Ethical Frameworks and Governance for Generative Artificial Intelligence in Public Services: A Multidimensional Analysis

Ting Sing Hoi[1]*, Tan Shi Yee[2], Abhijit Ghosh[3], Normal Mat Jusoh[4]

[1]Azman Hashim International Business School, Universiti Teknologi Malaysia, Malaysia

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.100601247

Received: 02 July 2026; Accepted: 07 July 2026; Published: 16 July 2026

ABSTRACT

The public sector is coming into its own as new opportunities for Generative Artificial Intelligence (GenAI). It can help to streamline tasks, enhance interactions with citizens, and aid administrative processes. Meanwhile, there are significant issues of bias and privacy, transparency and accountability and absence of human judgment in day-to-day decisions. While there are technical solutions, there is no workable governance model in the public sector that meets its needs. This is the purpose of this research: to fill in this gap and offer a multidimensional approach to governance of GenAI in public services. It reviews 12 major AI governance frameworks, which were published between 2018 and 2026, and are based on the public values framework of Bozeman, the concept of digital bureaucracy of Bovens and the administrative values of Hood. The analysis is related to the technical, institutional and ethical aspects of these materials that are combined. Based on the results, it is found that the benefits of GenAI can only be realized with a Three-Dimensional Governance Model, which is a Layered Responsibility Framework between regulators, GenAI developers, public organizations and frontline officials. Then, the model is extended to civil administration, health and education. At the heart of the recognition is a recognition of long-term practical protection (such as the right to intervene in automated systems), the right to appeal decisions made by an AI and requiring human oversight. These safeguards build on each other and provide a credible alternative to ensure that GenAI supports the public interest and is held accountable and trusted.

Keywords: Generative Artificial Intelligence (GenAI), Algorithmic Accountability, Public Values, Ethical Governance, Public Services

INTRODUCTION AND BACKGROUND

Generative Artificial Intelligence (GenAI) is driving a significant technology shift in the public sector. In the past, most digital tools were employed to sort data, make predictions or process information in regular administrative situations. However, GenAI can create text, write code, respond to questions, and aid in complex decision making at scale. This makes governments much more opportunities to speed up the services delivery, cut down on repetitive administration work and enhance document management. However, there are institutionally and ethically pertinent challenges that cannot be ignored at the same time.

This change can be explained within the framework of Christopher Hood's (1991) administrative values approach. Hood separates theta type values (honesty, fairness, and prevention of administrative injustice) from sigma type values (efficiency, speed and cost reduction). GenAI is often adopted because of its sigma-type benefits. Rather, if public agencies deal quickly and efficiently rather than fairly, transparently and in accordance with due process, they undermine the very public values that legitimate administrative action. That's where the majority of governance issues start to appear today.

This basic administrative conflict goes beyond the operational concerns and serves as a challenge to the normative bases of public governance. In order to adequately assess this imbalance in a systematic way, Bozeman's (2007) public values framework should be taken into consideration alongside Hood's administrative

values framework. Bozeman believes that public values are the moral or ethical rights, duties and principles that should be upheld in a democratic state. These values in the context of digital government are non-discrimination, administrative accountability, procedural fairness and citizen dignity. Public agencies that blindly take the least-squares value of a sigma-type measure, generated by GenAI, could be part of what Bozeman calls a public values failure, in which efficiency-oriented and vendor-focused incentives start to supersede democratic values. This poses threats to transparency, accountability and public confidence in digital governance.

Emergence of Genai and Discretionary Displacement

The change is important, because it is not merely a change of procedure. Discretion is where general rules are modified to fit individual situations, in many public services. A citizen's case could be one in which records are incomplete, or the evidence is conflicting, hardship exists, language barriers are an issue, or local conditions are not evident in a database. The context factors may not be as significant when the actual starting point for action is the GenAI generated summary and recommendation. It's not if officials are being assisted by technology, it's if technology has become an addiction.

The impact of GenAI on administrative discretion is one of the biggest impacts of GenAI in the public sector. Bovens and Zouridis (2002) explain that the street level bureaucracy that requires decisions to be made by these first-line officials is being replaced by system level bureaucracy, in which decisions and rules become trapped in digital systems. GenAI goes even further; it's possible to use it to handle tasks that used to require human involvement and context.

Understanding rules of eligibility, policy notes, case file summary can be done by GenAI, and clinical text can be generated. However, it not only helps public officials. It can also be a stand-in for some of their judgements of commercial systems' output. As the officials' workload increases, the pressure of time increases, or trust in the system increases; their own discretion decreases. This may change the balance between state and citizen over time and increase the lack of accountability of human decision making.

It is not the only occasion when man's judgment will fail. It can build up overtime and can be referred to as ‘ethical debt' in this study. Just like technical debt, ethical debt has further-reaching social and institutional consequences, typically over a longer timeframe, when an organization decides to focus on short-term efficiency instead of public values. Algorithmic outputs, not human discretion, are used by public agencies, which results in a postponement of hard decisions over ethics, not a solution. What we have are more responsibility issues, equity and trust issues which are more difficult to rectify afterward.

Empirical Testbeds of the Study

The authors seek to illustrate this with three use cases in which public service is making sense with GenAI or soon will. These are actual test cases to be put into practice to test the suggested governance model outlined in this paper.

The importance of civil services and administration procurement in public administration is great due to its impact on resource allocation, supplier participation and citizens' choices. AI is already leveraged for creating procurement specs, collating policy submissions and answering public enquiries. The problem is that administrative bias can be continued, opportunities may be denied to those who are working as vendors, or it can affect the public resources' decisions without sufficient review.

Public healthcare and clinical informatics are a high-stakes environment, mistakes can have direct impact on the patient. These are just a few of the ways health-care specific GenAI tools might be applied, but each raises privacy, hallucination, accuracy and commercialization concerns over patient-sensitive public health data.

Education and pedagogical systems are also important in that they mold learning, judgements and personal development in school and at university. Although GenAI can assist with lesson planning, grading, and providing feedback to students, it also has the potential to create dependency, bias in assessment, and the commercialization of student learning data. Thus, it is necessary to have strong protection of cognitive agencies and developmental

privacy in education.

