LEX LOCALIS-JOURNAL OF LOCAL SELF-GOVERNMENT
ISSN:1581-5374 E-ISSN:1855-363X
VOL. 24, NO. S1 (2026)

AI-ENABLED PUBLIC GOVERNANCE IN DEVELOPING STATES: SERVICE DELIVERY GAINS, ACCOUNTABILITY RISKS, AND A PRACTICAL RISK-BASED REGULATORY MODEL

NAVEED RAFAQAT AHMAD[1]

[1]Director General, Punjab Sahulat Bazaars Authority (PSBA), Government of the Punjab, Pakistan

nrahmad@live.com[1]

Abstract
Governments are moving quickly from small artificial intelligence (AI) pilots to operational use in public administration especially in citizen services, compliance, fraud detection, and planning. This shift is no longer theoretical: the United States' consolidated federal inventory reported more than 1,700 AI use cases across agencies, including a significant subset classified as rights- or safety-impacting. At the same time, evidence from advanced administrations shows that well-designed "assistive" systems can produce measurable gains, such as sharply reduced response times in service workflows and time savings for public servants.
However, without clear governance, AI can weaken accountability through opaque decision pathways, biased outcomes linked to poor or unrepresentative data, staff over-reliance ("automation bias"), and weak or inaccessible channels for citizens to challenge outcomes. These risks are more acute in developing states where institutional capacity, procurement maturity, data governance, and independent oversight are often uneven. The central policy problem is therefore not whether governments should use AI, but how they can adopt it while preserving procedural fairness, explainability, and public trust.
This paper proposes a practical, risk-based governance model for the public sector that translates international principles into operational controls that resource-constrained administrations can implement. The model classifies government AI into low-, medium-, and high-risk uses, and aligns safeguards to impact. For high-risk systems such as eligibility and sanctions decisions, law-enforcement support, and biometric identification the paper specifies minimum deployment requirements: named accountability ownership, meaningful human oversight, pre-deployment impact assessment, proportionate explainability, data quality and fairness testing, security controls with audit trails, enforceable procurement clauses for vendor accountability, and accessible grievance and review mechanisms.
The paper also provides a phased implementation roadmap: (i) governance rules, procurement templates, and an AI registry; (ii) testing, monitoring, and audits; and (iii) stronger independent oversight, transparency, and redress. The paper's contribution is a governance framework that enables service delivery gains while preventing the most damaging failure mode in public administration: high-impact automation that becomes effectively unchallengeable.

Keywords: AI governance, digital government, accountability, risk-based regulation, public administration, developing states, trustworthy AI, algorithmic impact assessment, public sector innovation

1. Introduction

1.1 Background and problem statement

Artificial intelligence (AI) adoption in government has moved decisively beyond experimentation. What began as small pilot projects is now embedded in routine public administration, including citizen service delivery, back-office automation, compliance and fraud detection, and policy analytics. Recent consolidated reporting from the United States illustrates the scale of this shift: more than 1,700 AI use cases were reported across federal agencies as of December 2024, with a substantial subset classified as rights-impacting and/or safety-impacting. This evidence confirms that public-sector AI is no longer peripheral; it is becoming a normal component of high-stakes administrative decision-making.
International experience also shows that AI can deliver tangible administrative benefits when applied appropriately. Governments have reported faster case processing, reduced backlogs, improved targeting of audits and inspections, and measurable productivity gains for public

servants. These outcomes explain why AI adoption is accelerating across jurisdictions, including in countries with limited fiscal and human resources.

At the same time, the expansion of AI into core government functions has exposed a central governance challenge. When deployed without clear rules and accountability, AI can weaken public administration rather than strengthen it. Opaque decision pathways, biased outcomes driven by poor or unrepresentative data, over-reliance by frontline staff on automated outputs, and weak mechanisms for citizen challenge and correction have emerged as recurring risks. In public systems where decisions affect rights, benefits, safety, and livelihoods such failures carry legal, social, and political consequences.

1.2 Why developing states face higher governance risk

While these challenges are visible across all administrations, they are amplified in developing states. Many developing governments operate with fragmented data systems, limited capacity for model testing and assurance, asymmetric procurement relationships with technology vendors, skills shortages within the civil service, and weaker or over-stretched independent oversight institutions. In this context, AI systems may be adopted faster than the governance structures needed to control them.

These conditions increase the likelihood that AI's operational benefits speed, efficiency, and targeting are achieved at the expense of procedural fairness, transparency, and legitimacy. The risk is not simply technical error, but administrative failure: decisions that cannot be adequately explained, reviewed, or corrected. For governments already facing trust deficits, such outcomes can undermine confidence in public institutions and provoke resistance to further digital reform.

1.3 Research question

Q: How can developing-state governments use artificial intelligence to improve service delivery while protecting accountability, fairness, and citizen trust through a risk-based governance model?

1.4 Contribution and novelty

This paper makes three interrelated contributions to the literature on AI and public governance.

First, it proposes a public-sector risk classification framework that distinguishes low-, medium-, and high-risk AI uses based on administrative impact rather than on technology type or novelty. This shifts attention from abstract debates about AI to the concrete consequences of its use in government decision-making.

Second, it specifies a minimum, implementable safeguards package for high-risk government AI systems. These safeguards covering accountability ownership, human oversight, explainability, data quality and fairness, security and auditability, procurement obligations, and citizen redress are designed to be operational rather than aspirational.

Third, the paper presents a phased implementation roadmap that reflects the capacity constraints and sequencing realities faced by developing administrations. Rather than assuming advanced regulatory or technical infrastructure from the outset, the roadmap emphasizes incremental governance building that can be achieved within existing administrative systems.

2. Conceptual method and approach

This paper is a conceptual framework study grounded in applied public administration rather than in technical system design. Its approach combines two elements.

First, it synthesizes key themes from international governance standards and policy frameworks relevant to public-sector AI, including principles of trustworthy AI, administrative accountability, and structured risk management. These sources converge on

the idea that AI should be governed according to the level of risk it poses to individuals and society, with stronger controls applied where public impact is greatest.

