

Artificial Intelligence and the Transformation of Public Governance

Ali Anwar¹

Polina Kovaleva²

¹ Master in Public Policy, Nazarbayev University Graduate School of Public Policy (NU GSPP), Astana, Kazakhstan.

E-Mail: ali.anwaar@nu.edu.kz ORCID <https://orcid.org/0009-0004-6727-0184>

² Master in Public Policy, Nazarbayev University Graduate School of Public Policy (NU GSPP), Astana, Kazakhstan.

E-Mail: polina.kovaleva@nu.edu.kz

ABSTRACT

Artificial intelligence (AI) is already part of daily government work: chatbots, risk scores, forecasting models and tools for sorting applications faster than a human officer. This article looks at how that change is affecting public governance and how it's affecting older public values like legality, fairness, accountability and equal access. The paper is conceptual and literature based. It draws on peer-reviewed works in public administration and regulatory works, as well as official texts from the OECD, UNESCO, the European Union, the Council of Europe, Canada, the United Kingdom, Singapore and the United States. The argument is clear, if not so much so. AI can improve efficiency, support more systematic evidence use and free staff from repetitive processing. The same systems can also reproduce bias and hide reasons for decisions, strain privacy and leave officials uncertain who is responsible when something goes wrong. Digital inequality makes these problems worse, because the people who depend on public services are most likely to be the least equipped to challenge an automated outcome. Scholarship has very well mapped applications and risks, but it is still somewhat less clear how administrative institutions absorb AI without strangling political responsibility. The article reviews uses in service delivery, policy analysis, administrative decisions and prediction and then compares various jurisdictional responses. Technical safeguards are not enough. What we need is institutional work: impact assessment, meaningful human review, public inventories of tools, procurement that doesn't tie agencies into opaque systems, and investment in public sector skills. AI will keep spreading in government. And how much it improves public governance depends less on the models than on the legal and organisational arrangements built around them.

ARTICLE HISTORY

Received **01-05-2025**
Accepted **11-11-2025**
Published **30-12-2025**

KEYWORDS

AI; public governance; digital government; algorithmic decision making; accountability; administrative law; comparative regulation.

CORRESPONDING AUTHOR

Ali Anwar
Ali.anwaar@nu.edu.kz



<https://doi.org/10.68740/jjs11>

1. Introduction

1.1 Background

The role of digital government for a long time has been to put forms online and connect back-office systems to it. That already changed how citizens met the state, but it did not really change how decisions were reasoned. AI is different. Once the system starts classifying, ranking, predicting or generating the text officials depend on, then the machinery of government is no longer just a channel. It starts to participate in judgement. This change was not a sudden one. Dunleavy and Margetts (2023) have also identified data science and AI as part of a third wave of digital-era governance, following previous phases of e-government and platform-based reform. Mergel et al. (2019) make a similar point in that they differentiate digitisation (turning paper into files) from digital transformation and recognise it as a more complete redesign of processes and relationships. AI is closer to that second one. It's not just a new tool on an old desk. International organisations have also noticed the trend. The OECD Recommendation on Digital Government Strategies (2014) has already argued technology needs to be seen as a strategic tool for open and trustworthy government and not just an efficiency tool (OECD, 2014). Now more recently the OECD is explicitly addressing "governing with AI", showing use cases in core government functions and pointing to both enablers and guardrails (OECD, 2025). UNESCO's Recommendation on the Ethics of Artificial Intelligence (2021), adopted by 193 Member States, places human rights (e.g., dignity and human oversight) at the heart of any acceptable AI policy (UNESCO, 2022). Regulation (EU) 2024/1689, the so-called AI Act, has moved the debate from principles to binding, risk-based law (European Parliament and Council, 2024). So the background is not that governments are "about to" meet AI. Many already have. The harder question is what kind of public authority is being produced in the process.

1.2 Research problem

The research problem is very simple. Public institutions are adopting AI faster than they are rebuilding the doctrines, routines and oversight arrangements that used to make administrative power understandable. Efficiency gains are easy to advertise. It is much harder to say, in a concrete case, why a person was denied a benefit, flagged as high risk or pushed down a waiting list, and who can be asked to explain that result. It's not just a technical challenge of "explainable AI." It is also a governance issue. Bovens (2007) defined accountability as a relationship in which an actor must explain conduct to a forum that can question, judge and impose consequences. Algorithmic systems complicate all aspects of that relationship. The actor is mostly department, contractor and model. The forum may not see the relevant data. Consequences may be too late or not at all.

1.3 Research gap

There is a vast amount of literature on AI in the public sector already. Wirtz et al. (2019) drew out application areas and challenge dimensions. Zuiderwijk et al. (2021) discussed implications for public governance and pointed out that much of the work was exploratory and conceptual in nature, and practice-based. Yeung (2018) developed a useful taxonomy for algorithmic regulation. Pasquale (2015) and Eubanks (2018) showed how opacity and automation can have the greatest impact on people who have little institutional power themselves. The gap that still bothers me, and this is where the article tries to sit, is the space between principles and administrative reality. We now have high-level ethics texts and, in some places, hard law. We have case studies of particular tools. What is less developed is a combined view of how AI applications, public-value risks and institutional responses mesh across jurisdictions. Different approaches are discussed but national strategies are treated in practice as finished models and not as incomplete experiments. This paper does not fill this gap with new fieldwork. It tries at least to organise it more carefully.