The main aim of this research is to investigate these three areas and uncover the risks associated with the uncritical use of GenAI and the governance response. The paper outlines the core ethical issues, details the methodology used for the review of the available models and finally presents a practical governance model, and sector specific tips and policy recommendations.

Core Ethical Dilemmas and Risks

Before we can discuss these ethical issues and systemic risks, we must first explain the key ethical issues and systemic risks that come with using AI in the public sector. Establishing a robust governance model requires a foundational understanding of potential ethical dilemmas and systemic risks in the adoption of AI in public administration. The first step towards a robust governance model is to identify the ethical challenges and systemic risks that can arise from the integration of AI into public administration. Such risks are not simply technical problems, but structural problems and should be considered in this light and not as instances of a technical glitch. These risks should be considered structural problems, not just technical glitches, and should be analyzed as such considering the misalignment of technical parameters, institutional contexts and ethical values. To show how these systemic risks can be interrelated, the integration of Generative AI is depicted as a socio-technical feedback loop, with private infrastructure dependencies extracting public data, which leads to discretionary displacement, building ethical debt in the long term.

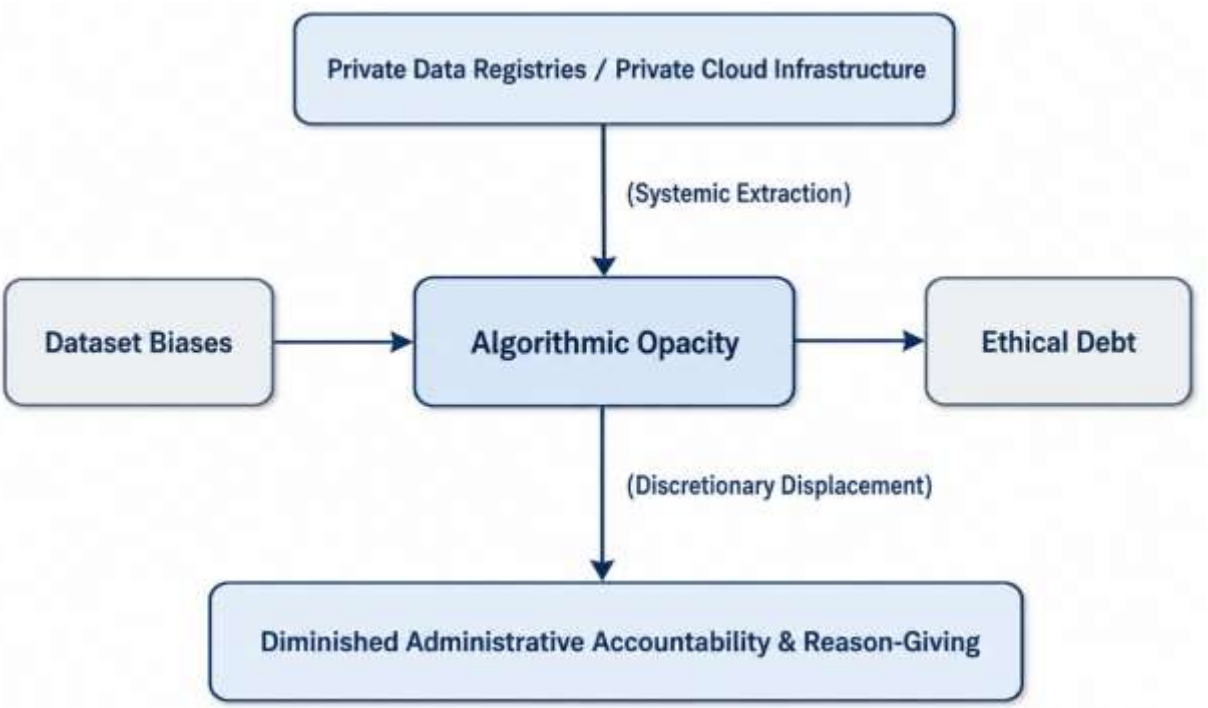

Figure 1. The Socio-Technical Cycle of Generative AI Integration: From Systemic Extraction to Discretionary Displacement

Dataset Bias and Systemic Administrative Inequalities

Public resource allocation, in the ethical sphere of public administration, must be completely nondiscriminatory and fair. However, generative models are just a mirror image of the data on which they have been trained. If the pretraining datasets contain historical, socio, political and structural biases, they convey and reinforce these biases as an objective mathematical neutrality.

The use of biased generative models in the public procurement and social service evaluation processes can lead to a disadvantage for minority groups, local businesses and marginalized groups, and be used to summarize the vendors' proposals. These historical biases lie in a nonlinear, multivariance weighing the neural network

architecture, and are therefore not apparent when using traditional administrative reviews. However, if they are not handled carefully and with caution, and if they are instead applied within the framework of public policy, they can institutionalize historical inequities and create historical trends of structural discrimination into an automated, self-renewing public policy that could be detrimental to vulnerable populations at the Base of the Pyramid. Their skewed and distorted calculated products are inputs to public administrative records, which in turn become inputs to other administrative processes, leading to a socio-technical cycle that reinforces inequality over time and makes it harder to correct the inequality using human efforts in later automation processes.

The Dilemma of Reidentification and Patient Data Sovereignty

One of the most obvious institutional threats to public sector generative AI implementation is the risk of breaching the privacy of citizen information. In the case of modern generative models, they are trained on large amounts of data and, when used in the public service, can handle a large amount of personal and sensitive information. However, in healthcare, this flaw is more critical since public clinical systems frequently hold electronic health records, medical history, treatment data, demographics, and other delicate data on patients. If commercial or cloud-based generative models are employed to summarize patient records or assist with clinical documentation, information about patients may be shared outside of public health institutions' direct management.

Two big governance issues arising from this operational path are patient re-identification and data sovereignty. Generative systems can pose reidentification risks even if clinical datasets are anonymized or deidentified, as they can still be matched with other datasets in the external environment or with other records available, for example when the systems leverage the same patient ID. This is particularly problematic in new public health informatics settings where public institutions may rely on external cloud systems and secure localized infrastructures may be limited. In these environments, vulnerable patient populations could be more vulnerable to data re-identification, commercial data profiling and loss of control over sensitive health information.

