MEDWIN PUBLISHERS
Committed to Create Value for Researchers

Open Access Journal of Data Science & Artificial Intelligence
ISSN: 2996-671X

Constitutional Intelligence as a Governance Architecture for Trustworthy Public-Sector AI

Celio Oliveira*

FedEthics Inc., Ottawa, Ontario, Canada

***Corresponding author:** Celio Oliveira, FedEthics Inc., Ottawa, Ontario, Canada, Tel: +1 (437) 777-9306; Email: Oliveira.celior@gmail.com

Research Article

Volume 4 Issue 1

Received Date: March 31, 2026

Published Date: July 30, 2026

DOI: 10.23880/oajda-16000167

Abstract

Public-sector adoption of artificial intelligence (AI) is accelerating, yet governmental use of AI unfolds within decision environments structured by legality, procedural fairness, privacy, equality, ministerial accountability, and democratic scrutiny. This review advances the concept of constitutional intelligence as a governance architecture for trustworthy public-sector AI. Using a narrative-comparative review of authoritative policy instruments, legal frameworks, and peer-reviewed scholarship from Canada, the United Kingdom, Europe, and major international standard-setting bodies, the article identifies a convergent set of design expectations for public AI: rights-sensitive scoping, ex ante impact assessment, documentation and traceability, human oversight, proactive transparency, and continuous institutional review. The analysis further shows that responsible AI adoption in government depends not only on model-level controls but also on the quality of data stewardship, interoperability, and administrative continuity. These findings are synthesized into a constitutional-intelligence model suited to Westminster-style institutions and other public administrations that must justify automated action under conditions of legal contestability and public accountability. The central claim is that constitutional and democratic requirements should not be treated as external compliance burdens. When embedded into institutional design, they operate as productive constraints that improve legitimacy, implementation discipline, and the long-term trustworthiness of AI-enabled public decision-making.

Keywords: Trustworthy AI; Ethical AI; Constitutional Intelligence; Algorithmic Accountability; Democratic Governance; Data Interoperability

Abbreviations

AI: Artificial Intelligence; ATRS: Algorithmic Transparency Recording Standard; AIA: Algorithmic Impact Assessment; UNESCO: United Nations Educational, Scientific and Cultural Organization; IPC-OHRC: Privacy Commissioner of the Ontario Human Rights Commission; NIST: National Institute of Standards and Technology; RMF: Risk Management Framework.

Introduction

Artificial intelligence is now firmly embedded in public-sector modernization agendas, but governmental adoption differs from private-sector adoption in one decisive respect: public institutions must justify how automated systems

CC BY

affect rights, obligations, access to services, and the exercise of public authority. In Canada, the Treasury Board of Canada Secretariat frames responsible governmental use of AI through explicit attention to values, law, and institutional accountability, while the Algorithmic Impact Assessment (AIA) remains the principal ex ante risk instrument supporting the Directive on Automated Decision-Making for automated administrative decisions [1-3]. In the United Kingdom, the Algorithmic Transparency Recording Standard (ATRS) has moved from voluntary experimentation toward mandatory application across specified central-government contexts, signalling a stronger expectation of proactive disclosure and named senior accountability [4,5].

These developments are part of a wider international shift toward rights-aware, lifecycle governance for AI. NIST's AI Risk Management Framework and its Generative AI Profile position trustworthiness, accountability, transparency, fairness, privacy enhancement, and human oversight at the centre of AI governance [6,7]. United Nations Educational, Scientific and Cultural Organization (UNESCO)'s Recommendation on the Ethics of Artificial Intelligence grounds governance in human rights, human dignity, and the prevention of harm [8]. The Organization for Economic Co-operation and Development (OECD)'s 2023 Digital Government Index emphasizes that durable digital transformation in the public sector depends on coherent governance arrangements, resilient infrastructure, and a prospective approach to emerging technologies such as AI [9]. The Council of Europe's Framework Convention on AI and the European Union's AI Act further reinforce a rights-based and risk-based model for governing AI across the system lifecycle [10,11].

Within Canada, the human-rights dimension of AI governance is becoming more explicit and institutionally actionable. The Ontario Human Rights Commission's Human Rights AI Impact Assessment and the joint Information and Privacy Commissioner of the Ontario Human Rights Commission (IPC-OHRC) Principles for the Responsible Use of Artificial Intelligence require organizations to examine discrimination risks, human oversight, privacy effects, and governance accountability before and during deployment [12,13]. Recent scholarship similarly suggests that trust in public AI depends less on technical performance in isolation than on institutional arrangements capable of rendering automated decision systems intelligible, contestable, and answerable to affected persons and oversight bodies [14-17].

