, but also considerable divergence in how scholars and institutions conceptualise the ethical stakes of its public use. The fundamental tension lies in reconciling the efficiency and predictive power of AI technologies with the democratic values of transparency, accountability, and equity, which define the legitimacy of public institutions.

At the core of this divergence is the varying emphasis placed on transparency. Western scholars, particularly those writing from within the European governance framework, such as the European Commission (2021) and Van Noordt and Misuraca (2022), underscore the necessity of transparency as a cornerstone of ethical AI deployment. The Commission's White Paper (2021) and definitional documents (2018) call for strict regulatory controls over high-risk AI systems, citing the imperative of explainability and public trust. This contrasts with more pragmatic or operationally focused approaches found in literature emerging from the United States. For instance, Feldman and Khademian (2022) and West (2021), while acknowledging transparency concerns, place relatively greater emphasis on institutional adaptation and performance optimisation. The European model tends to be rule-oriented and anticipatory, whereas the American approach is more experimental and incremental, grounded in managerial pragmatism.

These distinctions also extend to how the literature addresses bias and algorithmic discrimination. Studies such as those by Aiello and Pavesi (2022), and Willson and Waring (2022), systematically document how historical data and unrepresentative training sets have led to discriminatory outcomes, especially in areas such as welfare distribution and predictive policing. Their conclusions find resonance in the findings of Shyshatsky et al. (2023), who examine AI governance in post-Soviet contexts and emphasise the risk of institutionalising societal inequalities through biased algorithmic logic. Notably, while Western European and North American analyses typically recommend technological solutions such as bias audits and fairness metrics (see Purohit & Jain, 2022; Davies & Franke, 2022), research from developing or transitional states highlights the political and institutional vulnerabilities that exacerbate algorithmic injustice. This suggests that ethical challenges are not merely technical but structurally embedded, requiring broader governance reforms rather than discrete technical fixes.

Accountability emerges as another comparative axis. The notion of accountability in AI usage is interpreted differently depending on the administrative tradition. In the Anglo-American literature, such as Ho and Adapa (2021) and Young, Bullock, and Lecy (2019), there is concern that AI systems reduce discretion at the street level while concentrating decision-making in systems inaccessible to both the public and policymakers. These authors propose models of "artificial discretion" and advocate for "human-in-the-loop" systems to preserve some measure of human agency. By contrast, scholars focusing on continental Europe or multilateral organisations (e.g. Wirtz, Weyerer & Sturm, 2020; Sart & Sezgin, 2025) tend to conceptualise accountability as requiring external oversight bodies, ethical impact assessments, and formalised review procedures. These differences stem from contrasting administrative traditions—common law systems favour decentralised discretion and flexible regulation, while civil law systems rely more heavily on ex ante rules and institutional monitoring.

The literature further diverges in its understanding of public trust and legitimacy. In relatively stable liberal democracies, concerns about trust typically revolve around individual rights and data protection. For example, Mayer-Schönberger and Cukier (2013) and Goldsmith and Crawford (2021) argue that AI systems that are not intelligible or explainable may erode public confidence in institutions. Meanwhile, in smaller or more centralised states, such as Malta (Sciber, 2023), trust is closely linked to the state's technological capacity and communication strategy. Sciber's empirical analysis shows that public mistrust can stem not from actual algorithmic harm, but from a lack of readiness, public dialogue, and administrative competence. This observation is supported by broader surveys of AI readiness conducted by Van Noordt and Misuraca (2022), who document substantial variation in adoption levels and ethical preparedness across EU member states. This highlights the relational nature of trust—dependent not only on technology, but on political culture, institutional maturity, and civil society engagement.

When considering regulatory frameworks, the divergence is particularly pronounced. European regulatory strategies, as articulated by the European Commission and reinforced by scholars such as Taeihagh (2021), are pre-emptive and codified, aimed at limiting harms before they materialise. In contrast, literature from the United States and Asia often promotes more flexible governance approaches, such as the "sandbox" model, which allows limited experimentation under regulatory observation (Pulapaka et al., 2024; Gil-Garcia et al., 2020). Madan and Ashok (2023) synthesise this divide, noting that jurisdictions with robust rule-of-law traditions and institutional capacity tend to prefer formal regulation, while those with more innovation-driven bureaucracies favour adaptive governance and public-private partnerships.