Hence, patient data sovereignty should be considered as a public governance matter, instead of being a cybersecurity problem. To support generative AI systems in public healthcare institutions, it is essential to have secure, localized and auditable data architectures. Clinical data must not be sent to external systems without clear safeguards, legal accountability, human oversight and clear restrictions on data-use. If a patient's trust is eroded, and if these technologies are developed in isolation by private companies, there is a risk that they will become tainted and will lead to long-term institutional reliance on one vendor or another. This underscores local data protection, federated learning, and robust access control and compulsory human validation prior to the input of AI-generated clinical output into official health records.

Opacity, Black Box Architectures and the Crisis of Reason Giving

The law and ethics of reason-giving is a key element of the legitimacy of administrative action in democratic societies. Within Mashaw's (1983) notion of "bureaucratic justice," a decision is legitimate to the extent that it can be justified by a logical and rational progression of thought sustained by legal guidelines and values of public policy. In the same vein, Bovens' (2007) theory of public accountability puts forward the idea that public administrators are expected to be held accountable to the people by being open about their actions and offering clear, comprehensible and contestable reasons for what they do.

The use of very complex, and often uninterpretable, deep neural networks poses a significant challenge to this model of accountability. Some of the latest AI systems, especially those of the generative variety, are extremely large, and many models today use billions or even trillions of parameters to process and produce natural language. The level of complexity makes the internal logic of the model hard to understand, and the resulting “black box” effect makes it hard to explain with the traditional administrative audit trails.

This poses a particular challenge when dealing with decision-making related to social services benefits, procurement assessment, medical paperwork, schooling, licensing or public contracts. In such situations, it might not be possible to explain, either legally or in a coherent logical way, the source of a particular AI-generated output from public officials. This means that citizens might not have a foundation for understanding, challenging

or questioning automated public actions.

This opacity of the structures undermines the principle of giving reasons for administration and leads to a crisis of administrative legitimacy. If without justification, the state has to rely on automatically generated outputs, then there's an epistemic gap between the decision-making system and the affected citizen. This disparity can have a negative impact on the procedural justice of citizens, and on a form of epistemic injustice whereby citizens are set up against a wall of power by being unable to meaningfully contest decisions that impact on their rights, access to services or public entitlements.

Methodology: Integrative Conceptual Synthesis

The study utilizes an Integrative Literature Synthesis methodology (Torraco, 2005; Whittemore & Knafl, 2005) to integrate and synthesize all the different and disjointed frameworks into a cohesive public sector roadmap. In contrast to other fields where statistical meta-analysis has been conducted, the study of ethical and operational governance of Generative AI is still in its infancy and is represented in several disciplines of computer science and administrative law and public policy. Rather, the integrative synthesis enables a conceptual integration of different types of evidence, such as empirical studies, international ethical frameworks and the normative administrative theories, which yields a multi-faceted, comprehensive model.

Our research process consists of three-phases of analytical process as shown in Figure 2. This systematic pipeline builds on PRISMA guided corpus selection, towards multidimensional coding, leading to the theoretical reconstruction of an integrated socio-technical model.

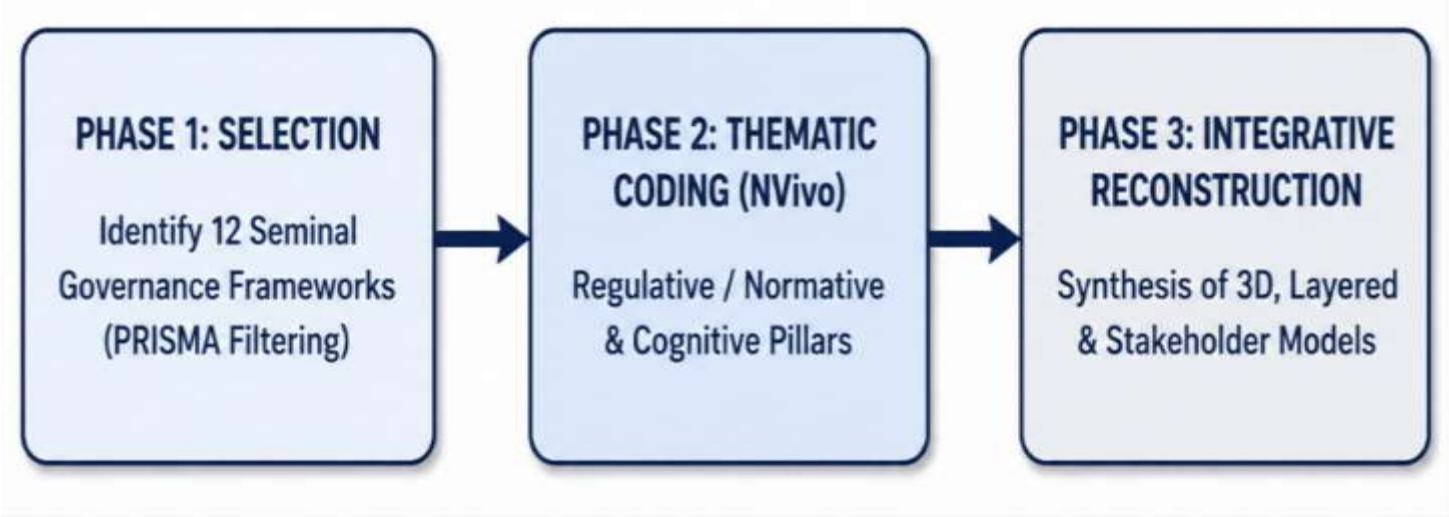

Figure 2. The Three-Phase Methodological Pipeline for Integrative Synthesis

FRAMEWORK SELECTION AND STRATEGY

A selection of twelve authoritative international, national and domain-specific AI ethics frameworks and public sector laws and policies from 2018 to 2026 were chosen and analyzed for establishing systemic validity. The frameworks were screened following the Preferred Reporting Items for Systematic Reviews and MetaAnalyses (PRISMA) and were included in the study only if they contained three items: (1) They were written by a national, regional or professional governing body recognized by experts; (2) They contained explicit details on how to use the automated systems; and (3) They contained explicit mention of public service allocation or clinical/pedagogical decision support.