A second Canadian current of thought is equally relevant. A recent CIRANO reflection paper by Dudoit, Labillois, and Oliveira argues that public-sector data interoperability should be treated as strategic infrastructure for responsible AI adoption, digital sovereignty, and coherent action across federal, provincial, and territorial jurisdictions [18]. That intervention expands the governance question beyond model supervision alone. It suggests that public-sector AI is only as trustworthy as the institutional capacity that sustains data quality, interoperability, continuity, and coordinated decision-making. The present article builds on that broader systems perspective while advancing a distinct conceptual proposition: public administrations require an integrated language that binds constitutional legitimacy to operational design. To that end, this review develops constitutional intelligence as a governance architecture for trustworthy public-sector AI. The concept draws from the author's broader framework development in Constitutional Intelligence: A Decision Architecture for Trustworthy AI Governance [19], but the present article is a standalone scholarly synthesis designed for journal publication.

Materials and Methods

This article employs a narrative-comparative review methodology suited to an emerging field in which legal instruments, policy guidance, technical standards, and peer-reviewed scholarship develop in parallel rather than sequentially. The objective was not to perform statistical meta-analysis, but to identify recurring governance expectations across jurisdictions and institutional traditions and to translate those expectations into a coherent public-sector governance model.

The review corpus was assembled according to two selection criteria: a) Context sources had to be either institutionally authoritative or peer reviewed; b) A key consideration was that they had to bear directly on public-sector AI governance, rights-centered digital governance, or institutional conditions for trustworthy AI deployment.

The core policy corpus therefore included Government of Canada guidance on responsible AI, the AIA and the scope guidance for the Directive on Automated Decision-Making, the United Kingdom's ATRS guidance and mandatory-scope policy, National Institute of Standards and Technology (NIST)'s AI Risk Management Framework (RMF) and Generative AI Profile, UNESCO's Recommendation on the Ethics of Artificial Intelligence, the OECD Digital Government Index, the Council of Europe's AI Framework Convention, the European Union (EU) AI Act, and Ontario human-rights and privacy guidance [1-13]. A supplementary scholarly corpus was used to test whether those instruments align with current research on AI governance, trustworthiness, implementation, and institutional design [14-18].

Copyright© Celio Oliveira.

The analysis proceeded in three stages. Pre-design, the corpus was reviewed for recurring governance expectations, including impact assessment, documentation, human oversight, transparency, data stewardship, interoperability, risk tiering, and mechanisms of institutional learning. During the review, these expectations were grouped into operational themes spanning legal authority, evidentiary traceability, rights assurance, administrative continuity, and reviewability. After validation, the themes were synthesized into a proposed constitutional-intelligence architecture informed by the author's original framework development [19]. The method is therefore interpretive and design-oriented: it seeks to transform a fragmented governance landscape into a more analytically coherent and operationally usable model for practitioners and researchers.

Results

Constitutional Conditions of Public-Sector AI

Across jurisdictions, public-sector AI governance is distinguished by a basic constitutional fact: automated systems used by state institutions do not merely optimize internal workflows; they participate in decision environments structured by legality, procedural fairness, equality, privacy, transparency, democratic accountability, and, in some settings, constitutional or quasi-constitutional rights review. The Government of Canada's generative AI guidance explicitly warns institutions to assess legal risks relating to human rights, privacy, intellectual property, and procedural fairness, and to engage legal, privacy, security, and service stakeholders prior to deployment [1]. The scope guidance for the Directive on Automated Decision-Making likewise clarifies that not all AI use is captured by the directive, but that automated support for administrative decisions gives rise to formal obligations [3].

This means that public-sector AI cannot be governed adequately through generic responsible-AI slogans alone. Government systems must be situated within a chain of public authority: what decision is being supported, under what legal authority, using which data, with what degree of human discretion, and through what route of explanation, review, or redress. The OHRC's impact-assessment model is especially useful in this regard because it treats AI harms not as abstract ethical possibilities but as context-dependent risks linked to protected grounds, sectors of public life, and affected populations [12]. The 2026 IPC-OHRC principles extend this logic by requiring safe, human-rights-affirming, privacy-protective, and accountable governance structures across the AI lifecycle [13].