1.4 Research questions

The article is structured around four questions:

1. In which domains of public governance is AI currently being deployed and what kind of tasks does it have to do?
2. What opportunities does the literature and official evidence point to for such uses?
3. What are the risks and institutional challenges, especially for bias, transparency, privacy, accountability and digital equality?
4. How have jurisdictions attempted to respond, and what does that say about a good governance framework?

1.5 Research objectives

The questions are the objectives. The aim of the paper is to (a) synthesise the current literature and the best-respected institutional sources on AI in public governance. (b) to distinguish applications, opportunities and risks without treating them as a single list of pros and cons. (c) to compare various regulatory and administrative approaches and (d) to provide only a few modest recommendations that public institutions would actually use.

1.6 Significance of the study

The significance is also academic and practical. AI is not another ICT topic for public administration research. It touches discretion, street-level bureaucracy, procurement, legal accountability and the political relationship between citizen and state. And the costs of doing this wrong for practitioners are not an abstract matter. A biased risk score, an unreviewable denial, or a procurement contract that no department can see can ruin trust for years. I should state at the beginning what this article is not. It is not an evaluation of one national programme. It does not test a model on original data. It is a doctrinal and conceptual review written from the perspective of a postgraduate researcher who is trying to keep the claims close to sources that can be checked.

2. Literature Review

2.1 Digital government

Digital government research has moved, rather unevenly, from e-government as service automation to a broader concern with institutional change. The OECD (2014) Recommendation is a useful marker here because it asked governments to see digital technology as a means of opening processes rather than speeding them up. Mergel et al. (2019), based on expert interviews, argued that public managers themselves often describe digital transformation as a redesign of processes, services and relationships rather than simply a technology upgrade. Dunleavy et al. (2006) already illustrated how dependent many states were on large systems-integrator contracts. That older political economy is still important. Margetts and Dunleavy (2024) also pointed out that the recent shift to data science and AI has been influenced by Silicon Valley firms and platform practices, not just by traditional government IT vendors. If that reading is correct, public governance is being drawn to tools and works that were never designed for public law in the first place. One implication is that “digital government” and “AI government” shouldn’t be collapsed. A portal that lets a citizen upload documents is not the same as a model that ranks that citizen’s risk. Both are digital. Only the second starts to occupy the space of administrative judgement.

2.2 AI in public administration

Wirtz et al. (2019) are a common starting point. They identified many application areas in the public sector and grouped challenges in a way that is still familiar: technology and data quality, law and ethics, organisational and social issues. Their overview is a few years old, but the map of use and

problems has not really been replaced. What has changed is the intensity of generative tools and the advent of harder legal instruments. Zuiderwijk et al. (2021) are more cautious about how much we actually know. Their systematic review revealed that the field was growing fast but still lacking explanatory work in the public sector and empirical work. They argued for research on specific forms of AI, not “AI in general”, and for more focus on implementation, risk management and governance modes. That call is still relevant. A chatbot in a tax office and a predictive model in child protection are not the same thing. Bullock and Chen (2024) have recently argued that public administration scholarship is just starting to catch up with the organisational and governance implications of more capable AI systems. Their focus on public values, equity, democracy, transparency, effectiveness is a reminder that adoption metrics are a poor substitute for public value analysis.

2.3 Algorithmic decision-making

Algorithmic decision-making is the point at which AI stops being a back-office convenience and becomes a public-power issue. Yeung (2018) defined algorithmic regulation as systems that manage risk or alter behaviour through continuous computational knowledge production. She distinguished reactive and pre-emptive forms and classified systems by how they set standards, gather information and impose or recommend sanctions. The taxonomy is helpful because it stops us treating “the algorithm” as one thing.

Mittelstadt et al. (2016) mapped the ethical debate around algorithms in inconclusive, inscrutable and misguided evidence, unfair outcomes, transformative effects and traceability. That map is still a good way to organise public sector worries. Officials are often asked to defend a score they cannot reconstruct. Citizens are asked to accept a classification whose evidential basis they cannot see.

The COMPAS debate is the example everyone cites, and for good reason. Angwin et al. (2016) brought the issue into public view. Larson et al. (2016) then set out the analysis of Northpointe’s COMPAS recidivism tool in Broward County, Florida, and reported that Black defendants who did not go on to reoffend were nearly twice as likely as White defendants to be misclassified as higher risk (45% versus 23%). The vendor disputed the interpretation of fairness. That disagreement is itself part of the literature: different statistical definitions of fairness can point in opposite directions, which is an awkward fact for administrators who want a single “unbiased” number.

2.4 Government efficiency

Efficiency is the official reason most AI projects get funded. The promise is familiar: faster processing, fewer repeat contacts, better targeting of inspections, lower cost per transaction. OECD work on governing with AI documents use cases across service delivery, justice, tax and other functions and argues that AI is already visible in many core tasks (OECD, 2025). The G7 Toolkit for AI in the Public Sector, prepared by the OECD and UNESCO, likewise treats service improvement and the redirection of staff time as legitimate aims, provided they sit inside a trustworthy-use framework (OECD & UNESCO, 2024).

The literature is less impressed by efficiency talk when it is detached from institutional capacity. Legacy systems, weak data governance and procurement delays keep coming up as constraints (OECD, 2025). Efficiency that depends on a vendor’s black box may also be a delayed liability. In that sense efficiency is not an opposite of risk. It is often the channel through which risk enters.