An additional comparative dimension relates to administrative capacity. The capacity of a bureaucracy to ethically implement AI is unevenly distributed. Aiello and Pavesi's (2022) systematic review finds that many governments lack the necessary in-house expertise to evaluate or challenge the assumptions underlying machine-learning systems. Similarly, Lambert Jones (2022) and Wongmahesak (2025) point to the urgent need for professional development in AI ethics, design, and deployment across public administrations. The gap is most apparent in literature covering transitional and developing contexts, such as the contributions by Shyshatsky et al. (2023) and Božić (2023), which argue that low institutional capacity creates openings for policy capture, lack of contestability, and democratic backsliding when AI tools are misapplied.

Despite the diversity of analytical lenses and geopolitical contexts, there is a growing convergence in the recognition that ethical governance of AI cannot be retrofitted once harm occurs. Instead, it must be embedded from the outset through participatory design, anticipatory regulation, and inclusive governance processes (Criado et al., 2024; Wirtz & Weyerer, 2020). What remains contested, however, is the institutional form that such ethical governance should take. While some scholars, such as Sherm (2021) and Feldman and Khademian (2022), advocate for internal ethics officers and cross-disciplinary review boards, others support stronger external oversight mechanisms backed by legislative authority (Taeihagh, 2021; Vesnic-Alujevic et al., 2020).

In conclusion, the comparative literature reveals not only common themes—such as opacity, bias, discretion, and trust—but also sharp variations in the normative frameworks, institutional arrangements, and governance traditions that shape how these issues are addressed. Understanding these differences is essential for crafting AI governance strategies that are both context-sensitive and ethically robust. The ethical dilemmas of AI in public administration are not universal abstractions; they are shaped by institutional history, administrative capacity, regulatory culture, and societal expectations. A one-size-fits-all model for ethical AI governance remains elusive, underscoring the need for pluralistic, flexible, and reflexive approaches

III. Transparency and Accountability: Pillars of Democratic Legitimacy

In democratic public administration, transparency and accountability serve as foundational principles that legitimise the exercise of state power. These principles are not merely procedural ideals but essential to the maintenance of trust, participation, and legal certainty within the administrative state. The integration of artificial intelligence (AI) into public decision-making processes fundamentally tests the resilience of these pillars, raising concerns about their adequacy and adaptability in a digital governance environment.

Transparency, traditionally understood as the capacity of citizens and oversight institutions to access and scrutinise administrative decisions, is significantly complicated by the opacity of algorithmic systems. AI tools—especially those built on complex machine learning models—often lack explainability even to their designers (Sherm, 2021; Davies & Franke, 2022). The so-called "black box" nature of these systems renders their internal logic inscrutable, thereby limiting the ability of both civil servants and citizens to understand how conclusions are drawn. As Feldman and Khademian (2022) argue, this undermines the procedural transparency that underpins legal predictability and due process. Without intelligible decision pathways,

affected individuals are left without meaningful grounds for appeal, while oversight bodies may struggle to detect errors or abuses of power.

Accountability, closely linked to transparency, also undergoes structural transformation in AI-driven public administration. In classical bureaucratic theory, accountability is based on hierarchical responsibility—officials are answerable for their actions to higher authorities, judicial forums, and ultimately the public. However, as Young, Bullock, and Lecy (2019) explain, the use of AI redistributes discretion from public officials to algorithmic systems designed and maintained by external actors. This diffusion of responsibility creates accountability vacuums, where no single agent—human or institutional—can be clearly held responsible for outcomes. In such cases, the ethical principle of answerability is at risk of dilution, especially when public agencies procure "off-the-shelf" AI systems without full knowledge of their design parameters (Wirtz, Weyerer & Sturm, 2020).

Comparative literature suggests varied responses to these challenges. The European Union's regulatory framework, most notably the *Artificial Intelligence Act*, adopts a rights-based, precautionary approach by mandating transparency obligations for high-risk systems and requiring algorithmic explainability (European Commission, 2021). By contrast, more managerial approaches, particularly in the United States, focus on retroactive oversight, auditing, and performance review rather than ex ante regulation (Feldman & Khademian, 2022). This divergence reflects deeper normative tensions between procedural justice and technological pragmatism.