Second, the paper translates these principles into a practical administrative model suitable for ministries and agencies. The focus is on concrete governance instruments risk classification, impact assessment, procurement clauses, testing and audit routines, monitoring indicators, and citizen recourse mechanisms rather than on abstract ethical statements or purely technical controls.

The design logic follows the risk-based governance direction increasingly emphasized in international practice, including guidance from multilateral organizations and public-sector risk management frameworks. By combining normative standards with operational tools, the paper aims to bridge the gap between global principles and day-to-day public administration, particularly in developing-state contexts where governance capacity must be built incrementally rather than assumed.

3. Defining "AI in Governance": Use-Cases and Risk Concentration

3.1 Common public-sector use-cases

Public-sector AI is best understood as a set of administrative applications (not a single technology) that support distinct government functions. International evidence shows that adoption is already widespread and increasingly operational. In the United States' consolidated federal reporting (as of 16 December 2024), agencies reported 1,700+ AI use cases, of which 227 were identified as rights-impacting and/or safety-impacting, indicating that AI is being used not only for internal efficiency but also in contexts with direct public consequences.

Across jurisdictions, government AI use clusters into three practical domains that align with public administration workflows:

A. Service delivery and citizen interaction (front-office systems)

These are tools that support citizen queries, case updates, and administrative communications often through natural language interfaces and "assistive drafting" features. OECD case documentation shows measurable service improvements when AI is used in controlled, human-supervised workflows: one recorded case reported average response time reduced from 19 days to 3 days, with 78% of public servants finding AI suggestions useful, 70% of suggested elements validated by officials, and improved citizen satisfaction (68% for AI-assisted responses vs 57% for traditional responses).

Implication: Front-office AI can produce real service gains, but it must remain assistive (draft/support), not authoritative.

B. Back-office efficiency, compliance, and fraud detection (operational systems)

These systems include document processing, routing/triage, anomaly detection, and audit prioritization tasks where governments face high volume and repetitive steps. The World Bank identifies recurring patterns of public-sector AI use cases including citizen engagement, compliance and risk management, fraud and anti-corruption, business process automation, service delivery, and analytics for policy design. A clear, documented example of administrative automation impact is the UK Department for Work and Pensions (DWP) use of automation in pensions processing: 12 software robots processed approximately 2,500 claims per week, clearing a 30,000-claim backlog in two weeks.

Implication: Back-office AI/automation often produces the fastest early wins, but requires controls to prevent "silent" processing errors.

C. Policy intelligence and planning (strategic systems)

Governments also apply AI to forecasting, early warning, targeting, and performance monitoring. OECD's foundational analysis of public-sector AI highlights that governments use AI to design better policies, improve decision-making, and improve the speed and quality

of public services, while noting that benefits depend on governance capacity and institutional context.

Implication: As AI shifts from automation into planning and targeting, governance needs increase because outputs influence priorities and resource allocation, not just internal speed.

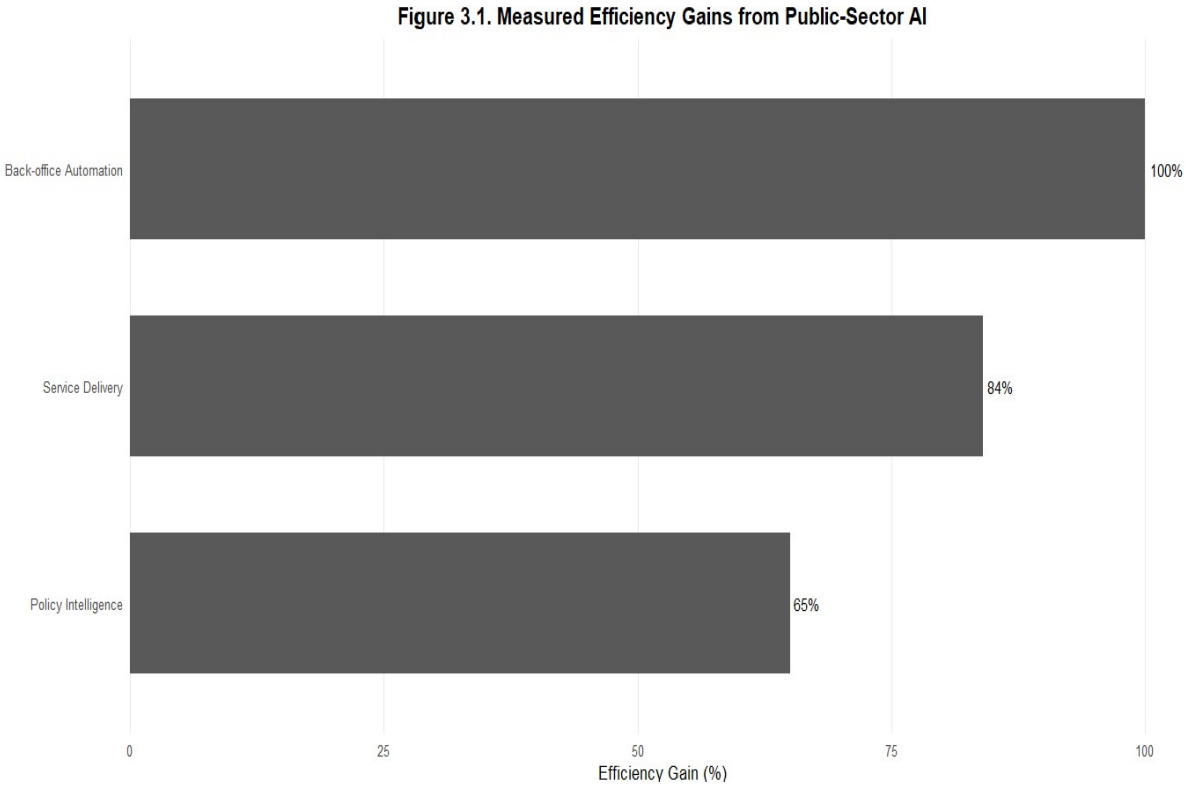


Figure 3.1 presents indicative efficiency gains associated with selected public-sector AI applications, demonstrating their capacity to enhance processing speed, administrative responsiveness, and service delivery outcomes. The comparison highlights the growing operational value of AI across core governance domains.