As a way of being transparent with the process and for replicability of the research, we conducted extensive literature searches in all the major interdisciplinary databases, such as Scopus, Web of Science and HeinOnline, and official institutional repositories. The search strategy used involved using the combinations of targeted Boolean search terms, which included (Generative AI or Large Language Models or automated decision making) and (public sector or public service or governance) and (ethics or accountability). The first search strategy

returned a total database return of candidate frameworks (N=47). After systematic screening of these documents, a multi-step screening funnel was used, excluding N=35 documents, which did not have any concrete administrative guidelines or any domain specific clinical/pedagogical execution paths. This systematic filtering process was able to produce the final analytical corpus (N=12 frameworks).

A sample of 12 complementary governance sources were chosen. It featured the European Union Artificial Intelligence Act, the UNESCO Recommendation on the Ethics of Artificial Intelligence, the OECD Principles on Artificial Intelligence, the United States Executive Order on Trustworthy Development and Use of Artificial Intelligence, the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, and the World Health Organization guidance on ethics and governance of AI for health. The HAIR model of pedagogical governance, the framework for government's use of Generative Artificial Intelligence in the United Kingdom, the draft continental strategy on the use of AI in Africa by the African Union, the Singapore Model AI Governance Framework, the Montreal Declaration for a Responsible Development of Artificial Intelligence, and the Tanzanian National Digital Health Strategy and its EHR governance guidance were also covered in the corpus. These documents provided a general, but manageable, blue print for making a technical, administrative and public value comparison of jurisdictions and sectors.

Analytical Coding and Synthesis

The 12 seminal frameworks obtained were entered into NVivo software which is qualitative data analysis software and analyzed rigorously and the themes extracted. Drawing on W. Richard Scott's (2014) institutional theory, we used this as our main coding framework which divided the operational needs of each framework into three institutional pillars:

The coding process was based on Scott's institutional theory and conceptualized governance as an interplay between the regulative, normative and cultural-cognitive dimensions. Regulative codes were found that described the technical aspects of the domain, such as explainability, logging APIs, retrieval augmented generation, computational boundaries, and traceability. Normative codes looked at the practices in institutional processes, including risk assessment, audit processes, procurement procedures, contractual liability, compliance mechanisms, etc. Cultural-cognitive codes that were centered around the ethical domain consisted of citizen dignity, equity, democratic trust, social justice, and meaningful human discretion.

We have applied for W. Richard Scott's pillars in NVivo and set up a very specific qualitative coding matrix to implement them and ensure intercoder consistency. In this matrix we linked some operational keywords and technical concepts directly to our target node structures. Technical keywords, such as explainability, API logging, retrieval augmented generation, and computational boundaries were used to fill the regulative nodes. Administrative compliance terms were used to populate the normative nodes including risk assessment, independent audit, procurement standards and contractual liability. To populate cultural cognitive nodes, human centered normative principles were followed, which mainly concerned the concepts of 'citizen dignity', 'administrative discretion', 'equity', 'social justice' and 'democratic trust'. We found this systematic classification of the text segments over these predefined nodes to be very transparent and reproducible, which served as our basis for the next step in model synthesis.

Conceptual Model Reconstruction

The final phase was to combine those codes that had been gathered to develop a public sector governance model that is not reserved. These three aspects of institutions were not isolated but were translated to the classic concepts of public administration, where the primary focus was on Mark Bovens' digital bureaucracy and Ansell and Gash’s (2008) collaborative governance. The Three-Dimensional Governance Model and the Layered Responsibility Framework and the Stakeholder Collaborative Loops in the following results were directly developed through this theoretical triangulation.

RESULTS AND DISCUSSION

In this section, we bring together the codes identified during the methodological review and introduce our

governance architecture of three dimensions and discuss how the elements of the governance architecture can directly contribute to tackling these ethical dilemmas in the integration of GenAI in the public sector.

The Three-Dimensional Governance Model

To control the deployment of generative architecture, we suggest a 'Three-Dimensional Governance Model'. This model fuses together technical requirements, institutional and ethical aspects to make sure that administrative automation remains accountable to democracy. The three domains are not linked in isolation but rather are intertwined: technical explainability facilitates institutional compliance, which can be used to support local ethical values and democratic trust. The three domains can be interpreted as the three pillars of institutional theory (Scott, 2014); regulative, normative and cultural which are then the pillars of institutionalization that are manifested in administrative procedures. The vertical shift from high level institutional pillars to the operational levels of developers and individual civil servants is what needs to be understood in this nested administrative structure as shown in Figure 3.

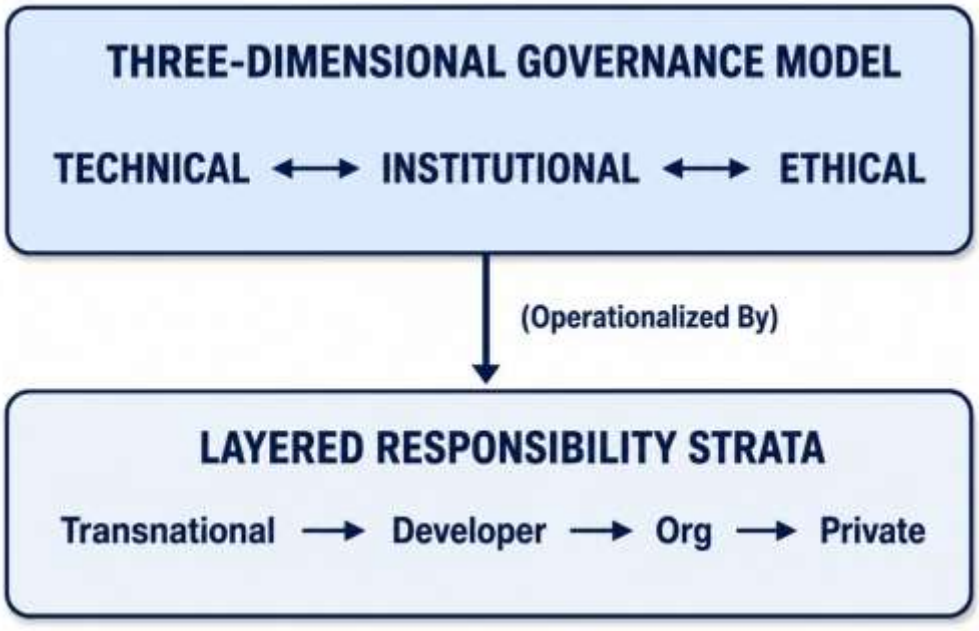

Figure 3. The Nested Governance Architecture

Technically, the model emphasizes aspects like locally explainable systems, retrieval-augmented generation, control of hallucination, audit trail, and culturally appropriate design (Kumar & Kumar, 2026). They are also important tools that enable public administrators to comprehend and interrogate the outputs of the system as opposed to relying on 'black box' vendor systems. A transparent technical layer also makes it easier to have the right to interrupt, make evidence available for future audits and make it possible to build ethical feedback loops that are based on system behavior, not trust.