2.5 Transparency and accountability

Transparency in this field means at least two different things, and they get mixed up. One is disclosure that a tool exists and is being used. The other is a meaningful explanation of a particular outcome. Pasquale (2015) described a “black box society” in which important scoring systems remain commercially and technically opaque. That diagnosis travels easily into government, especially where agencies buy proprietary models.

Accountability is thicker than transparency. On Bovens's (2007) account, someone must be obliged to explain, someone must be able to question, and something must be able to follow from the answer. Public-law systems already struggle with multi-actor administration. AI adds model developers, data brokers, cloud providers and, increasingly, general-purpose model vendors. The chain of explanation becomes long. The forum, a court, an auditor, a parliament, a citizen, may receive only a summary.

The United Kingdom's Algorithmic Transparency Recording Standard is one attempt to make the first kind of transparency routine by requiring central government bodies to publish records of in-scope algorithmic tools (Government Digital Service, 2023, 2025). That is useful. It is not, by itself, an explanation of a decision.

2.6 Existing scholarly debates

Several debates run through the field and never quite settle.

The first is optimism versus precaution. One side sees AI as the next stage of digital-era governance and a way to rebuild state capacity (Dunleavy & Margetts, 2023). The other side sees automation as a way of scaling administrative harm, particularly against poor and minoritised people (Eubanks, 2018). Both can be true in different sites. The literature sometimes writes as if one has to choose a camp.

The second debate is principles versus enforceable rules. UNESCO (2022) and the OECD AI Principles (OECD, 2019/2024) offer values that almost everyone can sign: human rights, transparency, robustness, accountability. The EU AI Act takes a different route and attaches legal duties to risk categories (European Parliament and Council, 2024). Singapore has preferred adaptable model frameworks rather than a single comprehensive statute (Infocomm Media Development Authority & Personal Data Protection Commission, 2020). Canada built a directive aimed specifically at automated administrative decisions (Treasury Board of Canada Secretariat, 2021).

The third debate is human-in-the-loop as solution or slogan. Human oversight is now a standard recommendation. Whether a busy officer clicking "accept" on a model recommendation counts as oversight is another matter. The literature is better at demanding a human than at specifying what that human must be able to do.

A fourth debate, less often named, is about the unit of analysis. Should we study models, organisations, legal regimes, or political economy? Zuiderwijk et al. (2021) wanted more public sector-focused and multidisciplinary work. That still seems right. A paper that discusses neural networks without procurement, or procurement without administrative law, will miss how government actually adopts AI.

3. Methodology

This article uses a doctrinal and conceptual approach combined with qualitative, literature-based analysis. That choice is partly principled and partly practical. The object is not one dataset or one agency. It is a set of institutional problems that appear across systems. For that kind of question, close reading of scholarly and official texts is a legitimate method, provided the texts are real and the claims stay inside what they can support.

Sources were selected in three layers. The first layer is peer-reviewed work in public administration, regulatory governance and related fields, with priority given to widely cited syntheses and conceptual papers (for example Wirtz et al., 2019. Zuiderwijk et al., 2021. Yeung, 2018. Dunleavy & Margetts, 2023. Mergel et al., 2019. Mittelstadt et al., 2016. Bovens, 2007). The second layer is book-length critical studies that have shaped the debate on opacity and social sorting (Pasquale, 2015. Eubanks, 2018) and investigative analysis that remains central to discussions of risk scoring (Angwin

et al., 2016. Larson et al., 2016). The third layer is authoritative institutional material: OECD digital government and AI texts, UNESCO's ethics recommendation, the EU AI Act, the Council of Europe Framework Convention, Canada's Directive on Automated Decision-Making, the UK Algorithmic Transparency Recording Standard, Singapore's Model AI Governance Framework, and the NIST AI Risk Management Framework.

I did not run a formal PRISMA-style systematic review. That would have been cleaner in one sense, but it would also have excluded the official instruments that now structure practice. The analytical framework is organised around four moves that public administration research has used before: map the uses. identify claimed public value. identify value tensions. compare institutional responses. In the background I keep Bovens's accountability relationship and Yeung's account of algorithmic regulation as reference points. They are not "tested". They are used to keep the discussion from sliding into a technology catalogue.

Two limitations should be admitted. First, a literature-based paper inherits the biases of the literature, including the over-representation of European and North American cases. Second, official strategies describe intentions better than implementation. Where I cannot verify an outcome, I treat the instrument as an approach, not as proof that the approach works.

4. AI Applications in Public Governance

4.1 Public service delivery

The most visible uses are in service delivery. Chatbots and virtual assistants handle first-line questions. Document classification systems sort incoming correspondence. Translation tools help multilingual administrations. Recommendation systems point citizens toward benefits or licences they may be eligible for. OECD mapping work finds public service among the functions where AI use cases are comparatively numerous (OECD, 2025).

On paper this looks uncontroversial. A well-designed assistant can reduce waiting time and make opening hours less important. The difficulty starts when the assistant becomes a gate. If the only practical way to reach a service is through an automated interface, people with low digital literacy, disability, or limited connectivity are not being "served more efficiently". They are being rearranged out of the queue. Eubanks (2018) is blunt about this pattern in US social policy: high-tech tools can profile and punish the poor even when the official story is modernisation.