3.2 Where risk increases: high-impact domains

A consistent finding across international practice is that risk is driven by administrative impact, not by whether a tool is labeled "AI." Risk becomes acute when AI influences decisions that affect rights, safety, access to benefits, or reputation, and when errors are difficult to detect or challenge.

High-impact domains include:

1. Eligibility, benefits, sanctions, and adverse administrative outcomes

When AI influences eligibility, termination, sanctions, or enforcement decisions, the administrative stakes are highest. US federal reporting explicitly distinguishes rights- and safety-impacting AI uses, reinforcing the principle that some government AI systems require elevated governance and risk management practices.

Risk mechanism: wrong denials/approvals, unequal treatment, and decisions that become effectively unreviewable in practice.

2. Compliance, fraud, and anti-corruption systems that trigger adverse action

These applications can generate high value but also high harm if outputs are treated as determinations rather than leads. The World Bank's GovTech evidence documents a Brazilian federal use case where AI identified more than 500 firms owned by public servants working at the same agency executing the contracts, involving over R$ 4.5 billion in contracts.

Risk mechanism: false positives can damage reputations and livelihoods unless outputs are governed as *risk signals* requiring human investigation and due process.

3. Public procurement analytics and red-flagging

Procurement is a governance-critical area because it directly relates to public funds and integrity. OECD OPSI documents "Robot Alice," a procurement analytics system that in 2023 analysed 190,923 acquisitions, and generated 203 audit jobs covering R$ 27 billion in procurement value.

Risk mechanism: biased targeting (always flagging the same regions/vendors), opacity in criteria, and institutional over-reliance without periodic validation.

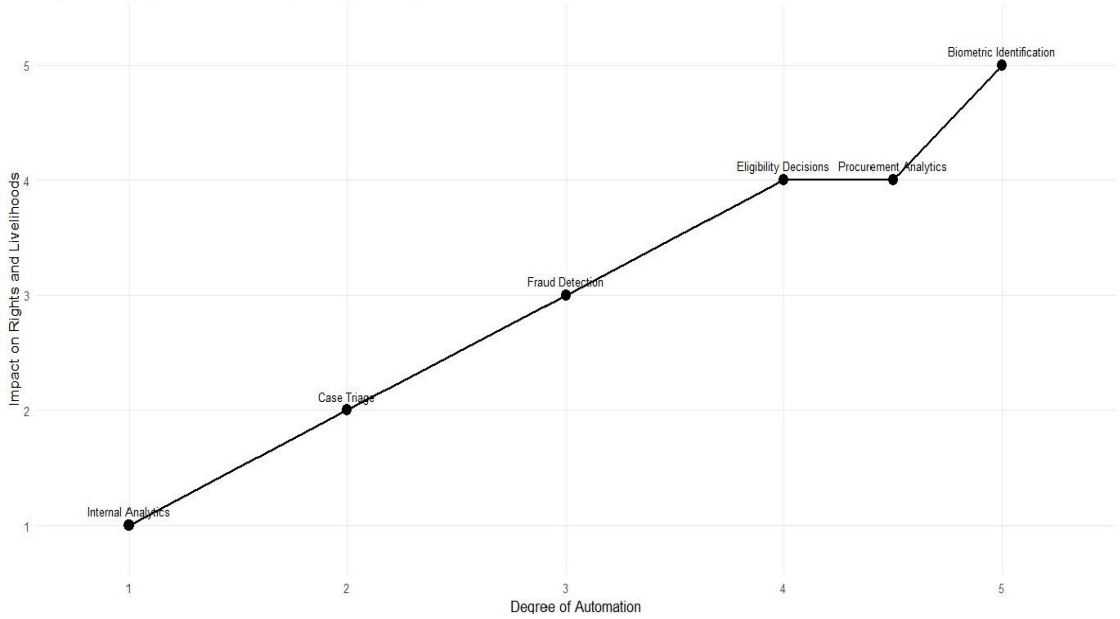
4. Biometrics and identity matching

Biometric identification is widely recognized as high-impact because of privacy and fundamental rights implications. The EU AI Act establishes a structured risk approach and identifies sensitive areas such as biometrics and law enforcement as requiring strict controls under harmonised rules.

Risk mechanism: misidentification, disparate error rates across groups, and disproportionate consequences in public settings.

Figure 3.2 illustrates the escalation of governance risk as the degree of automation increases within administrative decision-making. It emphasizes that applications exercising greater decision authority require stronger institutional oversight and accountability safeguards.

Figure 3.2. Risk Concentration in Public-Sector AI



3.3 Practical take-away for the framework (why this classification matters)

International practice supports a simple administrative conclusion: AI governance must be impact-based. Tools used for internal analytics and workflow support can often be governed

with baseline controls, while systems influencing rights, benefits, enforcement, procurement integrity, or biometrics require stronger safeguards, independent checks, and meaningful citizen recourse.

To operationalise this, the next section of the paper should formalise:

- a use-case-to-risk mapping (what governments deploy, and where risk concentrates), and
- a minimum safeguards threshold for high-impact deployments (what must be true before go-live).

Table 3.1 – Public-Sector AI Use-Cases: Documented Examples, Measured Outcomes, and Governance Risks

Use-Case Cluster	Government / Institution	AI Function	Documented Outcome	Administrative Benefit	Principal Governance Risk	Recommended Risk Tier
Citizen communication	OECD case documentation	AI-assisted drafting and response support	Response time reduced from 19 days to 3 days; higher citizen satisfaction	Faster service delivery; reduced backlog	Automation bias; insufficient review of AI-generated communication	Medium
Pension processing automation	UK Department for Work and Pensions	Robotic process automation for claims handling	30,000-claim backlog cleared; ~2,500 claims processed weekly	Rapid throughput; operational continuity	Silent processing errors at scale	Medium
Fraud and conflict-of-interest detection	Brazilian Federal Government	Pattern detection across procurement and employment data	500+ firms identified, involving over R$ 4.5 billion in contracts	Strengthened integrity oversight; improved audit targeting	False positives; reputational harm without due process	High
Procurement analytics	OECD OPSI ("Robot Alice")	Automated contract analysis and red-flag generation	190,923 acquisitions analysed; 203 audit jobs covering R$ 27 billion	Enhanced transparency; risk-based auditing	Opaque criteria; biased vendor targeting	High