The focus in institutional will shift from technical features to rules and arrangements which make the features enforceable. Public institutions should establish processes of approval, audits, specifications on procurement, liability mechanisms and secure local data platforms, based on the risk. In a sensitive service where citizens' access to essential data is crucial, such as the health care sector, these arrangements are of particular importance, they could leave citizens vulnerable to re-identification, commercial profiling, and lack of public control over essential data.

In the ethical dimension, governance is not only responsible for ensuring consistency of GenAI's technical functions, but also its consistency with public values – which is the area where achieving this is the most complex and challenging. This involves ensuring citizens' dignity, equity, social justice and democratic trust, and safeguarding against automated systems that discriminate against vulnerable groups (Tu & Tang, 2025). So, ethical governance is not an after-thought when deployed. It is the one which determines whether technical and

institutional decisions are legitimate.

Layered Responsibility Frameworks

The embedded public governance theory underpinned, where technological systems are linked to their social and institutional contexts (Bovens & Zouridis, 2002), an effective regulatory approach needs a multi-layered responsibility system. This is because it does not rely on a single point of accountability, but rather it distributes accountability throughout the generative AI lifecycle to audit, manage and reduce long-term ethical debt. Public institutions can avoid the social and technical burden by distributing the responsibility at various administrative levels. These layers work together and are interdependent, with platform-level compliance impacting on the capacity of civil servants to use meaningful professional discretion.

At the national and transnational level, legislation should set public-sector standards and establish clear procurement limits and prohibit high-risk generative practices that have serious impacts on fundamental rights, as noted by Carrasco (2025). This layer establishes a baseline level of responsibility for the use of generative AI in public sector work and prohibits the use of systems with high risks without adequate regulation.

There should be a legal and ethical duty on the AI developer to bear responsibility at the platform and developer level for the pretraining datasets, system design, documentation, and post-deployment monitoring. This encompasses the regular auditing of algorithms, technical security measures like auditable API designs, human-in-the-loop review points, and useful tools to facilitate the right to interrupt automated processing processes (Liu, 2025).

Public agencies should, at the organizational level, have internal review boards, define acceptable administrative use cases, establish continuous processes for ethical assessment and establish clear procedures for officials to interrupt or challenge automated outputs. They should be backed up by formal administrative safe harbours to prevent civil servants from being victim to career retribution for raising concerns or reporting errors in good faith.

Individual civil servants need to be educated on maintaining active oversight, developing professional skepticism and understanding the dangers of algorithmic bias. Individual accountability should not be applied to avoid placing responsibility on system failures in the design of the vendor and/or in the specifications or institutional policy. While there is a place for meaningful professional discretion for civil servants, they should not be blamed for failures stemming from higher levels of the governance system.

These four layers of responsibility correspond directly to the Three-Dimensional Governance Model, with the aim of conceptual integration. The layers of Individual Civil Servant and Platform/Developer and Ethical and Institutional share the operational processes necessary to implement safeguards within the Technical Domain, while the layers of National/Transnational and Ethical/Institutional establish the normative boundaries of the Ethical and Institutional layers. This vertical mapping puts technical standards into practice through developer API logging and liability, helps ensure institutional compliance with organizational audit checklists, and safeguards ethical values by giving civil servants the technical capacity and institutional authority to exercise meaningful discretion.

This multi-layered approach is also beneficial for practical review and assessment of accountability. In case biased or harmful outputs present, auditors can investigate if the problem lies in the training data, retrieval source, procurement specification, organizational review process or final human assessment. So it is not a matter of whether it's the machine's fault or the officer's fault. Rather, the key issue is how the governance system went wrong, and what remedial action is necessary to avoid further failures in the context of public service in the future.

The Hair Model and Universal Standards

Fragmentation of ethics can be prevented when local arrangements are included in universal services like education. To demonstrate the layered approaches, the “Human Artificial Intelligence Rights (HAIR)” model

(Abuzir, 2026) will be used in this study. The HAIR model respects the concept of the student's cognitive agency as well as the idea of developmental privacy as rights of the student, not to be delegated to automated systems, and offers a way to oppose algorithmic deskilling and leave the student free to learn independently, through local, non-commercial retrieval architectures.

At the same time, universal ethical principles make it harder for public institutions to be captured by corporations by allowing administrators to audit for bias and create an inclusive dataset in a consistent manner. Taking together, therefore, all these are protective measures, and once put together, they create a (protective) framework which does not let students' data be commercialized and reminds us that universal 'guard rails' can be constructed on top of local priorities for education.

The Role of Stakeholders in Ethical Implementation

The application of GenAI in public services must be more than an administrative issue and must be done with a careful and ethical approach to business. Instead, it relies on a multi-stakeholder governance process, which offers checks and balances across the Three-Dimensional Governance Model, promotes involvement of other stakeholders and creates a process of overlapping and overlapping responsibilities. Instead, it is a multi-stakeholder and collaborative governance system that brings checks and balances with it across the Three-Dimensional Governance Model, encourages participation of different actors, and creates a system of overlapping and overlapping responsibilities. To depict these co-operative paths, the intertwined processes of the public agencies, the technology developers and the democratic public continually interacting are displayed in Fig. 4, and the administration, technology transparency, and civic contestation are continually calibrated.