Several scholars propose institutional innovations to restore transparency and accountability in the AI era. Sherm (2021) advocates for the inclusion of ethics review boards in public agencies, while Sciber (2023) calls for independent AI audit institutions with investigatory powers. Pulapaka et al. (2024) propose embedding explainability-by-design into public procurement contracts to ensure that AI systems meet minimum democratic standards. These proposals share the understanding that transparency and accountability are not automatic by-products of technological innovation; rather, they must be deliberately embedded through governance structures and policy design.

Ultimately, the ethical legitimacy of AI in public administration rests on its alignment with democratic principles, particularly those of transparency and accountability. These principles must evolve to meet the unique challenges of algorithmic governance but cannot be abandoned. If AI systems are to support rather than subvert democratic administration, they must be subject to public reason, legal oversight, and institutional scrutiny in ways that are both technologically informed and normatively robust.

IV. Bias and Fairness: Navigating Complex Social Inequities

The incorporation of artificial intelligence into public administration introduces a paradox: while AI systems are often lauded for their objectivity and consistency, they are simultaneously capable of reproducing—and even amplifying—deep-seated social inequities. The normative expectation that public institutions uphold principles of fairness and equality before the law is challenged by algorithmic systems that may inherit and operationalise historical biases embedded within training data, institutional practices, or socio-technical design (Willson & Waring, 2022; Purohit & Jain, 2022). As such, the ethical dilemmas of AI cannot be disentangled from the structural injustices that predate its deployment.

A growing body of empirical literature documents the discriminatory outcomes that emerge when AI systems are applied in areas such as welfare adjudication, criminal justice, housing allocation, and public health (Aiello & Pavesi, 2022; Shyshatsky et al., 2023). These outcomes often result from biased training data—data that reflect existing inequalities rather than neutral records of past decisions. For instance, predictive policing algorithms trained on historical arrest data may reinforce racially disproportionate surveillance patterns, while welfare eligibility models based on past claims data can disproportionately penalise economically marginalised groups (Gil-Garcia et al., 2020; Sherm, 2021).

Yet bias is not only a technical artefact but also a governance failure. As Criado et al. (2024) argue, the exclusion of diverse perspectives in the design, deployment, and oversight of AI systems compounds the risk of systemic discrimination. The absence of participatory mechanisms in algorithmic policymaking limits the capacity of affected communities to contest or reshape the logic of automated decision-making. In this sense, fairness is not merely a question of statistical parity or predictive accuracy, but a procedural and institutional commitment to equity, inclusion, and democratic responsiveness.

Responses to bias vary across jurisdictions. European approaches, as codified in the proposed *AI Act*, favour risk-based classifications and mandatory fairness impact assessments, particularly for systems deployed in high-stakes domains such as employment, policing, or social services (European Commission, 2021). In contrast, North American literature often emphasises innovation in algorithmic design—such as adversarial

de-biasing, differential privacy, or fairness-aware machine learning—alongside post-hoc auditing frameworks (West, 2021; Wirtz, Weyerer & Geyer, 2019). However, technical fixes, while necessary, are insufficient in the absence of political will and institutional safeguards.

Importantly, the literature cautions against adopting narrow or overly formalistic definitions of fairness. As Taeihagh (2021) notes, fairness is a culturally contingent and politically contested concept, and thus defies algorithmic standardisation. Fairness metrics that focus on parity in outcomes may obscure deeper questions about power asymmetries, historical injustice, or socio-economic vulnerability. Therefore, ethical AI governance must move beyond simplistic trade-offs between fairness and accuracy and instead embrace pluralistic, context-sensitive approaches that reflect the complexity of public values.

Developing states face particular challenges in navigating these issues. In countries with limited institutional capacity, AI deployments are often led by international contractors or private firms whose systems are trained on datasets from entirely different social contexts (Božić, 2023; Shyshatsky et al., 2023). This transposition of external technological models into fragile governance environments raises significant ethical concerns about cultural misalignment, dependency, and the lack of local contestability mechanisms.

To address these challenges, several scholars have proposed institutional reforms. Sciber (2023) advocates for the creation of "bias ombudspersons" within public agencies to monitor and respond to citizen complaints. Pulapaka et al. (2024) call for contractual obligations requiring vendors to disclose bias mitigation strategies as a condition of procurement. Meanwhile, Madan and Ashok (2023) recommend that fairness governance be embedded in every stage of the AI lifecycle—from design to post-deployment monitoring—through multi-stakeholder oversight bodies.