Federal AI deployment inventory	United States Federal Agencies	Cross-agency AI applications	1,700+ use cases, including 227 rights- or safety-impacting	Strategic visibility; improved coordination	Governance fragmentation; uneven safeguards	High (varies by use)
Policy analytics and forecasting	OECD public-sector AI analysis	Predictive modelling for policy design	Improved decision speed and service quality (context-dependent)	Better resource allocation; anticipatory governance	Over-reliance on predictive outputs	Medium–High
Citizen engagement and service access	World Bank GovTech cases	AI-enabled engagement platform	Expanded access channels and improved responsiveness	Greater administrative reach	Digital exclusion; data privacy concerns	Medium
Identity verification / biometrics	EU regulatory classification	Automated identity matching	Recognized as high-risk under harmonised rules	Stronger security; fraud reduction	Misidentification; fundamental rights impacts	Very High

4. International Governance Standards for Public-Sector AI: Convergence, Not Fragmentation

4.1 Why international standards matter for public administration

As artificial intelligence becomes embedded in routine government functions, international institutions have moved rapidly to articulate governance standards for public-sector use. These standards are not abstract ethical debates; they respond to concrete administrative failures observed in early deployments, including unexplainable decisions, biased outcomes, unclear responsibility between agencies and vendors, and weak mechanisms for appeal and correction.

Although developed by different institutions and for different legal systems, the leading international frameworks converge on a common core: AI used by governments must remain accountable, transparent, and subject to human control especially when it affects rights, benefits, safety, or livelihoods. This convergence is particularly important for developing states, which often rely on international guidance to shape domestic policy and procurement practices.

This section reviews the most influential international governance instruments and extracts their shared operational logic, rather than treating them as competing or optional approaches.

4.2 OECD: Trustworthy AI as an administrative obligation

The OECD Principles on Artificial Intelligence, adopted in 2019 and reaffirmed through subsequent guidance, represent the most widely endorsed international standard for AI

governance, formally supported by over 45 countries. The principles are explicitly framed for public policy and public administration, not only for private-sector innovation.
The OECD identifies five core requirements for trustworthy AI:

1. Inclusive growth, sustainable development, and well-being
2. Human-centred values and fairness
3. Transparency and explainability
4. Robustness, security, and safety
5. Accountability

For public administration, these principles translate into concrete administrative duties. OECD case analysis shows that AI systems in government must be designed so that public officials remain responsible for outcomes, can understand and explain system behaviour at an appropriate level, and can intervene when results appear incorrect or unfair. Importantly, the OECD explicitly warns against over-reliance on automated outputs, noting that productivity gains depend on embedding AI into administrative workflows with clear oversight rather than treating it as a substitute for judgment.
Recent OECD work on Governing with Artificial Intelligence reinforces this position by documenting that governments achieving measurable gains are those that combine AI adoption with risk management, staff training, auditability, and accountability mechanisms. The OECD does not recommend blanket restrictions; instead, it consistently endorses risk-based governance, with stronger controls applied where public impact is greatest.

4.3 UNESCO: Human rights, legality, and public accountability

UNESCO's Recommendation on the Ethics of Artificial Intelligence (2022) provides the most comprehensive global articulation of AI governance grounded explicitly in human rights law. While often discussed in ethical terms, the Recommendation contains direct implications for public administration.
For government use of AI, UNESCO emphasizes:

- Legality and proportionality in administrative decision-making
- Human oversight and responsibility, particularly in public authority contexts
- Transparency and explainability, enabling affected individuals to understand decisions
- Access to remedies, including the right to challenge and seek correction

UNESCO's guidance is especially relevant for developing states because it links AI governance to existing international human rights obligations rather than to advanced technical capacity. It makes clear that public authorities cannot delegate responsibility to automated systems or private vendors. When AI is used in public functions, the state remains accountable under administrative and constitutional law. The Recommendation therefore reinforces the principle that AI governance in government is not optional innovation policy, but a matter of lawful public administration.

4.4 NIST: Risk management as an operational discipline

The United States National Institute of Standards and Technology (NIST) published the Artificial Intelligence Risk Management Framework (AI RMF 1.0) in 2023. Although not legally binding, the framework has become a global reference point because it translates abstract principles into operational risk management functions.
The NIST framework organizes AI governance into four functions:

- Govern (assign responsibility, policies, and oversight)
- Map (identify context, stakeholders, and potential harms)
- Measure (assess performance, bias, and risk)
- Manage (mitigate risks and monitor over time)

For public-sector use, this structure is particularly valuable because it mirrors existing administrative practices such as internal controls, audit cycles, and risk registers. NIST explicitly recognizes that AI risks evolve over time due to data drift, changing social contexts, and system updates making continuous monitoring and reassessment a core requirement. While NIST is technology-neutral, its framework strongly supports the idea that higher-impact systems require deeper governance effort, aligning closely with risk-based approaches adopted elsewhere.

4.5 The European Union: Risk-based legal regulation

The European Union's Artificial Intelligence Act (Regulation (EU) 2024/1689) represents the most comprehensive binding legal framework for AI governance to date. Its central innovation is a risk-based regulatory structure, which classifies AI systems into:

- Unacceptable risk (prohibited uses)
- High risk (strict pre-deployment and operational obligations)
- Limited risk (transparency obligations)
- Minimal risk (largely unregulated)

For public administration, the EU AI Act is especially significant because it explicitly identifies many government use cases such as biometric identification, law enforcement support, migration and border control, and access to public services as high-risk. High-risk systems are subject to mandatory requirements, including:

- documented risk management and data governance
- technical documentation and record-keeping
- human oversight measures
- accuracy, robustness, and cybersecurity standards

Although designed for EU member states, the Act has global relevance. It provides a clear demonstration that advanced legal systems are moving away from voluntary guidelines toward enforceable, proportionate controls, while still allowing innovation in low-risk contexts.