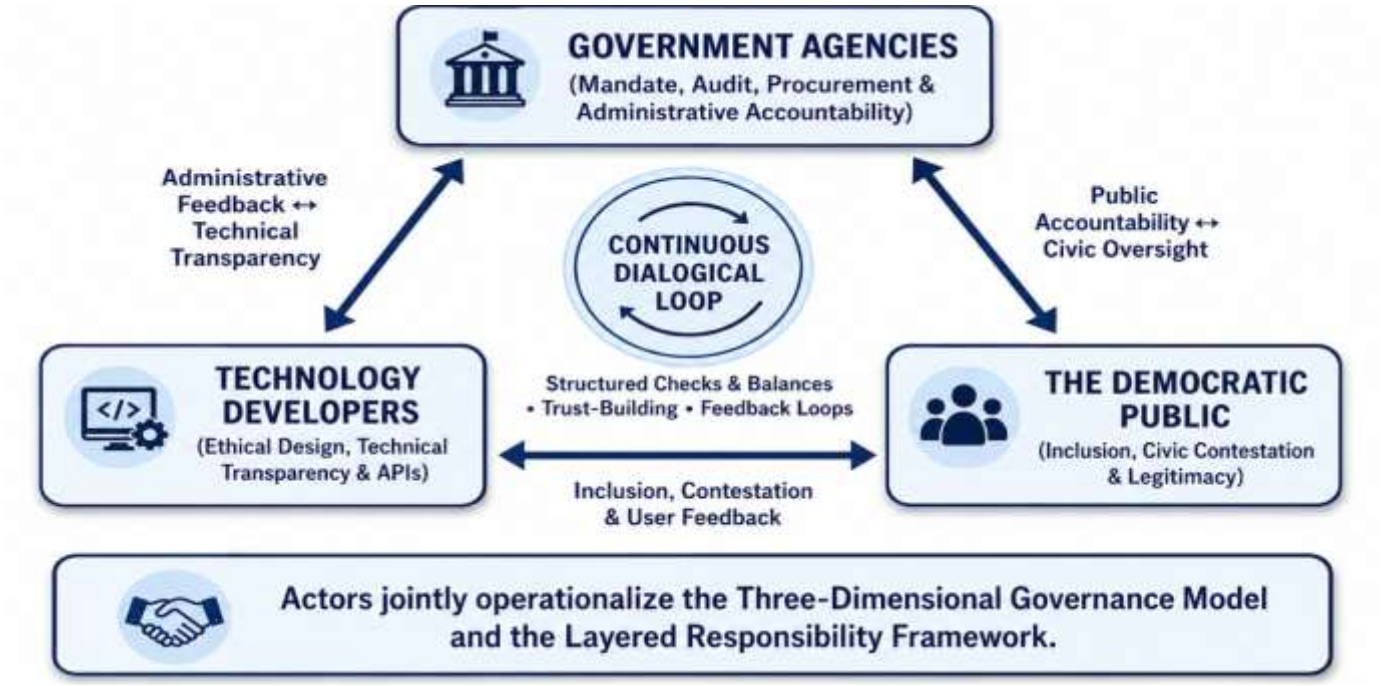

Figure 4. The Collaborative Stakeholder Deliberation Loop

The role should be proactive and not reactive, with the government agencies being the ones to carry out institutions and ethics. They can use procurement leverage to ensure open architecture, independent audit logs, security of data handling, and enforceable procurement requirements of vendors (Goyal et al., 2026). That way, procurement may be utilized as a governance instrument to prevent public agencies from unwittingly unethical projects that are driven by poorly managed automation. Corporate and academic technology developers should move away from rapid deployment without adequate safeguards and avoid relying on corrections only after problems occur.

Instead, an ethical design from the onset of the development, including explainability, protection of privacy and human override systems, and engineering teams with varied engineering backgrounds who can detect dataset

bias prior to deployment is more appropriate for public sector uses, as described by Kumar and Kumar (2026). Developers must also give administrators and citizens adequate transparency to be able to have meaningful oversight.

The last source of administrative legitimacy is the democratic public. The use of generative systems can be influenced by local values, as reflected in public consultation, citizen advisory panels, clear appeals and explanations. The civic sensors are also able to expose recurring errors, challenged results, and decision paths that fall into the automated trap if they do not participate in such activity.

Domain-Specific Strategic Insights

For public administrators, it is necessary to create a highly context-specific and sector-specific operational guidelines that directly translate the Three-Dimensional Governance Model and Layered Responsibility Framework into sector specific protocols to bring the conceptual frameworks to the operational and implementation in the field. As in public administration, there are two hazards of social manipulation:

External vulnerabilities are vulnerabilities found when malicious text, images, audio or deepfakes can disrupt municipal election workflow, crisis communication, public health campaigns or other activities which rely on trust. Such risks are not insignificant because they have the potential to undermine the citizens' trust and confidence in official information during critical moments.

Internal vulnerabilities are also important because public agencies may be just as misinformed as citizens when the wrong information is fed into the system through unchecked generative paraphrasing, from regulatory documents, policy briefings, procurement notes, or just citizen-facing explanations. This damage is not the same as if citizens are intended to be manipulated but could be the same if they can't distinguish between verified information provided by the administration and synthetic or poorly verified information.

When the gap between real and artificial reality, created by human beings, and synthetic reality, created by machines, is fully erased, the public sphere is highly vulnerable to systematic manipulation and leads to loss of confidence in the institutions that guarantee democratic societies.

Institutional Barriers and Layered Responsibility in Public Administration

The Technology Organization Environment framework (TOTE) (Goyal et al., 2026) can help public managers evaluate and tackle institutional issues related to safe integration of GenAI. This model directly relates to the Three-Dimensional Governance Model, as technical, institutional and ethical is directly connected to our model's technology, institutions and ethics, respectively, ensuring full theoretical compatibility. This systematic evaluation implemented the National, Platform and Organizational layers of responsibility, setting up regulatory boundaries prior to system deployment. An agency should not completely and automatically implement road maps but should have a timeline for their deployment. According to Scott's second cultural cognitive pillar, early tasks should focus on low-risk internal processes, such as formatting draft documents, summarizing documents, etc., and the allocation of public resources using high discretion should be left, as an administrative process, to human civil servants to preserve the administrative process as a cultural cognitive phenomenon.

Patient Trust, Technical Safeguards and Data Sovereignty in Healthcare

Technical and institutional measures need to be taken to ensure patient confidence in public clinical settings. Carries out the Technical and Institutional domains directly. Local federated learning is needed to avoid the risk of patient reidentification, which could occur in public health institutions that do not have an external cloud-based API to access patient data (Deng et al., 2025; Mwogosi, 2025). This single Technical Domain deployment addresses the critical patient data sovereignty issues and clinical re-identification risks that were already seen in new clinical settings, like python databases in Tanzania's public health system. Also, all generative clinical summaries should be double blind and humanly validated by licensed medical professionals as an ethical barrier before entering the EHR system (Individual Civil Servant layer).