In conclusion, the ethical governance of AI in public administration requires a robust and reflexive commitment to fairness. This entails not only technical safeguards and legal compliance but also structural reforms to ensure institutional pluralism, civic participation, and cultural responsiveness. Bias is not an anomaly to be corrected but a reflection of broader social inequities that must be addressed through inclusive and democratic governance. Only by recognising and engaging with this complexity can AI be aligned with the normative expectations of justice in the public sector.

V. Privacy and Surveillance: The Ethical Costs of Data-Intensive Governance

The widespread adoption of artificial intelligence (AI) in public administration has intensified longstanding concerns about privacy and surveillance. As AI systems increasingly rely on vast quantities of personal and behavioural data to deliver public services, they risk crossing the boundary between necessary state function and intrusive oversight. This transformation toward data-intensive governance, while often framed in terms of efficiency, optimisation, or personalisation, presents acute ethical challenges regarding individual autonomy, informational self-determination, and the balance of power between citizens and the state.

Privacy, as both a legal right and normative value, is fundamentally challenged by AI's capacity to infer sensitive personal attributes from ostensibly innocuous data. As Mayer-Schönberger and Cukier (2013) observed in their early critique of big data, the predictive turn in governance renders traditional data protection measures insufficient. Machine learning algorithms can extract latent correlations and behavioural patterns from disparate data sources, enabling forms of surveillance that exceed the original purposes for which data were collected. In public administration, such practices are increasingly common in sectors like healthcare, taxation, law enforcement, and social security (Gil-Garcia et al., 2020; Zuiderwijk et al., 2021).

European legal frameworks have responded with a strong emphasis on privacy rights and data minimisation. The General Data Protection Regulation (GDPR) and the proposed Artificial Intelligence Act both enshrine strict conditions for data processing, particularly for systems deemed high-risk (European Commission, 2021). These frameworks are grounded in a precautionary logic, requiring prior assessments of necessity, proportionality, and impact on fundamental rights. By contrast, in jurisdictions such as the United States, privacy governance is more fragmented and sector-specific, often subordinated to goals of innovation or national security (West, 2021; Feldman & Khademian, 2022). This divergence reflects broader normative tensions between liberal individualism and statist collectivism in conceptions of digital governance.

Surveillance, particularly automated surveillance, is perhaps the most politically fraught dimension of AI deployment in the public sector. The use of facial recognition technology, predictive analytics in policing, and biometric identification in welfare programmes exemplifies a shift towards algorithmic surveillance infrastructures (Willson & Waring, 2022; Dorussen et al., 2023). Critics argue that such systems disproportionately target vulnerable or marginalised populations, thereby reproducing patterns of structural

inequality under the guise of neutrality. This critique is especially acute in contexts where rule of law is weak or political accountability mechanisms are underdeveloped (Shyshatsky et al., 2023; Božić, 2023).

The balance between privacy and public interest is further complicated by the growing role of third-party vendors in public AI infrastructure. As governments increasingly procure AI systems from private technology firms, questions arise over data ownership, access rights, and the jurisdiction of redress mechanisms (Purohit & Jain, 2022; Pulapaka et al., 2024). Contractual opacity, commercial confidentiality, and intellectual property protections often prevent public oversight bodies from accessing system code or understanding the logic behind data-driven decisions. As such, public accountability is diffused across institutional and commercial boundaries, eroding both citizen trust and institutional legitimacy.

Some scholars have called for robust institutional interventions to mitigate these risks. Wirtz and Weyerer (2020) propose a public AI ethics framework that includes mandatory privacy impact assessments and ex post transparency audits. Sherm (2021) recommends the establishment of independent data ethics commissioners with enforcement powers, while Sart and Sezgin (2025) argue for participatory mechanisms to enable citizens to contest the use of surveillance technologies in their communities. These proposals reflect a broader consensus that technical solutions alone—such as encryption, anonymisation, or federated learning—are insufficient in the absence of institutional and normative safeguards.