There is also a quieter organisational effect. Once a front-end tool absorbs routine contact, the remaining human work becomes the hard cases. That can be good. It can also leave departments with less everyday knowledge of how ordinary claims look, which then makes the hard cases even harder.

4.2 Policy analysis

AI is also entering policy analysis: summarising consultations, clustering public comments, scanning large administrative datasets, and generating scenarios. Dunleavy and Margetts (2023) argue that data science reduces the old bureaucratic need to compress information into a few indicators. In principle, that could enrich policy debate.

In practice, two cautions keep returning. Policy analysis is political. A model that ranks "what works" is never only describing the world. it is ranking some outcomes as more important than others. And generated text can sound like analysis without being grounded in evidence the analyst has actually checked. For government, the risk is not only error. It is the production of fluent but shallow briefing material that travels too easily up a hierarchy.

4.3 Administrative decision-making

This is the legally sensitive zone. Automated or semi-automated systems now support decisions about benefits, licensing, tax risk, immigration processing, hiring and, in some systems, aspects of justice administration. The EU AI Act treats many such uses as high-risk when they affect access to essential public services, law enforcement, migration or the administration of justice (European Parliament and Council, 2024). Canada's Directive on Automated Decision-Making was written precisely because federal institutions were starting to automate administrative decisions and needed a common set of transparency, quality and recourse duties (Treasury Board of Canada Secretariat, 2021).

It is important not to exaggerate how "fully automatic" most systems are. Many are decision-support tools. Officers still sign the letter. But if the interface presents a score as the reasonable answer, discretion can shrink without anyone formally abolishing it. That is an old street-level bureaucracy problem in a new form.

4.4 Predictive systems

Predictive systems try to estimate future events: recidivism, benefit fraud, child maltreatment risk, infrastructure failure, emergency demand. Yeung (2018) would place many of these on the pre-emptive side of algorithmic regulation. They do not only record what happened. They try to act on what might happen.

Prediction is attractive to managers because it promises to allocate scarce attention. It is also where some of the most serious fairness disputes have arisen. The COMPAS findings (Larson et al., 2016) are one illustration. Eubanks's (2018) discussion of predictive tools in child welfare and homelessness services is another. Even when a model is statistically "accurate" in one sense, it may still distribute error in ways that track race, class or neighbourhood.

Predictive systems also change time. Government starts treating people as risks before a legal event has occurred. That may be reasonable for flood defence. It is much more uncomfortable in policing or social work, where prediction can become a substitute for evidence.

5. Opportunities

5.1 Efficiency

The efficiency case is the easiest to state and still worth stating. Public organisations face backlogs, ageing workforces and rising demand. AI can automate repetitive classification, extract information from documents, and help schedule inspections. OECD and G7 public sector guidance treats this as a legitimate objective when savings are used to improve services rather than only to cut them (OECD & UNESCO, 2024. OECD, 2025).

Efficiency is an opportunity only if quality holds. A faster wrong decision is not a public-value gain. That sounds obvious, but project business cases often measure throughput and ignore downstream complaint and litigation costs.

5.2 Better service delivery

Better service delivery is not the same as speed. It can mean more consistent answers, 24-hour availability, clearer language, and services that meet people in the channel they already use. For some groups, for example people who find office visits difficult, a well-designed digital pathway is an access gain.

The condition is design. “Better” has to be tested with the users who currently fail in the system, not only with the users who already manage digital services well. Otherwise AI becomes a renovation of the front door that leaves the stairs unchanged.

5.3 Data-driven policy

Governments sit on large administrative datasets that were under-used for years. AI methods can help detect patterns that ordinary reporting misses: emerging service bottlenecks, geographic inequality, or unexpected interactions between programmes. Dunleavy and Margetts (2023) see this as part of a shift toward less compressed, more data-intensive information regimes.

Used carefully, that can support more honest policy. Used carelessly, it can give a scientific appearance to correlations that should not drive intervention. Data-driven policy is an opportunity only where analysts remain able to say “the model is not a good reason to do this”.

5.4 Resource allocation

Scarce inspectors, social workers, auditors and emergency teams cannot be everywhere. Predictive allocation can point them toward higher-risk sites or higher-need households. In infrastructure and disaster management this is relatively easy to defend. In human services it is more complicated, because the “risk” being predicted is often entangled with poverty and prior contact with the state.

Still, refusing to use better information is not a neutral alternative. The opportunity is real if allocation models are treated as hypotheses under review, not as queues that no official is allowed to reorder.

6. Risks and Challenges

6.1 Bias

Bias is the risk that has travelled furthest into public debate, and not without cause. Training data record past administrative behaviour. If that behaviour was uneven. more policing in some neighbourhoods, more benefit investigation of some groups. the model can learn the unevenness and present it as prediction.

Larson et al. (2016) showed how error rates in a justice risk tool can differ by race even when overall accuracy looks similar. Mittelstadt et al. (2016) noted that unfair outcomes are only one of several ethical concerns, but they are the concern citizens feel first. Wirtz et al. (2019) already listed ethical and social challenges among the main obstacles to public sector AI.

A difficulty for government is that “removing bias” is not a single technical task. Fairness metrics conflict. A department can satisfy one definition and violate another. That does not mean nothing can be done. It means bias work has to be treated as an ongoing institutional duty. documentation, testing, contestability. rather than a certificate obtained at procurement.