Safeguarding Cognitive Agency and Pedagogical Governance in Education

The key is to keep critical thinking and cognitive socialization alive in public schools, an issue that should be addressed by education policymakers. The pedagogical (micro) ethical safeguards have to correlate directly with the Platform and Organizational responsibility layers to make the HAIR model's macrolevel ethics principles operationally meaningful in the classroom. In practice, some of the issues that need to be addressed include formal teacher training on generative methodologies, establishment of academic integrity guidelines that incorporate the use of AI as a cognitive collaborator, rather than an author, and implementing strict data boundaries with the incorporation of localized and non-commercial retrieval architectures to protect student learning profiles from commercial exploitation (Ofosu Asare, 2026). One of the most active countermeasures public schools can take against the dangers of automated cognitive deskilling and loss of student developmental privacy is to create a protective sociological shield; one in which commercial extraction of data is limited. One of the ways public schools can offer a protective sociological shield to limit commercial extraction of data is to be active.

DISCUSSION

Bridging Technical Excellence and Social Responsiveness

The main challenges in public sector generative governance are how to balance technical excellence (improving model performance, processing speed and administrative efficiency) with social responsiveness (ensuring protection of public values, human dignity and democratic accountability). Public institutions that do not consider the ethical dimension, but only the technical, find themselves in 'ethical debt'. Ethical debt occurs when the ethical component of the first procurement and implementation of automated systems is deferred and becomes long-term (compounding) social and institutional obligations. No one else is going to solve this technology vendor debt issue other than self-regulation. Instead, we suggest three important, binding administrative steps be put in place: These validation checks are a key feature of this real-time override path, and they are also illustrated in a real-time automated workflow in Figure 5, where the routing of decision making out of the primary processing engine, into the human administrators, is shown.

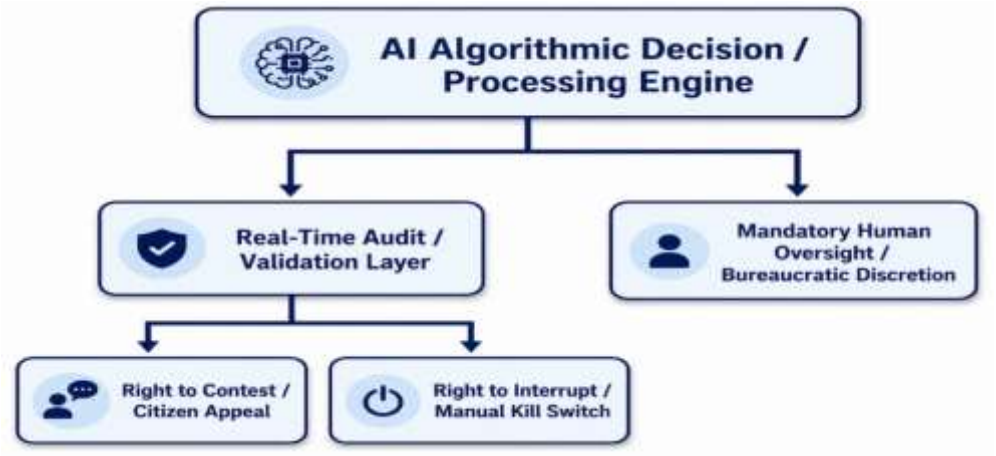

Figure 5. Real-Time Administrative Override Safeguard Architecture

The Right to Interrupt

The Right to Interrupt is a fundamental institutional protection which will restore epistemic authority and control to street-level administrators. The Right to Interrupt comes from the legally protected and technically enabled power of a civil servant to halt, override or terminate an automated generative process when the result of the process is inconsistent with the set of administrative, legal or ethical requirements.

It should be integrated into the system's technical API architecture and be an active manual override and NOT a post-hoc process. In high-stakes public environments, like the public clinical settings that have been analyzed in

section of patient trust, technical safeguards and data sovereignty in healthcare, this protection is implemented in practice when a generative summary system generates clinical hallucinations or alerts for data privacy reidentification in public EHR databases. It can be done by public organizations by offering this system level kill switch, which is available in real-time. It is a policy instrument that is essential to regaining the discretionary flexibility of the street level bureaucrats and a fundamental balancing counter to the centralized automated aspects of system level bureaucracy.

The Right to Contest

The Right to Interrupt is the right to defend the system from an inside out (Administrator) point of view, while the Right to Contest is the right to defend the system from an outside in (Citizen) point of view. It is a right to contest, namely a clear, legally binding right to get a full, readable explanation of the automated public action from a human being, and to contest and re-evaluate the action with a human being as the administrator.

Public systems provide a very robust administrative safe harbor to challenge; therefore, no resource allocation shall ever be automatic. Pedagogical governance section provides that the assessment of the student's learning or the evaluation of the student's cognition carried out with an automated tool is formally challenged by the teacher and the student's parent. This bottom-up challenge is an important democratic sensor which can sense and expose errors in the system and bring it back into the review cycle of the organization as empirical feedback.

Mandatory Human Oversight

Human resources need to be put in place to support human oversight. It requires time and training, an escalation process, and authorization from the organization to take a different view than the automaton's recommendation. If not, the oversight can turn into a ceremonial. One of the members of public service can decide and formally approve a decision but the member cannot check into it in detail. That is why it is important to conceive of mandatory oversight within an institutional framework, rather than just ‘off' on a software box.

Overarching protection is the process of learning; this is continuous and active intervention by humans and learning by humans. In that way, the role of the generative system is confined to supporting a decision of the human licensed administrator and no final and binding administrative action can be taken without direct and active approval from the human licensed administrator.

Human in the loop is not only a programmatic requirement, but also an ethical commitment to the final barrier of human judgement for democratic ideals. This is an implementation of individual accountability layer, and it allows flexible and adaptive administrative acts in the context of evolving public values by means of automated systems.