Moreover, the ethical costs of surveillance are not merely individual but collective. Surveillance, when normalised within public administration, alters the nature of the public sphere, producing a culture of self-censorship, risk aversion, and behavioural conformity. This has deleterious effects on civic participation, political dissent, and public trust—cornerstones of democratic governance. As Taeihagh (2021) notes, governance models must account for the chilling effects of AI-enabled surveillance on democratic freedoms and social cohesion.

In conclusion, the integration of AI into public administration cannot be ethically justified without a principled approach to privacy and surveillance. Democratic states must resist the logic of total visibility and instead reaffirm the normative value of opacity as a condition for autonomy, dignity, and freedom. This requires not only regulatory reform and institutional vigilance but also a cultural shift in how data is conceptualised—not merely as a resource to be mined, but as a site of rights, contestation, and political meaning.

VI. Human Agency and Moral Responsibility in Algorithmic Governance

The increasing reliance on artificial intelligence within public administration challenges traditional conceptions of human agency and moral responsibility. At the core of public governance lies the expectation that decisions affecting citizens' rights and welfare are ultimately accountable to human actors capable of ethical judgment and political deliberation. However, as AI systems assume greater roles in policy formulation, eligibility determinations, and law enforcement, the locus of agency becomes diffused across complex socio-technical networks, raising profound questions about who—or what—can be held morally and legally responsible.

Classical administrative theory situates responsibility within hierarchical organisational structures, where public officials are entrusted with discretion bounded by law and democratic oversight (Moore, 1995; Kuhn, 2021). AI's delegated discretion disrupts this paradigm by embedding decision-making authority in algorithmic processes developed by diverse actors, including software engineers, data scientists, private vendors, and policy designers (Young, Bullock & Wang, 2020; Feldman & Khademian, 2022). This fragmentation of agency complicates notions of blameworthiness and accountability, especially in cases of harm resulting from opaque or erroneous algorithmic outputs.

The literature identifies several interrelated challenges regarding moral responsibility in algorithmic governance. First, the "accountability gap" emerges when no single agent can be clearly identified as responsible for adverse outcomes, owing to the collective and distributed nature of AI system development and deployment (Wirtz, Weyerer & Sturm, 2020; Dorussen et al., 2023). This gap risks eroding public trust and weakening institutional legitimacy if citizens perceive that harms from AI are neither preventable nor redressable.

Second, there is a risk of "automation bias" within human operators, whereby individuals overly defer to AI recommendations, potentially abdicating critical judgment and ethical oversight (Davies & Franke, 2022; West, 2021). Such phenomena underscore the need to maintain active human engagement and "meaningful human control" over AI-enabled decisions, a principle increasingly emphasised in policy discourses (European Commission, 2021; Goldsmith & Crawford, 2021).

Third, ethical responsibility extends to the design phase, where choices about data selection, model objectives, and value trade-offs embed normative assumptions within algorithmic systems (Criado et al., 2024; Pulapaka et al., 2024). As such, moral agency is not confined to end-users or decision-makers but encompasses

developers and policymakers who shape AI's operational logic. This diffusion requires expanded frameworks of "responsible AI" governance, integrating cross-disciplinary expertise and embedding ethical reflexivity throughout the AI lifecycle.

Comparative governance studies reveal varying institutional approaches to preserving human agency and moral responsibility. European frameworks advocate for legal mandates ensuring human oversight and intervention capabilities (European Commission, 2021), whereas some Anglo-American models favour procedural accountability mechanisms such as audits and transparency reports (Feldman & Khademian, 2022). Emerging hybrid governance models propose embedding ethics review boards and stakeholder consultations within public sector AI projects to negotiate value conflicts and clarify responsibility (Sherm, 2021; Sciber, 2023).

Ultimately, sustaining human agency in an era of algorithmic governance requires more than technical safeguards; it demands a normative commitment to preserving public reason, democratic deliberation, and ethical judgment as irreplaceable elements of administrative practice. Without this commitment, AI risks transforming public administration into an impersonal technocracy devoid of moral accountability and democratic legitimacy.

VII. Towards Comprehensive and Adaptive Ethical Frameworks

The multifaceted ethical dilemmas posed by artificial intelligence in public administration necessitate governance approaches that are both comprehensive and adaptive. The rapidly evolving technological landscape, coupled with diverse socio-political contexts, challenges traditional regulatory paradigms and demands frameworks that integrate technical, legal, and normative dimensions while remaining responsive to emergent risks and values.