4.6 Canada: Administrative impact assessment and recourse

Canada offers one of the most practical public-sector governance tools through its Directive on Automated Decision-Making and the associated Algorithmic Impact Assessment (AIA). Rather than regulating AI broadly, Canada focuses on administrative decision systems and their impact on individuals.

The AIA requires departments to assess systems before deployment based on:

- the nature of the decision
- the degree of automation
- the impact on rights, benefits, or entitlements

Depending on the assessed impact level, agencies must implement graduated requirements, including human review, public notice, explanation to affected individuals, and formal recourse mechanisms. The Canadian approach is widely cited because it operationalizes risk-based governance in a way that is directly usable by line ministries and agencies. For developing states, the key lesson is that impact assessment can be implemented through policy and procurement rules, without requiring new legislation or advanced technical infrastructure.

4.7 Convergence across frameworks: a shared governance logic

Despite differences in legal form and institutional origin, the international standards reviewed above converge on a common governance logic for public-sector AI:

1. Risk-based differentiation: not all AI systems require the same level of control.
2. Human accountability: public officials remain responsible for outcomes.
3. Proportional transparency and explainability: stronger where impact is higher.

4. Pre-deployment assessment: risks must be identified before systems go live.
5. Ongoing monitoring and auditability: governance does not end at deployment.
6. Access to recourse: affected individuals must be able to challenge decisions.

This convergence provides a strong empirical and normative foundation for the governance framework developed in this paper. Rather than proposing a new or competing standard, the paper builds on what the international system already agrees upon and translates it into an implementable model for developing administrations.

5. A Risk-Based Governance Model for Artificial Intelligence in Public Administration

5.1 Rationale for a risk-based model

International experience demonstrates that blanket approaches to governing artificial intelligence in government either permissive or restrictive are ineffective. Governments that treat all AI systems as equally risky tend to over-regulate low-impact uses and under-govern high-impact ones. Conversely, administrations that rely on informal discretion or vendor assurances face elevated risks of opaque decision-making, biased outcomes, and loss of public trust.

Across OECD guidance, NIST's risk management framework, the European Union's AI Act, and Canada's Directive on Automated Decision-Making, a consistent governance logic emerges: AI systems should be governed according to their potential impact on individuals, society, and public institutions. This risk-based approach allows governments to scale beneficial uses of AI while applying proportionate safeguards where administrative consequences are most serious.

For developing states, a risk-based model is especially appropriate. It does not require comprehensive new legislation or advanced technical capacity at the outset. Instead, it relies on administrative instruments policy rules, procurement conditions, accountability assignments, and audit processes that can be implemented incrementally within existing public management systems.

5.2 Defining risk tiers for government AI use

The proposed model classifies public-sector AI systems into three risk tiers based on administrative impact, not on the complexity of the technology used.

Level A — Low-risk AI (baseline controls)

Low-risk systems support internal government functions and do not directly affect individual rights, benefits, or obligations.

Typical use cases include:

- internal analytics for staffing, inventory, or logistics planning;
- routing or prioritising citizen inquiries without affecting outcomes;
- summarisation or drafting tools used solely for internal purposes.

Governance rationale:

Errors in these systems may reduce efficiency but are unlikely to cause direct harm to individuals or undermine legal rights.

Minimum controls:

- basic data governance and quality checks;
- cybersecurity and access controls;
- routine performance monitoring.

These controls align with OECD guidance that low-impact AI applications can be governed through standard digital governance and IT risk management practices.

Level B — Medium-risk AI (standard controls)

Medium-risk systems influence administrative processes or support decision-making but do not independently determine outcomes.

Typical use cases include:

- case triage or prioritisation tools;
- automated drafting of notices or decisions reviewed by officials;
- decision-support systems used by frontline staff.

Governance rationale:
While human judgment remains central, these systems shape workflow, attention, and recommendations, creating a risk of automation bias or uneven treatment if left unchecked.

Minimum controls:

- documented purpose and scope of use;
- named accountability owner within the agency;
- mandatory human review before final decisions;
- testing for accuracy and bias prior to deployment;
- audit logs documenting system use and overrides.

This tier reflects practices observed in OECD and World Bank case studies, where decision-support systems delivered productivity gains only when embedded within accountable administrative processes.

Level C — High-risk AI (enhanced safeguards)
High-risk systems influence or materially shape decisions that affect rights, safety, livelihoods, or legal status.

Typical use cases include:

- eligibility or termination of public benefits;
- sanctions, penalties, or enforcement actions;
- law-enforcement or regulatory profiling and targeting;
- biometric identification or identity verification;
- any system whose output can cause adverse administrative consequences.

Governance rationale:
Errors, bias, or opacity in these systems can produce serious harm, including unlawful denial of services, discriminatory outcomes, reputational damage, or erosion of procedural fairness. International frameworks including the EU AI Act and Canada's AIA consistently classify such systems as requiring the strongest governance controls.

5.3 Minimum safeguards for high-risk government AI
For Level C systems, deployment should be conditional on the implementation of the following minimum safeguards, which together form a non-negotiable governance baseline.

1. **Named accountability ownership**
   A senior public official must be formally designated as accountable for system outcomes. Responsibility cannot be delegated solely to vendors or technical teams.
2. **Meaningful human oversight**
   A human decision-maker must be able to review, override, and correct AI-supported outcomes. Oversight must be practical and protected from organisational pressure to "follow the system."
3. **Pre-deployment impact assessment**
   A structured assessment must evaluate the system's purpose, affected groups, potential harms, and mitigation measures before deployment. Canada's Algorithmic Impact Assessment provides a widely cited benchmark for this practice.
4. **Explainability proportionate to impact**
   Agencies must be able to explain, in plain language, how the system influences decisions, what data is used, and how errors are addressed especially to affected individuals.
5. **Data quality and fairness controls**

Controls must include checks for representativeness, error rates, and drift over time, as well as fairness testing across legally relevant groups where permissible.

6. **Security and auditability**
   Systems must incorporate access controls, logging, incident reporting, and secure data handling consistent with public-sector cybersecurity standards.
7. **Procurement and vendor accountability**
   Contracts must require disclosure of system limitations, testing evidence, audit access, change-control procedures, and clear liability arrangements.
8. **Citizen grievance and redress mechanisms**
   Affected individuals must have accessible channels to request review, challenge outcomes, correct records, and receive explanations. Complaints data should be treated as a governance signal.