CONCLUSION AND POLICY RECOMMENDATIONS

The ethical challenges of Generative AI in public services are not easy to navigate, and a significant paradigm shift is needed to be successful. As opposed to technologically/performance-oriented interventions, the public sector organizations should have a more comprehensive and human-centered public administration, which acknowledges the discretion of street level public administration (Bovens & Zouridis, 2002). To support human rights and institutional legitimacy, robust and multidimensional governance systems, specifically in line with democratic transparency, institutional accountability and social inclusion can be used to harness the transformative potential of GenAI. This change is crucial to bring the functional and functional cultural aspects of the norm and culture of governance to life – not as an instrument of automated centralization, but as an adaptive tool of stewardship.

Concrete Policy Recommendations for Public Administrators

To bring this paradigm shift into practice, public agencies and the transnational organizations should shift from general ethical statements to policy instruments that can be implemented. These are three obvious and necessary policy demands that we think are soon: First, public organizations need to include in their procurement contracts

that the vendor will create secure, localized and auditable architectures. These systems need to be designed to accommodate real-time manual override, tamper evidence logs, transparent data protection duties, and bias audit before use.

Second, Governments need to create algorithmic registers at national level for high-risk generative systems used in public services. These registries should be more informative, and should contain the purpose of the system, who is responsible for the system, a description of the documentation of the model, the data sources, the structure for accessing the data, an audit schedule and the results of the audit system performance. These registries should be more informative, such as a statement of the purpose of a system, the agency responsible for a system, documentation of the model, sources of data, structure for accessing the data, an audit schedule, and results of the audit system performance, thereby giving public accountability and judicial review a firmer foundation. Thirdly, there must be some safeguards for the civil servants who, in good faith according to their rules of practice, intervene in and thwart automated decisions. If left unchecked, the right to interrupt could be a futile gesture as frontline staff could worry about facing repercussions in their career or within their institution if they question an approved system.

Study Limitations and Future Directions

This study contains a governance framework integrated across various dimensions, but the major drawback of it is the conceptualization and synthesis-based approach. The model has yet to be tested in empirical studies of active public agencies because it is based on qualitative coding of policy frameworks. In this regard, further research is needed at the field level to explore the practical, cognitive and organizational issues that arise when implementing the 'Right to Interrupt' and 'Right to Contest' in the field. There is a need for future research to employ ethnographic and longitudinal methods to gain a deeper understanding of the wider social impact of algorithmic decision support systems, especially how civil servants handle algorithmic bias. It is also important to create transnational governance institutions to guarantee the sovereignty of data, with the aim of harmonizing international standards for the responsible use of AI, as described in the transnational responsibility layer of this model. The benefits of GenAI in the public sector will go beyond technology and its efficacy to the democratic principles and values that underlie its design and governance.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to the Azman Hashim International Business School (AHIBS) and Universiti Teknologi Malaysia (UTM) for providing the institutional support and research facilities to conduct this research. The authors would also like to thank the anonymous reviewers for their qualitative and constructive feedback during conceptual design process, which was helpful during initial conceptual stage of this multi-dimensional framework and helped in clarifying the institutional and institutionalization theories and institutional and institutionalization accountability theories.

Ethical Considerations

Ethical Approval

This study is based on an integrative conceptual synthesis of published literature and policy frameworks. It does not involve human participants, animals, or personally identifiable data. Therefore, formal ethical approval was not required.

Conflict of Interest

The authors declare that there is no conflict of interest.

Data Availability Statement

The data supporting this study are derived from publicly available literature, policy documents, and governance frameworks cited in the reference list.

REFERENCES

1. Abuzir, Y. (2026). Guiding pedagogical safety in the digital age: An implementation of the Human Artificial Intelligence Rights (HAIR) model in public education. International Journal of Educational Technology in Higher Education, 23(1), 89–104.
2. Ansell, C., & Gash, A. (2008). Collaborative governance in theory and practice. Journal of Public Administration Research and Theory, 18(4), 543–571.
3. Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. European Law Journal, 13(4), 447–468.
4. Bovens, M., & Zouridis, S. (2002). From street-level to system-level bureaucracy: How information and communication technology copes with discretion. Public Administration Review, 62(2), 174–184.
5. Bozeman, B. (2007). Public values and public interest: Counterbalancing economic individualism. Georgetown University Press.
6. Carrasco, L. B. (2025). Towards responsible AI: A conceptual framework for mitigating ethical risks in generative artificial intelligence systems. AMCIS 2025.
7. Deng, M., Yang, R., Zheng, X., Deng, Y., & Jiang, J. (2025). Artificial intelligence in diabetes care: From predictive analytics to generative AI and implementation challenges. Frontiers in Endocrinology, 16, Article 1620132. https://doi.org/10.3389/fendo.2025.1620132
8. Goyal, R., Deshmukh, S. G., & Bolia, N. (2026). Navigating barriers to GenAI adoption in public administration: A systematic evaluation and policy roadmap. Socio-Economic Planning Sciences, 105, Article 102461. https://doi.org/10.1016/j.seps.2026.102461
9. Hood, C. (1991). A public management for all seasons? Public Administration, 69(1), 3–19.
10. Kumar, P., & Kumar, S. (2026). Designing trustworthy artificial intelligence: Ethical frameworks, transparency and accountability in AI systems. Journal of Trust Research, 14(1), 45–63.
11. Liu, J. Y. (2025). Generative artificial intelligence: Ethical challenges and trust mechanisms. IT Professional, 27(2), 34–41.
12. Mashaw, J. L. (1983). Bureaucratic justice: Managing social security disability claims. Yale University Press.
13. Mwogosi, A. (2025). Ethical and privacy challenges of integrating generative AI into EHR systems in Tanzania. Digital Health, 11, 23–38.
14. Ofosu Asare, Y. (2026). Guiding the future: Developing an ethical framework for generative AI use in education. International Journal of Information and Learning Technology, 43(1), 56–72.
15. Scott, W. R. (2014). Institutions and organizations: Ideas, interests and identities (4th ed.). SAGE Publications.
16. Tu, J., & Tang, M. (2025). The ethical reconstruction of generative AI: Social value, governance challenges and pathways to digital empowerment. Lecture Notes in Computer Science, 14500, 210–226.
17. Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. Human Resource Development Review, 4(3), 356–367.
18. Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. Journal of Advanced Nursing, 52(5), 546–553.