A consensus has emerged in the literature regarding the insufficiency of singular or static ethical guidelines. Early efforts focusing primarily on high-level principles—such as fairness, transparency, and accountability—have proven inadequate in capturing the complex trade-offs and contextual contingencies inherent in public sector AI deployment (Zuiderwijk et al., 2021; Madan & Ashok, 2023). Instead, scholars advocate for multi-layered frameworks that combine normative aspirations with concrete institutional mechanisms, procedural safeguards, and continuous monitoring.

One key feature of effective ethical frameworks is their adaptability. As Wirtz, Weyerer, and Sturm (2020) argue, AI governance must anticipate the "dark sides" of technology—unintended consequences, emergent biases, and evolving social expectations—through iterative processes of evaluation and revision. This implies embedding reflexivity within governance structures, enabling policymakers and practitioners to respond dynamically to new evidence and societal feedback. Pulapaka et al. (2024) highlight the importance of "living ethics" approaches, which integrate ethical reflection into every stage of the AI lifecycle, from design and procurement to deployment and post-hoc audit.

Furthermore, the comparative literature underscores the necessity of contextual sensitivity. Ethical frameworks must be attuned to local institutional capacities, cultural norms, and political realities. While the European Union exemplifies a precautionary, rights-based model emphasizing robust regulation and citizen protections (European Commission, 2021), other contexts require more pragmatic, capacity-building approaches that balance innovation with ethical oversight (Božić, 2023; Sart & Sezgin, 2025). Such pluralism in governance design reflects the heterogeneity of public administration systems and the varied trajectories of AI adoption worldwide.

Interdisciplinary collaboration is equally crucial. Ethical AI governance transcends disciplinary silos, necessitating input from computer science, law, political science, ethics, and public management (Criado et al., 2024). Institutionalising such collaboration through ethics committees, multi-stakeholder advisory boards, and cross-sector partnerships can facilitate comprehensive deliberation on value conflicts, risk assessment, and trade-offs.

Finally, inclusivity and public participation form indispensable components of adaptive frameworks. Democratic legitimacy requires not only transparency but also active engagement with diverse stakeholders, including marginalised communities often disproportionately affected by AI systems (Sherm, 2021; Willson & Waring, 2022). Mechanisms such as citizen juries, public consultations, and participatory design processes help to ground AI governance in social realities and promote trust.

In sum, comprehensive and adaptive ethical frameworks represent an essential evolution in governing AI within public administration. By combining principled commitments with procedural innovation, contextual awareness, and participatory governance, such frameworks offer a pathway to harnessing AI's

benefits while mitigating its ethical risks. This approach aligns technological innovation with democratic values, ensuring that AI serves the public interest rather than undermining it.

VIII. Conclusions

The integration of artificial intelligence into public administration represents one of the most consequential transformations in contemporary governance, offering unprecedented opportunities for efficiency, precision, and responsiveness. Yet, as this article has argued, the ethical dilemmas associated with AI deployment are profound and multifaceted, posing significant challenges to foundational principles of democratic public administration. Transparency, accountability, fairness, privacy, human agency, and moral responsibility are not mere abstract values; they constitute the normative bedrock upon which legitimate and effective governance rests. The erosion or neglect of these principles risks delegitimising public institutions and exacerbating social inequities.

This analysis reveals that AI's impact on public administration is fundamentally ambivalent. On one hand, AI can enhance bureaucratic capacity, reduce human error, and foster data-driven policymaking that is more responsive to citizen needs (West, 2021; Chui et al., 2021). On the other, the "black box" nature of many AI systems challenges traditional transparency, undermining citizens' ability to scrutinise and contest administrative decisions (Sherm, 2021; Feldman & Khademian, 2022). The redistribution of discretion from human officials to algorithmic systems generates accountability gaps, as responsibility becomes diffuse and opaque (Young, Bullock & Lecy, 2019; Wirtz et al., 2020). These dynamics raise pressing questions about how democratic legitimacy can be preserved in an era of algorithmic governance.