These safeguards reflect convergent requirements across OECD, UNESCO, NIST, EU, and Canadian frameworks and are designed to prevent the most common governance failures observed in early public-sector AI deployments.

5.4 A practical "no-go" rule for public administration

To prevent blind automation, the model establishes a clear administrative rule:
AI systems must not make final, unreviewable decisions in cases affecting rights, benefits, sanctions, safety, or reputation.
Instead, such systems may flag, prioritise, or recommend but a human official must decide, record reasons, and enable review. This rule directly addresses the most damaging failure mode identified in international practice: automated decisions that become effectively unchallengeable.

5.5 How the model aligns with international practice

The proposed risk-based model does not introduce new normative principles. Rather, it operationalises what international governance frameworks already require:

- OECD guidance supports proportional safeguards linked to impact.
- NIST provides the operational structure for risk management.
- The EU AI Act demonstrates enforceable differentiation by risk.
- Canada's AIA shows how impact-based governance works in day-to-day administration.
- World Bank and UNDP evidence confirms that governance capacity, not algorithmic sophistication, determines outcomes.

By translating this convergence into a simple administrative model, the framework provides a practical pathway for developing states to govern AI effectively without delaying adoption or compromising accountability.

6. Implementing Risk-Based AI Governance in Public Administration

6.1 Why implementation not principles is the binding constraint

International experience shows that the primary barrier to effective AI governance in the public sector is not the absence of principles, but the gap between high-level guidance and day-to-day administrative practice. Governments frequently adopt ethical guidelines or strategy documents without embedding them into procurement rules, operational workflows, audit systems, and accountability structures. As a result, AI systems are deployed faster than the institutions required to control them.
Evidence from OECD, World Bank, and UNDP assessments consistently indicates that governments achieving sustained benefits from AI are those that sequence governance reforms alongside deployment. These administrations do not wait for comprehensive legislation or advanced technical capacity; instead, they rely on incremental institutional

measures that can be implemented within existing public management systems. This section sets out a phased implementation roadmap designed explicitly for developing-state contexts.

6.2 Phase I (0–6 months): Establish basic governance controls

The first phase focuses on visibility, accountability, and procurement discipline. These measures require limited technical investment but produce immediate governance benefits.

Key actions include:

1. **Adopt a formal AI policy note for the public sector**
   Governments should issue a short, binding policy instrument defining AI use in government, establishing risk categories (low, medium, high), and setting minimum governance requirements per category. International practice shows that early clarification of scope and responsibility reduces ad hoc and vendor-driven deployments.
2. **Mandate impact assessment for medium- and high-risk systems**
   Before deployment, agencies should complete a structured impact assessment covering purpose, affected groups, decision points, and mitigation measures. Canada's experience demonstrates that such assessments can be implemented through administrative directives without new legislation and serve as an effective gatekeeping tool.
3. **Create a centralized AI registry**
   A simple registry should record all AI systems used by government agencies, including system purpose, risk level, owner, and deployment status. The United States' federal AI inventory illustrates how registries improve transparency and enable oversight by revealing where AI is used and which systems affect rights or safety.
4. **Update procurement templates and contracts**
   Procurement documents should require vendors to disclose system limitations, provide testing evidence, allow audit access, and comply with change-control procedures. World Bank procurement guidance emphasizes that many AI governance failures originate at the contracting stage rather than during technical deployment.

Implementation outcome:

By the end of Phase I, governments should know what AI systems exist, who is responsible for them, and which systems require stronger controls.

6.3 Phase II (6–12 months): Build assurance and monitoring capacity

The second phase focuses on operational assurance ensuring that AI systems perform as intended over time and that risks are detected early.

Key actions include:

1. **Introduce standard testing protocols**
   Agencies should test AI systems for accuracy, error rates, bias, and security before deployment and at defined intervals thereafter. NIST's risk management framework provides a practical structure for embedding such testing into routine operations.
2. **Establish ongoing monitoring mechanisms**
   Monitoring should track performance drift, data changes, unusual outcome patterns, and incident reports. OECD case evidence shows that AI systems degrade over time if monitoring is absent, particularly in dynamic administrative environments.
3. **Train managers and frontline staff**
   Training should focus on when to rely on AI outputs, when to question them, and how to document overrides. Studies of public-sector AI pilots demonstrate that productivity gains depend heavily on staff understanding AI as a support tool rather than an authoritative decision-maker.

4. **Integrate AI oversight into internal audit functions**
   Internal audit units should review compliance with governance requirements, including documentation quality, override practices, and complaint trends. This approach leverages existing institutional structures rather than creating new oversight bodies prematurely.

Implementation outcome:
By the end of Phase II, governments should be able to demonstrate evidence of performance, bias control, and active human oversight for deployed systems.

6.4 Phase III (12+ months): Strengthen independent oversight and public accountability
The third phase consolidates governance maturity by strengthening independent review, transparency, and citizen protection.

Key actions include:

1. **Establish a multi-disciplinary review mechanism**
   High-risk AI systems should be periodically reviewed by a body combining legal, audit, technical, and domain expertise. International practice shows that cross-functional oversight reduces responsibility gaps and improves institutional learning.
2. **Publish transparency information for high-impact systems**
   Where legally feasible, governments should publish summaries of high-risk AI systems, their purpose, safeguards, and avenues for redress. Transparency requirements under the EU AI Act reflect growing international expectations for public disclosure.
3. **Formalize citizen grievance and redress channels**
   Citizens must be able to request explanations, challenge outcomes, and correct data. Canada's administrative framework demonstrates that effective recourse mechanisms reduce harm and improve trust while also serving as a quality signal for agencies.
4. **Use complaints and appeals as governance data**
   Complaint volumes, appeal success rates, and override patterns should inform system improvements and policy adjustments. OECD analyses emphasize that feedback loops are essential for sustainable AI governance.

Implementation outcome:
By the end of Phase III, AI governance becomes a routine element of public accountability, rather than a special or experimental activity.