Cross-Jurisdictional Convergence Around Accountable AI

The review corpus shows strong convergence around several governance expectations even where legal traditions and regulatory mechanisms differ. To begin with, institutions are expected to identify and assess risk before deployment. Canada's AIA operationalizes this through structured questionnaires and impact-level scoring [2], while Ontario's human-rights framework adds anti-discrimination scrutiny [12]. Then, institutions are expected to document how and why algorithmic tools are used. The UK ATRS requires public bodies to disclose the purpose, ownership, context, data use, and risk profile of in-scope tools, and links transparency to named organizational responsibility [4,5]. Finally, institutions are expected to maintain meaningful human oversight and traceable accountability. NIST frames this within trustworthy-AI design [6,7], while UNESCO, the Council of Europe, and the EU AI Act connect oversight to broader protections of dignity, rights, democracy, and the rule of law [8,10,11].

Fourth, institutions are increasingly expected to treat data governance and digital capability as foundational rather than peripheral. The OECD's Digital Government Index emphasizes the importance of coherent governance arrangements and resilient digital capacity for technology adoption in the public sector [9]. This matters because weak metadata discipline, fragmented data ownership, undocumented workflows, or poor interoperability can render even technically strong AI systems institutionally untrustworthy. Fifth, institutions are increasingly expected to support proactive transparency and public justification rather than relying exclusively on reactive disclosure. The UK model is notable in this respect because it treats transparency not merely as an accountability obligation, but as a practical mechanism for reducing informational asymmetry and strengthening public confidence [4].

Interoperability, Digital Sovereignty, and Institutional Capacity

The reviewed sources also point toward a system-level governance condition that is often under-theorized in mainstream responsible-AI discussions: trustworthy public AI depends on the quality of the institutional substrate in which it operates. The CIRANO reflection paper by Dudoit, Labillois, and Oliveira is particularly important here because it argues that federal-provincial-territorial data interoperability is not simply an administrative convenience; it is a strategic infrastructure for responsible AI adoption,

Copyright© Celio Oliveira.

public-sector productivity, and digital sovereignty [18]. In that account, fragmented data architecture produces duplication, weakens coordination, obscures accountability, and limits the ability of institutions to act coherently at scale. Conversely, interoperable and governable public data create systemic capacity for planning, coordination, and rights-aware digital transformation.

This insight is analytically significant for constitutional intelligence. If AI systems are deployed into fragmented information environments, the burden of justification increases while the quality of evidence often declines. Public institutions may then find themselves unable to explain how data moved across systems, why one dataset was preferred over another, who owned key decisions, or whether downstream effects were monitored consistently. Under such conditions, technical sophistication does not compensate for institutional fragility. Interoperability, therefore, should be understood not merely as an efficiency level but as a condition of evidentiary integrity and constitutional defensibility in public-sector AI.

Constitutional Intelligence as an Operational Architecture

Constitutional intelligence is proposed here as the governance architecture that binds these fragmented requirements together. The concept treats constitutional and democratic obligations as productive design infrastructure for AI-enabled public decisions. Under this model, trustworthy public AI is not defined solely by model performance, but by whether the surrounding institutional system can explain, justify, supervise, and learn from automated or AI-supported action. The concept is especially relevant to Westminster-style systems, where ministerial responsibility, administrative law, parliamentary scrutiny, access-to-information exposure, and public reporting obligations intensify the demand for intelligibility and reviewability.

Five operational components follow from the review. The first is constitutional scoping: before technical design begins, institutions identify the legal basis, decision context, affected populations, and likely rights exposure of the use case. The second is evidentiary traceability: data provenance, model purpose, documentation quality, and grounds for deployment must be explicit and reviewable. The third is stakeholder and rights assurance: privacy, human-rights, legal, service-design, security, and operational expertise should be engaged early enough to shape architecture rather than merely approve it retrospectively. The fourth is stewardship and administrative continuity: trustworthy AI depends on ownership clarity, metadata discipline, version control, monitoring protocols, and institutional memory that survive staff turnover and leadership change. The fifth is contestable oversight: people affected by AI-supported processes must be able to understand the role of automation and, where appropriate, trigger meaningful review or intervention.

These components align closely with the author's broader framework family: DECIDE-PLUS™ for structured decision logic, FIVE-LAYER™ for institutional accountability and stakeholder integration, IMPACT™ for executive-grade knowledge transfer, and the CHARTER RIGHTS DECISION TREE™ for rights-sensitive AI governance [19]. The contribution of the present article is not to reproduce the book's argument in full, but to translate that architecture into a journal-ready synthesis anchored in current public-sector governance instruments and peer-reviewed scholarship. Viewed in this way, constitutional intelligence is not simply a rhetorical device. It is a systems-design proposition: public-sector AI becomes more trustworthy when institutions require themselves to answer constitutional questions ex ante — who is affected, what authority justifies the use, what evidence supports the tool, what records exist, who is accountable, how the decision can be challenged, and what happens if the system fails.