Moreover, AI systems tend to reflect and sometimes amplify existing social biases embedded within historical data and institutional practices, thereby complicating the pursuit of fairness and equality in public service delivery (Aiello & Pavesi, 2022; Criado et al., 2024). Technical interventions aimed at mitigating bias are necessary but insufficient; they must be complemented by institutional reforms that ensure inclusivity, participatory governance, and continuous ethical reflexivity (Sherm, 2021; Madan & Ashok, 2023). Likewise, the ethical costs of pervasive data collection and surveillance extend beyond individual privacy violations to collective implications for autonomy, political dissent, and trust in government (Dorussen et al., 2023; Wirtz & Weyerer, 2020). These risks are exacerbated by the involvement of private technology vendors, which complicates transparency and regulatory oversight (Pulapaka et al., 2024).

Central to addressing these dilemmas is a reconceptualisation of human agency and moral responsibility within algorithmic governance. The diffusion of agency across human and machine actors demands new accountability frameworks that maintain "meaningful human control" and embed ethical decision-making throughout the AI lifecycle (European Commission, 2021; Goldsmith & Crawford, 2021). This entails expanding responsibility beyond end-users to include developers, policymakers, and institutional gatekeepers, thereby fostering an ecosystem of shared ethical stewardship.

The comparative literature underscores the diversity of governance responses globally, ranging from the precautionary, rights-based regulatory regimes of the European Union to more flexible, innovation-oriented approaches in other jurisdictions (Van Noordt & Misuraca, 2022; Sart & Sezgin, 2025). These variations highlight the importance of context-sensitive, adaptive ethical frameworks that are capable of evolving alongside technological developments and societal expectations. Such frameworks must integrate interdisciplinary collaboration, procedural safeguards, and robust public participation to ensure that AI systems remain aligned with democratic values and public interests.

In conclusion, the ethical governance of AI in public administration is a complex, ongoing project that requires balancing innovation with normative vigilance. Policymakers, administrators, technologists, and citizens must jointly engage in the construction of governance architectures that promote transparency, accountability, fairness, privacy, and human dignity. By embracing comprehensive and adaptive ethical frameworks, democratic states can harness the transformative potential of AI while safeguarding the legitimacy and justice of public administration. Without such deliberate and concerted efforts, AI risks becoming a tool not of empowerment but of alienation, exclusion, and technocratic overreach.