6.2 Transparency

Opacity has several sources. Some models are hard to interpret even for their designers. Some are interpretable in principle but hidden by commercial confidentiality. Some are hidden simply because no one in the department wrote down what the tool does.

Pasquale (2015) argued that secrecy around scoring systems is itself a form of power. In government that power is public power, which makes secrecy harder to justify. The UK ATRS tries to reduce one layer of secrecy by creating a public record of tools (Government Digital Service, 2023, 2025). The EU AI Act goes further for high-risk systems, including transparency duties and, for many

public deployers, registration and fundamental rights impact assessment (European Parliament and Council, 2024).

Transparency still fails if the disclosed text is unreadable or if explanation arrives only after the decision has already shaped a person's life. A register is a start. It is not the whole answer.

6.3 Privacy

AI systems in government are hungry for data: tax records, health flags, education histories, location traces, previous applications. Linking those sources can improve targeting. It can also produce a state that knows citizens in a way they never agreed to be known.

UNESCO (2022) treats privacy and data protection as part of a human-rights approach to AI. The OECD AI Principles likewise include fairness and privacy among the values that AI actors should respect (OECD, 2019/2024). Legal privacy regimes already exist in many jurisdictions, so the issue is often not the absence of law but the way AI projects strain older notice-and-purpose rules. Once data are reused to train or fine-tune a model, the original purpose can quietly expand.

There is a further problem with generated systems. Staff may paste personal case details into external tools that were never approved for that use. That is not a sophisticated privacy harm. It is a very ordinary one, and it is already happening in more than one administration.

6.4 Accountability

Accountability is the institutional heart of the problem. If a model contributes to a decision, who is the actor in Bovens's (2007) sense? The minister? The senior responsible owner? The data scientist? The vendor? If the forum cannot obtain the features that drove a score, the right to reasons becomes thin.

The EU AI Act tries to pin duties on providers and deployers and requires human oversight for high-risk systems (European Parliament and Council, 2024). Canada's Directive requires impact assessment, notice, quality measures and recourse for automated administrative decisions (Treasury Board of Canada Secretariat, 2021). These are serious attempts. They still leave a practical gap: oversight that is only formal, and recourse that is slow.

I find it hard to see how public accountability survives if departments cannot, in-house, explain the logic of the tools they use. Outsourcing can buy capability. It should not buy the right to say "the system decided".

6.5 Digital inequality

Digital inequality is easy to mention and easy to under-analyse. It is not only about who has a smartphone. It is about who can read an automated notice, who can assemble digital evidence, who has the time and literacy to appeal, and who is already over-represented in the datasets used to train enforcement models.

Eubanks (2018) shows how automated welfare and child-welfare systems can intensify scrutiny of poor families. If AI is rolled out first in high-volume, low-status services. benefits, immigration, minor enforcement. inequality is not an accidental side effect. It is a predictable distributional outcome.

UNESCO (2022) and the OECD principles both stress inclusion and the reduction of inequalities. Those statements matter, but they do not implement themselves. Without offline alternatives, assistance with appeals, and monitoring of who is disadvantaged by automation, AI governance will be formally universal and practically skewed.

7. Institutional and Regulatory Responses

Responses have arrived in layers rather than as one global model.

At the level of principles, the OECD AI Principles, first adopted in 2019 and updated in 2024, set out values of inclusive growth, human rights and democratic values, transparency and explainability, robustness and safety, and accountability, plus recommendations to governments on policy and international cooperation (OECD, 2019/2024). UNESCO's 2021 Recommendation added a global ethics instrument with policy action areas and later work on ethical impact assessment (UNESCO, 2022). The Council of Europe Framework Convention on Artificial Intelligence, opened for signature in September 2024, is the first legally binding international treaty aimed at keeping AI activities consistent with human rights, democracy and the rule of law (Council of Europe, 2024).

At the level of risk management, the NIST AI Risk Management Framework offers a voluntary, use-case-agnostic cycle of govern, map, measure and manage (National Institute of Standards and Technology, 2023). It has been influential beyond the United States because agencies needed something operational while statutes were still being written.

At the level of binding public sector rules, three patterns stand out. The EU has chosen horizontal, risk-based legislation that treats many public-service, justice and law-enforcement uses as high-risk and attaches duties of data governance, documentation, human oversight, transparency and, for specified public deployers, fundamental rights impact assessment (European Parliament and Council, 2024). Canada has chosen an administrative-law route: a Treasury Board directive that conditions automated decision-making inside federal institutions (Treasury Board of Canada Secretariat, 2021). The United Kingdom has combined principle-based policy with a transparency standard that is now mandatory for in-scope central government tools (Government Digital Service, 2023, 2025).

Singapore illustrates a fourth pattern: iterative model frameworks aimed at practical organisational controls, later extended toward generative and agentic systems, with public agencies encouraged to adapt the same baseline (Infocomm Media Development Authority & Personal Data Protection Commission, 2020).

The United States has relied more on executive direction, voluntary standards and sectoral law than on a single AI statute. Executive Order 14110 of 30 October 2023 set a broad federal agenda on safe, secure and trustworthy AI (Exec. Order No. 14110, 2023). NIST's framework remains a central practical reference (National Institute of Standards and Technology, 2023). Subsequent administrations have continued to issue executive instruments on AI, which is a reminder that US federal policy in this area is politically mobile.