Implementation Sequence for Public Institutions

A practical implementation sequence follows directly from the reviewed instruments and the constitutional-intelligence model. Step one is use-case scoping: define the decision context, legal basis, intended objective, affected populations, and institutional accountabilities. Step two is layered impact assessment: complete an AIA or equivalent risk assessment and pair it, where warranted, with privacy and human-rights review [2,12,13]. Step three is documentation and disclosure design: record the system's purpose, inputs, outputs, ownership, oversight, limitations, and review mechanisms in a form that supports both internal governance and public transparency [4,5]. Step four is multidisciplinary validation: assess performance, robustness, fairness, procedural fit, and operational consequences with legal, privacy, human-rights, technical, and service-delivery expertise before deployment. Step five is monitored operation and institutional review: establish monitoring protocols, escalation paths, and periodic review cycles so that both operational problems and governance gaps can be detected and remedied over time (Table 1).

Copyright© Celio Oliveira.

Source / Instrument	Core Governance Expectation	Implementation Orientation	Constitutional-Intelligence Relevance
Government of Canada: Responsible AI use and AIA [1-3]	Risk-tiered ex ante assessment; legal, privacy, fairness, and stakeholder review prior to deployment	Structured questionnaires linked to impact levels; directed at administrative decision systems	Provides the Canadian regulatory foundation for constitutional scoping and evidentiary traceability
UK ATRS [4,5]	Proactive disclosure of purpose, ownership, data use, and risk profile; named senior accountability	Mandatory scope for specified central-government tools; transparency as operational standard	Operationalizes proactive transparency and named accountability within the constitutional-intelligence model
NIST AI RMF and Generative AI Profile [6,7]	Trustworthiness, accountability, transparency, fairness, privacy-enhancement, human oversight across the lifecycle: valid, safe, transparent, privacy-enhanced, fair	Risk management integrated across design, testing, deployment, and monitoring rather than post hoc review	Supplies a cross-domain technical-operational vocabulary for trustworthy implementation
UNESCO, Council of Europe, and EU AI Act [8,10,11]	Human-rights-centered and risk-based governance across the AI lifecycle	Treat oversight, proportionality, and remedies as system requirements	Links AI governance to democracy, legality, and the rule of law
OECD and Ontario human-rights / privacy guidance [9,12,13]	Digital capability, rights assessment, privacy safeguards, and organizational readiness	Align data stewardship, rights review, and institutional preparedness	Ensures accountability is sustained institutionally, not only procedurally
CIRANO reflection paper [18]	Interoperability and digital sovereignty as strategic infrastructure for responsible AI	Treat cross-jurisdictional data capability as a condition of coherent and defensible AI adoption	Extends constitutional intelligence from tool governance to systems-level capacity

Table 1: Cross-Jurisdictional Governance Frameworks and Their Constitutional-Intelligence Relevance.

Discussion

The analysis suggests that public-sector AI governance is entering a more mature phase. Earlier debates often cast innovation against regulation, or efficiency against accountability. The policy landscape reviewed here is more nuanced. Across Canada, the United Kingdom, the European governance sphere, and major international standard-setting bodies, the prevailing direction is toward governance models that combine risk management with rights protection, transparency, and lifecycle accountability [1-13].

Contemporary scholarship helps explain why this convergence matters. Lahusen, Maggetti, and Slavkovik argue that trust and trustworthiness in public AI are governance problems, not merely technical properties [14]. Batool, Zowghi, and Bano's systematic review shows that AI-governance solutions are proliferating, yet remain fragmented across legal, organizational, and technical domains [15]. Research on how public organizations implement AI governance likewise finds that process design, stakeholder capability, standards, and institutional arrangements are central to success [16]. The World Economic Forum's recent playbook reinforces this implementation emphasis by encouraging responsible innovation through practical governance enablers rather than through principle statements alone [17]. Recent Canadian reflection adds a further systems dimension by locating trustworthy AI within the broader strategic problem of public-sector data interoperability and digital sovereignty [18].