6.5 Institutional roles and responsibilities
Clear allocation of responsibility is critical to prevent governance gaps. International practice supports the following division of roles:

- **Central policy authority (e.g., cabinet office or coordinating ministry):**
  Sets rules, approves high-risk deployments, and maintains the AI registry.
- **Line ministries and agencies:**
  Own AI systems, service outcomes, and compliance with governance requirements.
- **Digital/IT units:**
  Provide technical support, cybersecurity, and monitoring tools.
- **Audit and legal units:**
  Review compliance, fairness, and handling of complaints and appeals.
- **Frontline staff:**
  Retain decision-making authority, document overrides, and ensure procedural fairness.

This structure reflects established public management principles and avoids over-centralization or diffusion of responsibility.

6.6 Managing capacity constraints realistically

International evidence consistently shows that governments face skills shortages, data limitations, and legacy systems. The appropriate response is sequencing, not delay. UNDP and World Bank experience indicates that governments should:

- begin with low-risk, high-volume use cases (e.g., backlog clearance, drafting support);
- build assurance capacity incrementally; and
- defer high-impact applications until governance mechanisms are in place.

This approach avoids both reckless deployment and institutional paralysis.

6.7 Measuring success: governance indicators

Governments should assess AI governance success using measurable indicators rather than policy statements. International guidance suggests monitoring:

- service performance (processing times, backlog reduction);
- decision quality (error rates, appeal and reversal rates);
- fairness (outcome disparities where legally permissible);
- trust signals (complaints, satisfaction surveys);
- security (incidents and recovery time).

These indicators align with audit and performance management standards and provide defensible evidence of governance effectiveness.

7. Discussion and Policy Implications

7.1 Reframing AI in government: from innovation tool to regulated public instrument

The evidence reviewed in this paper supports a clear conclusion: artificial intelligence in public administration should be understood and governed as a public instrument, not as a neutral efficiency technology. While much of the early policy discourse framed AI primarily as an innovation opportunity, international experience now shows that its most significant impacts arise where it shapes administrative discretion, allocates public resources, or influences decisions affecting rights and livelihoods.

Across jurisdictions, the core governance failures have not stemmed from the technical novelty of AI, but from institutional gaps unclear accountability, weak oversight, poor integration into administrative law principles, and insufficient avenues for review and correction. The risk-based model developed in this paper responds directly to these failures by re-centering governance on administrative impact rather than technological sophistication.

This reframing has important implications for policy design. It suggests that governments do not need to "solve AI" in the abstract. Instead, they must ensure that AI-enabled processes comply with the same standards of legality, fairness, transparency, and reviewability that already apply to other high-impact administrative tools.

7.2 Resolving the efficiency–accountability tension

A recurring concern in the literature is the perceived trade-off between efficiency gains and accountability safeguards. Some policy debates imply that stronger governance will slow innovation or reduce administrative productivity. The international evidence examined in this paper does not support this assumption.

Cases documented by the OECD, World Bank, and national governments indicate that efficiency gains are most durable when governance is explicit, not when it is relaxed. AI systems that deliver sustained improvements such as reduced response times, backlog clearance, and improved targeting are those embedded in workflows with clear human oversight, documentation, and auditability. By contrast, systems deployed without such controls are more likely to generate errors, provoke legal challenges, or be withdrawn after public controversy.

The implication is that accountability mechanisms should not be treated as constraints on performance, but as enablers of reliable scale. Risk-based governance allows low-impact uses

to proceed quickly while ensuring that high-impact systems are subject to proportionate scrutiny. This differentiation is central to avoiding both regulatory paralysis and uncontrolled automation.

7.3 Why developing states require a distinct governance approach

The literature often assumes governance capacities typical of advanced administrations, including mature data infrastructures, strong regulatory agencies, and abundant technical expertise. This assumption limits the applicability of many proposed AI governance models to developing-state contexts.

The framework advanced in this paper responds to this gap by emphasizing administrative feasibility. It relies on policy instruments that are already familiar to public administrations impact assessments, procurement rules, internal audits, and complaint-handling mechanisms rather than on new technical authorities or complex regulatory regimes. This design choice reflects empirical findings from multilateral institutions showing that sequencing and institutional discipline matter more than technological sophistication.

For developing states, the policy implication is clear: AI should not be used as a shortcut around institutional reform. Instead, AI deployment should be paced to reinforce existing administrative controls and to strengthen accountability over time. Where governance capacity is weak, high-impact applications should be deferred until minimum safeguards are in place.

7.4 Implications for procurement and vendor relationships

One of the most consistent findings across international case studies is that governance failures often originate at the procurement stage. Contracts that lack clear requirements for documentation, testing, audit access, and change control effectively transfer power over public decision processes to vendors.

The risk-based model presented here directly addresses this issue by embedding governance obligations into procurement and contract management. This approach aligns with World Bank and OECD guidance emphasizing that public-sector AI governance is inseparable from procurement discipline. From a policy perspective, this shifts responsibility back to the state, reinforcing the principle that governments cannot outsource accountability.

For policymakers, the implication is that AI governance reforms should prioritize procurement templates and contracting practices early in the reform sequence, rather than focusing exclusively on downstream technical audits.

7.5 Human oversight as an institutional design problem

Much of the literature calls for "human-in-the-loop" oversight without specifying how it should function in practice. The evidence reviewed in this paper suggests that oversight fails when it is symbolic rather than operational. Effective human oversight requires institutional design choices: protected authority to override automated outputs, clear documentation requirements, escalation rules for high-impact cases, and performance metrics that value judgment rather than blind compliance. Without these elements, frontline staff may default to automation bias even when formal oversight provisions exist.

The policy implication is that governments must design oversight mechanisms that change incentives and workflows, not merely add review steps on paper. This insight is particularly relevant for developing administrations where staffing pressures and hierarchical cultures may otherwise encourage deference to automated systems.

7.6 Contribution to the AI governance literature

This paper contributes to the growing literature on AI governance in three ways.

- First, it provides a public administration–centred framework that integrates AI governance into existing principles of administrative law and public management, rather than treating AI as a separate regulatory domain.