OECD (2025) usefully splits what governments need into enablers (data, infrastructure, skills, procurement, partnerships) and guardrails (rules, guidance, transparency, oversight bodies). That split is more helpful than the usual "ethics versus innovation" slogan. Without enablers, rules become paper. Without guardrails, enablers become an acceleration of harm.

8. Comparative Discussion of Selected Jurisdictions

Comparison is risky if it becomes a beauty contest. The jurisdictions below were chosen because they are documented in official texts and because they represent different legal styles: comprehensive legislation, administrative directive, transparency standard, model framework, and executive-plus-standards. They are not a representative sample of the world. That limitation should be kept in view.

Table 1

Selected jurisdictional approaches to AI in public governance

Jurisdiction	Main instrument(s)	Legal character	How public-sector use is treated	Distinctive emphasis
European Union	Regulation (EU) 2024/1689 (AI Act)	Binding regulation, risk-based	Many uses in essential public services, justice, law enforcement and migration classed as high-risk; some practices prohibited	Duties on providers and deployers; fundamental rights impact assessment for specified public deployers; public registration of many high-risk systems
United States	EO 14110 (2023); NIST AI RMF 1.0; sectoral statutes	Executive policy plus voluntary framework; no single comprehensive AI statute	Federal agencies directed through executive instruments; RMF available across sectors	Safety, security, innovation and risk management rather than one horizontal licensing regime
United Kingdom	Algorithmic Transparency Recording Standard; broader digital/AI policy	Mandatory transparency standard for in-scope central government; otherwise principles-led	Public algorithmic tools with significant public effect should be recorded and published	Public explanation that a tool exists and how it is used, in a two-tier record
Canada	Directive on Automated Decision-Making (Treasury Board), with later amendments	Binding administrative directive for federal institutions	Applies to systems that automate or support administrative decisions, including some internal services after amendment	Impact assessment, notice, quality, recourse and reporting inside administrative law
Singapore	Model AI Governance Framework (2019/2020) and later generative/agentive updates	Non-binding model framework with strong official promotion	Public agencies encouraged to adapt the framework; engineering translation by public technology agencies	Practical organisational controls, testing and human accountability rather than a single AI Act
Council of Europe	Framework Convention on AI (2024)	International legally binding treaty for parties	Covers public authorities and private actors acting for them, with a human-rights, democracy and rule-of-law frame	Principles and graduated measures rather than product-style risk annexes

Sources: European Parliament and Council (2024). Exec. Order No. 14110 (2023). National Institute of Standards and Technology (2023). Government Digital Service (2023, 2025). Treasury Board of Canada Secretariat (2021). Infocomm Media Development Authority and Personal Data Protection Commission (2020). Council of Europe (2024).

A few comparative observations follow from the table.

First, almost everyone now agrees on the vocabulary: transparency, human oversight, risk, accountability. The disagreement is about hardness. The EU has been willing to use product-safety style law for AI systems. Canada has used the quieter but sometimes more immediately effective channel of internal administrative control. Singapore has bet on adaptable guidance and tooling. The UK has made publicity of tools a duty before it has made model design a licence condition.

Second, public authorities appear in these regimes in two roles, and the texts do not always keep them apart. Government is a regulator of AI markets. Government is also a deployer. and sometimes a provider. of AI systems. The EU Act is explicit that public administrations can be both (European Commission, n.d.. European Parliament and Council, 2024). That dual role creates conflicts. A state that wants a domestic industry to grow may go gently on the same tools its welfare agency is using on residents.

Third, enforceability is uneven. A regulation with fines looks strong. A directive with reporting and recourse can be stronger in daily administration if someone actually audits it. A model framework can be ignored. It can also be implemented more quickly than a statute. It is too early, on the sources reviewed here, to declare a winner.

Fourth, international instruments are doing boundary work. The OECD principles supply a shared definitional and value language that later laws have used (OECD, 2019/2024). UNESCO supplies an ethics and policy-action template (UNESCO, 2022). The Council of Europe Convention tries to lock AI governance to existing human-rights obligations rather than invent a new catalogue of rights (Council of Europe, 2024). For public governance, that last move is important. It says AI does not suspend administrative justice. It arrives inside it.

9. Discussion

If the sections above are pulled together, a slightly uncomfortable picture appears. AI is already part of public governance. The opportunities are not imaginary. Neither are the risks. The institutional responses are real, but they are younger than the tools they are meant to discipline.

One theme is the relocation of discretion. Street-level officers used to exercise judgement in conditions of scarcity and incomplete information. Predictive and ranking systems do not abolish that condition. They hide it in a score. The officer may still be formally responsible. The practical discretion may have moved to the people who chose the training data, the threshold and the interface. Public administration research has talked about discretion for decades. AI makes the old question sharper: discretion by whom, and under what visibility?

A second theme is the incompleteness of transparency remedies. Publishing that a tool exists is better than silence. It does not tell a claimant why her file was routed to investigation. Explainability research tries to fill that gap, but administrative law needs reasons that a person and a court can use, not only a saliency map that a data scientist finds interesting. I am not convinced the field has solved this. It has named it.

A third theme is political economy. Dunleavy et al. (2006) showed how governments became dependent on large IT contractors. Margetts and Dunleavy (2024) argue that platform firms now shape the next conversion. If public agencies become users of general-purpose models they do not control, the older vendor problem returns in a more concentrated form. Procurement then stops being a technical annex and becomes a constitutional issue of sorts: who may structure the information environment of the state?