The principal contribution of constitutional intelligence is therefore integrative. It offers public-sector vocabulary that joins legitimacy and operations, constitutional expectations and administrative methods. For Westminster systems in particular, this matters because ministerial responsibility, parliamentary scrutiny, access-to-information exposure, and rights review produce governance demands that are

Copyright© Celio Oliveira.

poorly captured by corporate AI-governance templates. A public administration that treats those demands as peripheral compliance tasks is likely to struggle with trust, auditability, and implementation coherence. By contrast, an administration that treats them as design infrastructure is more likely to produce AI-enabled systems that are reviewable, explainable, and resilient under scrutiny.

This article has limitations. It is a conceptual synthesis rather than an empirical validation study. It does not test the model against a single live deployment or measure outcomes statistically across agencies. Future research should examine whether constitutional-intelligence practices improve implementation quality, audit readiness, dispute resolution, public trust, or intergovernmental coordination in specific public-sector settings. Comparative case studies across Westminster and non-Westminster jurisdictions would be particularly valuable, as would applied work examining the relationship between interoperability maturity and explainability or audit performance in complex administrative environments.

Conclusion

Trustworthy public-sector AI requires more than technical controls and more than generalized ethical aspiration. It requires governance architecture capable of translating constitutional, legal, and democratic obligations into operational design requirements. The policy and scholarly record reviewed here show strong convergence around a practical core: ex ante assessment, rights-sensitive scoping, traceability, transparency, accountable human oversight, data stewardship, and continuous institutional learning. The concept of constitutional intelligence provides a way to organize those expectations coherently for public administrations.

In that sense, constitutional intelligence is both normative and operational. It is normative because it insists that public AI remain compatible with rights, democracy, and the rule of law. It is operational because it converts those commitments into repeatable governance practices that can shape procurement, deployment, monitoring, and review. For public institutions seeking to adopt AI without weakening legitimacy, that dual function is not a supplementary benefit. It is the central design requirement.

Acknowledgments

The author thanks the wider public-sector and research communities whose work on democratic governance, rights protection, public-sector data interoperability, and trustworthy AI informed this article.

Funding

No external funding was received for this work.

Conflict of Interest

The author declares no conflict of interest.

References

1. (2026) Responsible use of artificial intelligence in government. Government of Canada, Treasury Board of Canada Secretariat.
2. (2026) Algorithmic Impact Assessment tool. Government of Canada, Treasury Board of Canada Secretariat.
3. (2025) Guide on the Scope of the Directive on Automated Decision-Making. Government of Canada, Treasury Board of Canada Secretariat.
4. (2025) Algorithmic Transparency Recording Standard: guidance for public sector bodies. Government Digital Service, United Kingdom.
5. (2024) Algorithmic Transparency Recording Standard (ATRS) Mandatory Scope and Exemptions Policy. Government Digital Service, United Kingdom.
6. Tabassi E (2023) Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1, National Institute of Standards and Technology.
7. Autio C, Schwartz R, Dunietz J, Jain S, M. Stanley (2024) Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. NIST AI 600-1, National Institute of Standards and Technology.
8. (2021) Recommendation on the Ethics of Artificial Intelligence. UNESCO.
9. (2024) 2023 OECD Digital Government Index: Results and key findings. OECD Public Governance Policy, pp: 44.
10. (2024) Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. Council of Europe.
11. (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.
12. (2024) Human Rights AI Impact Assessment. Ontario Human Rights Commission.

Copyright© Celio Oliveira.

13. (2026) Principles for the Responsible Use of Artificial Intelligence. Human Rights AI Impact Assessment.

14. Lahusen C, Maggetti M, Slavkovik M (2024) Trust, trustworthiness and AI governance. Scientific Reports 14: 20752.

15. Batool A, Zowghi D, Bano M (2025) AI governance: a systematic literature review. AI and Ethics 5: 3265-3279.

16. Viale Pereira A, Oliveira AS, Pinto TC, Cunha T (2024) Artificial intelligence governance: Understanding how public organizations implement it. Government Information Quarterly 41(4): 102003.

17. (2025) World Economic Forum, Advancing Responsible AI Innovation: A Playbook. Geneva.

18. Dudoit A, Labillois T, Oliveira C (2026) Public Sector Data Interoperability and Responsible Adoption of AI in Canada: Strategic Overview and Call to Action, 2026PR-02, For Reflection, CIRANO, Montréal, QC, Canada.

19. Oliveira C (2026) Constitutional Intelligence: A Decision Architecture for Trustworthy AI Governance. Altona, MB, Canada: FriesenPress.

Copyright© Celio Oliveira.