- Second, it advances a risk-based operational model grounded in documented international practice, bridging the gap between high-level principles and implementable controls.
- Third, it offers a developing-state–appropriate pathway, emphasizing sequencing, feasibility, and institutional learning rather than idealized governance capacity.

Together, these contributions move the discussion beyond normative aspiration toward actionable governance design.

8. Conclusion

Artificial intelligence is no longer a marginal or experimental feature of public administration. Evidence from multiple jurisdictions demonstrates that governments are already using AI at scale to manage service delivery, compliance, fraud detection, and policy planning, including in contexts that directly affect citizens' rights, safety, and access to public benefits. These developments make AI governance an immediate administrative concern rather than a future regulatory question.

The analysis in this paper shows that the primary risks associated with public-sector AI do not arise from algorithmic complexity alone, but from institutional weaknesses: unclear accountability, insufficient oversight, opaque decision processes, and inadequate mechanisms for review and correction. These risks are present across all administrations but are particularly acute in developing states, where capacity constraints, fragmented data systems, and procurement asymmetries heighten the likelihood of administrative failure.

In response, this paper has advanced a practical, risk-based governance model tailored to public administration. Drawing on convergent international standards and documented practice, the model differentiates AI systems according to administrative impact and aligns safeguards accordingly. It specifies minimum, non-negotiable requirements for high-risk uses, including named accountability ownership, meaningful human oversight, pre-deployment impact assessment, proportionate explainability, data quality and fairness controls, security and auditability, enforceable procurement obligations, and accessible citizen redress. Importantly, the model is operational rather than aspirational: it relies on governance instruments already familiar to public administrations and can be implemented incrementally within existing institutional frameworks.

The phased implementation roadmap presented in the paper demonstrates that effective AI governance does not require immediate comprehensive legislation or advanced technical infrastructure. Instead, it depends on sequencing reforms that prioritize visibility, accountability, assurance, and oversight. International experience indicates that governments which adopt this approach are more likely to sustain efficiency gains while avoiding the most damaging failure mode in public administration high-impact automation that becomes effectively unreviewable.

The central contribution of this paper lies in reframing AI in government as a regulated public instrument rather than an informal efficiency shortcut. By embedding AI within established principles of administrative law and public management, the proposed framework offers a credible pathway for governments particularly in developing contexts to harness AI's benefits while preserving legality, fairness, and public trust. As AI continues to reshape state capacity, the durability of its impact will depend less on technological advancement than on the strength of the institutions that govern its use.

References

Berryhill, J., Bourgery, T., & Hanson, A. (2019). *Hello, world: Artificial intelligence and its use in the public sector*. Organisation for Economic Co-operation and Development.

https://www.oecd.org/content/dam/oecd/en/publications/reports/2019/11/hello-world_7734f030/726fd39d-en.pdf
European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
Government of Canada, Treasury Board of Canada Secretariat. (2019). *Directive on automated decision-making*. Government of Canada. https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/guide-scope-directive-automated-decision-making.html
Government of Canada, Treasury Board of Canada Secretariat. (2020). *Algorithmic impact assessment (AIA)*. Government of Canada. https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html
National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)*. U.S. Department of Commerce. https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf
Organisation for Economic Co-operation and Development Observatory of Public Sector Innovation. (2024). *Robot Alice: Bid, contract and notice analyser*. OECD OPSI. https://oecd-opsi.org/innovations/robot-alice-bid-contract-and-notice-analyser/
Organisation for Economic Co-operation and Development. (2019). *OECD principles on artificial intelligence*. OECD Publishing. https://www.oecd.org/en/topics/sub-issues/ai-principles.html
Organisation for Economic Co-operation and Development. (2025a). *Governing with artificial intelligence: Are governments ready?* OECD Publishing. https://www.oecd.org/en/publications/2025/06/governing-with-artificial-intelligence_398fa287.html
Organisation for Economic Co-operation and Development. (2025b). *AI in public service design and delivery*. In *Governing with artificial intelligence: Are governments ready?* OECD Publishing. https://www.oecd.org/en/publications/2025/06/governing-with-artificial-intelligence_398fa287/full-report/ai-in-public-service-design-and-delivery_09704c1a.html
UNESCO. (2022). *Recommendation on the ethics of artificial intelligence*. United Nations Educational, Scientific and Cultural Organization. https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence
United Kingdom, Department for Work and Pensions. (n.d.). *Using automation for good: Clearing pension backlogs with automation*. UK Government. https://careers.dwp.gov.uk/using-automation-for-good/
United Nations Department of Economic and Social Affairs. (2024). *E-government survey 2024: Addendum on artificial intelligence and digital government*. United Nations. https://desapublications.un.org/sites/default/files/publications/2024-10/Addendum%20on%20AI%20and%20Digital%20Government%20%20E-Government%20Survey%202024.pdf
United Nations Development Programme. (2025a). *AI for the next generation of public services*. UNDP. https://www.undp.org/sites/g/files/zskgke326/files/2025-12/ai-for-the-next-generation-of-public-services.pdf
United Nations Development Programme. (2025b). *Artificial intelligence landscape assessment (AILA): Methodology note*. UNDP Digital AI Hub.

https://www.undp.org/sites/g/files/zskgke326/files/2025-12/undp-dai-hub-aila-methodology-note-eng-2025.pdf
United States Office of Management and Budget. (2024). *Consolidated federal artificial intelligence use case inventory (as of December 16, 2024)*. Executive Office of the President. https://www.cio.gov/policies-and-priorities/Executive-Order-13960-AI-Use-Case-Inventories-Reference/
World Bank. (2021a). *Artificial intelligence in the public sector: Maximizing opportunities, managing risks*. World Bank Group. https://documents1.worldbank.org/curated/en/809611616042736565/pdf/Artificial-Intelligence-in-the-Public-Sector-Maximizing-Opportunities-Managing-Risks.pdf
World Bank. (2021b). *Artificial intelligence in the public sector: Summary note*. World Bank Group. https://documents1.worldbank.org/curated/en/746721616045333426/pdf/Artificial-Intelligence-in-the-Public-Sector-Summary-Note.pdf