A fourth theme is inequality. The literature on bias often starts with demographic error rates. That matters. Digital inequality is wider. It includes who is asked to deal with machines in the first place. If automation is concentrated in mass, low-status services, then “modern government” will be experienced as personal and flexible by some publics and as remote and unchallengeable by others.

There is also a methodological lesson. Zuiderwijk et al. (2021) were right that the field needs more empirical, public sector-specific work. Conceptual papers like this one can organise the problem. They cannot tell us whether Canada’s impact assessments are changing decisions, whether ATRS records are being read, or whether high-risk duties under the AI Act will be absorbed by public bodies or outsourced as compliance paperwork. Those are research questions for the next few years.

Finally, I think it is a mistake to treat “trustworthy AI” as a property of models alone. OECD (2025) is closer to the mark when it talks about enablers and guardrails together. A safe model in an unsafe institution will still produce unsafe government. An unsafe model in a careful institution may be caught. That is an institutionalist conclusion, and it is the one this review supports.

10. Recommendations

The recommendations below are deliberately modest. They are written for public institutions rather than for model developers.

1. Treat high-impact uses as administrative actions, not as IT projects. If a system ranks, denies, targets or predicts in ways that affect rights or access to services, it should pass through the same kind of legal, equality and records scrutiny that a new administrative programme would face. Canada's directive logic is useful here even for governments that do not copy the instrument (Treasury Board of Canada Secretariat, 2021).
2. Publish inventories before celebrating pilots. A public register of algorithmic tools, in the spirit of the UK ATRS, should become normal (Government Digital Service, 2023, 2025). Inventories will be incomplete at first. That is not a reason to avoid them.
3. Require fundamental-rights or equity impact assessment for high-risk public deployments. The EU Act's insistence on fundamental rights impact assessment for specified public deployers is a good institutional habit, not only a European legal detail (European Parliament and Council, 2024). Assessments should be specific about who is likely to be harmed, not copied from a template.
4. Keep a human who can disagree. Human oversight is empty if the officer cannot access case-level reasons, cannot record a departure from the model, and is measured only on throughput. Oversight needs time, training and permission to override.
5. Rewrite procurement so opacity is a defect. Contracts should secure documentation, audit access, exit options and limits on secondary use of government data. If a vendor cannot explain a system well enough for an official to defend it, the system is not ready for public power.
6. Invest in public sector data and skills as seriously as in tools. OECD work keeps repeating that legacy infrastructure and weak skills block trustworthy use (OECD, 2025). That is unglamorous and correct. Without in-house capacity, every safeguard is negotiated from weakness.
7. Build appeal routes that do not assume digital fluency. Recourse has to work for the person who received an automated denial. That means plain-language notices, human contact points, and help with evidence. Otherwise accountability exists only for those who already know how to use it.
8. Monitor distributional effects after deployment. Bias testing at purchase is not enough. Departments should review who is delayed, denied, flagged or sanctioned after a tool goes live, and they should be willing to switch it off.
9. Use international principles as a floor, not as a substitute for local law. The OECD principles, UNESCO's recommendation and the Council of Europe Convention provide a shared language (OECD, 2019/2024. UNESCO, 2022. Council of Europe, 2024). They do not replace administrative procedure, equality law or data-protection duties already in force.
10. Do not automate first where contestation is weakest. If a government wants to learn how to use AI responsibly, high-volume welfare and immigration systems are a poor classroom. Starting in domains with stronger professional counterweights and clearer audit trails is slower and wiser.

11. Conclusion

This article has tried to look at AI in public governance without the usual two speeches: the one that promises a smarter state, and the one that announces the end of public responsibility. Neither speech is adequate.

AI is already changing how services are delivered, how policy material is produced, how administrative decisions are prepared, and how future risk is imagined. The opportunities. efficiency, better access for some users, richer analysis, more focused allocation. are documented in official and scholarly sources. The risks. bias, opacity, privacy strain, broken accountability chains, and digital inequality. are documented too, and not only in speculative ethics papers. Investigative and ethnographic work has shown how automated systems can land on the people least able to push back.

Jurisdictions are responding in different legal dialects. The European Union has written a risk-based statute. Canada has tightened administrative control over automated decisions. The United Kingdom has made algorithmic transparency a publishing duty for much of central government. Singapore has invested in practical frameworks and engineering translation. The United States has

mixed executive direction with voluntary risk-management standards. International organisations have supplied principles and, in the Council of Europe's case, a treaty frame tied to human rights and the rule of law.

The comparative lesson is not that one of these models is finished. It is that public governance of AI is becoming a layered craft: values, risk methods, legal duties, transparency devices, procurement, skills and oversight bodies. Missing any one layer makes the others decorative.

I began by saying that AI participates in judgement. That is the point I want to end on. If machines are going to help the state judge, then the old public-law questions remain in force. Who decided? On what basis? With what chance of challenge? And who is answerable when the answer is wrong? Those questions are not anti-innovation. They are the difference between a more capable public administration and a more automated one.