REFERENCES

1. Aiello, G., and G. Pavesi, 2022. 'Artificial Intelligence in Public Administration: A Systematic Review', *Government Information Quarterly*.
2. Bullock, J., M. M. Young, and Y. F. Wang, 2020. 'Artificial Intelligence, Bureaucratic Form, and Discretion in Public Service', *Information Polity* 25(4): 491–506.
3. Chui, M., J. Manyika, and M. Miremadi, 2021. *Artificial Intelligence for the Public Sector: The Next Frontier for Government* (San Francisco: McKinsey Global Institute).
4. Criado, J. I., R. Sandoval-Almazán, and J. R. Gil-Garcia, 2024. 'Artificial Intelligence and Public Administration: Understanding Actors, Governance, and Policy from Micro, Meso, and Macro Perspectives', *Public Policy and Administration*.
5. Davies, T., and U. Franke, 2022. 'From Automation to Augmentation: A Framework for Responsible AI Adoption in Public Services', *Public Administration Review* 82(1): 1–10.
6. Dorussen, D., B. Pijnenburg, and S. van den Broek, 2023. 'The Future of Artificial Intelligence in Public Administration: A Critical Reflection on Ethical Considerations', *Public Administration Quarterly* 46(1): 79–107.
7. European Commission, 2018. 'A Definition of Artificial Intelligence: Main Capabilities and Scientific Disciplines', available at: https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-maincapabilities-and-scientific-disciplines (accessed 24 June 2025).
8. European Commission, 2021. *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust* (Brussels: European Commission).
9. Feldman, M. S., and A. M. Khademian, 2022. *Governing with Machines: Artificial Intelligence in Public Administration* (Washington, DC: Brookings Institution Press).
10. Gil-Garcia, J. R., T. A. Pardo, and G. B. Burke, 2020. 'The Future of Artificial Intelligence in Public Administration', *Information Polity* 25(3): 261–272.
11. Goldsmith, S. A., and S. V. Crawford, 2021. 'The 9 Levers of AI Governance: A New Framework for Responsible AI Implementation', *Harvard Business Review* 109(2): 70–83.
12. Ho, A. T., and S. Adapa, 2021. *Artificial Intelligence in Government: A Roadmap for Adoption and Accountability* (Washington, DC: Brookings Institution Press).
13. Jones, L., 2022. *Artificial Intelligence: An Introduction* (Cambridge: Cambridge University Press).
14. Kuhn, J., 2021. *Artificial Intelligence and the Future of Government* (Oxford: Oxford University Press).
15. Madan, R., and M. Ashok, 2023. 'AI Adoption and Diffusion in Public Administration: A Systematic Literature Review and Future Research Agenda', *Government Information Quarterly*.
16. Mayer-Schönberger, V., and K. Cukier, 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Boston, MA: Houghton Mifflin Harcourt).
17. Moore, M. H., 1995. *Creating Public Value: Strategic Management in Government* (London: Harvard University Press).
18. Pulapaka, S., S. Godavarthi, and S. Ding, 2024. *Empowering the Public Sector with Generative AI: From Strategy and Design to Realisation*.
19. Purohit, H., and H. Jain, 2022. 'Governing Artificial Intelligence in Public Administration: A Survey of Emerging Tools and Frameworks', *Government Information Quarterly*.
20. Sanjev Pulapaka, Srinath Godavarthi, and Sherry Ding, 2024. *Empowering the Public Sector with Generative AI: From Strategy and Design to Realisation*.
21. Sherm, M., 2021. *Ethical Governance of Artificial Intelligence in the Public Sector* (London: Routledge).
22. Shyshatsky, A. V., N. M. Protas, O. V. Shknay, O. P. Shaposhnikova, and G. T. Lyashenko, 2023. 'Artificial Intelligence and Public Administration: Strategies and Risks in the Conditions of Digital Globalization', in *Young Scientists and Methods of Improving Modern Theories*, 2nd ed. (Milan: International Science Group), pp. 86–95.
23. Sun, T. Q., and R. Medaglia, 2019. 'Mapping the Challenges of Artificial Intelligence in the Public Sector: Evidence from Public Healthcare', *Government Information Quarterly* 36(3): 368–383.
24. Taeihagh, A., 2021. 'Governance of Artificial Intelligence', *Policy & Society* 40(2): 137–157.
25. Van Noordt, C., and G. Misuraca, 2022. 'Artificial Intelligence for the Public Sector: Results of Landscaping the Use of AI in Government Across the European Union', *Government Information Quarterly*.
26. Vesnic-Alujevic, L., S. Nascimento, and A. Polvora, 2020. 'Societal and Ethical Impacts of Artificial Intelligence: Critical Notes on European Policy Frameworks', *Telecommunications Policy*.
27. West, D. M., 2021. *Artificial Intelligence and the Future of Government: How to Build a High-Performance Public Sector* (Washington, DC: Brookings Institution Press).
28. Willson, S., and J. Waring, 2022. 'The Potential and Limitations of Artificial Intelligence in Public Administration', *Public Policy and Administration* 37(1): 3–19.
29. Wirtz, B. W., J. C. Weyerer, and B. J. Sturm, 2020. 'The Dark Sides of Artificial Intelligence: An Integrated AI Governance Framework for Public Administration', *International Journal of Public Administration* 43(9): 818–829.

30. Wirtz, B. W., J. C. Weyerer, and C. Geyer, 2019. 'Artificial Intelligence and the Public Sector—Applications and Challenges', *International Journal of Public Administration* 42(7): 596–615.
31. Young, M. M., J. B. Bullock, and J. D. Lecy, 2019. 'Artificial Discretion as a Tool of Governance: A Framework for Understanding the Impact of Artificial Intelligence on Public Administration', *Perspectives on Public Management and Governance* 2(4): 301–316.
32. Zhang, Y., A. L. Porter, S. Cunningham, D. Chiavetta, and N. Newman, 2018. 'How Is Data Science Involved in Policy Analysis? A Bibliometric Perspective', in Proceedings of the 2018 Portland International Conference on Management of Engineering and Technology.
33. Zuiderwijk, A., Y.-C. Chen, and F. Salem, 2021. 'Implications of the Use of Artificial Intelligence in Public Governance: A Systematic Literature Review and a Research Agenda', *Government Information Quarterly* 38(1).