References

- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias. ProPublica. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal*, 13(4), 447–468. <https://doi.org/10.1111/j.1468-0386.2007.00378.x>
- Bullock, J. B., & Chen, Y.-C. (2024). The brave new world of AI: Implications for public sector agents, organisations, and governance. *Asia Pacific Journal of Public Administration*. <https://doi.org/10.1080/23276665.2024.2356540>
- Council of Europe. (2024). Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (CETS No. 225). <https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence>
- Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2006). *Digital era governance: IT corporations, the state, and e-government*. Oxford University Press.
- Dunleavy, P., & Margetts, H. (2023). Data science, artificial intelligence and the third wave of digital era governance. *Public Policy and Administration*. <https://doi.org/10.1177/09520767231198737>
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- European Commission. (n.d.). AI Act. Shaping Europe's digital future. <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>
- European Parliament and Council. (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>
- Exec. Order No. 14110, 88 Fed. Reg. 75191 (2023). Safe, secure, and trustworthy development and use of artificial intelligence.
- Government Digital Service. (2023). Guidance for organisations using the Algorithmic Transparency Recording Standard. GOV.UK. <https://www.gov.uk/government/publications/guidance-for-organisations-using-the-algorithmic-transparency-recording-standard>
- Government Digital Service. (2025). Publish a record of algorithmic tools using the ATRS. GOV.UK. <https://www.gov.uk/guidance/publish-a-record-of-algorithmic-tools-using-the-atrs>
- Infocomm Media Development Authority & Personal Data Protection Commission. (2020). *Model artificial intelligence governance framework* (2nd ed.).
- Larson, J., Mattu, S., Kirchner, L., & Angwin, J. (2016, May 23). How we analyzed the COMPAS recidivism algorithm. ProPublica. <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>
- Margetts, H., & Dunleavy, P. (2024). The political economy of digital government: How Silicon Valley firms drove conversion to data science and artificial intelligence in public management. *Public Money & Management*. <https://doi.org/10.1080/09540962.2024.2389915>
- Mergel, I., Edelmann, N., & Haug, N. (2019). Defining digital transformation: Results from expert interviews. *Government Information Quarterly*, 36(4), 101385. <https://doi.org/10.1016/j.giq.2019.06.002>
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1–21. <https://doi.org/10.1177/2053951716679679>
- National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). <https://doi.org/10.6028/NIST.AI.100-1>
- OECD. (2014). Recommendation of the Council on digital government strategies (OECD/LEGAL/0406). <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0406>
- OECD. (2019). Recommendation of the Council on artificial intelligence (OECD/LEGAL/0449) (updated 2024). <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

- OECD. (2025). Governing with artificial intelligence: The state of play and way forward in core government functions. OECD Publishing. <https://doi.org/10.1787/795de142-en>
- OECD & UNESCO. (2024). G7 toolkit for artificial intelligence in the public sector. OECD Publishing.
- Pasquale, F. (2015). *The Black Box Society: The secret algorithms that control money and information*. Harvard University Press.
- Treasury Board of Canada Secretariat. (2021). Directive on automated decision-making. Government of Canada.
- UNESCO. (2022). Recommendation on the ethics of artificial intelligence. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—Applications and challenges. *International Journal of Public Administration*, 42(7), 596–615. <https://doi.org/10.1080/01900692.2018.1498103>
- Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance*, 12(4), 505–523. <https://doi.org/10.1111/rego.12158>
- Zuiderwijk, A., Chen, Y.-C., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly*, 38(3), 101577. <https://doi.org/10.1016/j.giq.2021.101577>

Declarations

◆ Funding	This research received no external funding.
◆ Conflict of Interest	The authors declare that they have no known financial or personal conflicts of interest that could have influenced the work reported in this paper.
◆ Data Availability Statement	Data sharing is not applicable to this article, as no new data were created or analysed in this study.
◆ Ethical Statement	Ethical approval was not required because the study is based exclusively on publicly available data and published literature and did not involve human participants.
◆ AI Use Disclosure	Generative artificial intelligence tools were used in a limited and assistive capacity during the preparation of this manuscript, primarily for idea development, language refinement, and grammatical correction. The authors retain full responsibility for the originality, accuracy, analysis, interpretation, and final content of the manuscript.

Journal Policies *JuriSocial Journal of Social Sciences*

OPEN ACCESS & LICENCE

All articles are freely available on publication under CC BY-NC 4.0. Readers may share and adapt the work for non-commercial purposes with proper attribution.

PEER REVIEW

Every research article undergoes double-blind peer review by at least two independent reviewers.

PLAGIARISM POLICY

All submissions are screened with similarity-detection software; plagiarism in any form leads to rejection.

CORRECTIONS & RETRACTIONS

Errors are corrected through published corrigenda; retractions follow COPE guidelines.

PUBLICATION CHARGES

No article publishing fee charged.

COPYRIGHT

Authors retain copyright and grant JJSS the right of first publication.

PUBLICATION ETHICS

JJSS follows the Committee on Publication Ethics (COPE) Core Practices and guidelines.

GENERATIVE AI POLICY

AI tools cannot be listed as authors. Any substantive use must be disclosed; authors remain fully responsible for the content.

ARCHIVING & INDEXING

Content is preserved through [PKP Preservation Network / LOCKSS]. DOIs are registered with Crossref.

DISCLAIMER

The views expressed are those of the authors and do not necessarily reflect the views of the editors or the publisher.

HOW TO CITE THIS ARTICLE *(APA 7th edition)*

Submissions & enquiries editor@jjss.jurisocial.org ◆ <https://socialsciences.jurisocial.org/>